HIGHLIGHTS OF IAB IMOSS SEB 2019 JOINT CONFERENCE

EDITED BY: Stefan A. Rensing, Jeffrey Graham Duckett, Bernard Goffinet and Meenu Kapoor

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-995-0
DOI 10.3389/978-2-88966-995-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

HIGHLIGHTS OF IAB IMOSS SEB 2019 JOINT CONFERENCE

Topic Editors:
Stefan A. Rensing, University of Marburg, Germany
Jeffrey Graham Duckett, Natural History Museum, United Kingdom
Bernard Goffinet, University of Connecticut, United States
Meenu Kapoor, Guru Gobind Singh Indraprastha University, India

Citation: Rensing, S. A., Duckett, J. G., Goffinet, B., Kapoor, M., eds. (2021). Highlights of IAB IMOSS SEB 2019 Joint Conference. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-995-0

Table of Contents

frontiers in Plant Science

EDITORIAL
published: 20 May 2021
doi: 10.3389/fpls.2021.694765

Editorial: Highlights of IAB IMOSS SEB 2019 Joint Conference

Meenu Kapoor[1]*, *Jeffrey G. Duckett*[2]*, *Stefan A. Rensing*[3]* and *Bernard Goffinet*[4]*

[1] University School of Biotechnology, Guru Gobind Singh Indraprastha University Sector 16C, Dwarka, India, [2] Department of Life Sciences, Natural History Museum, London, United Kingdom, [3] Department of Biology, Plant Cell Biology, University of Marburg, Marburg, Germany, [4] Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, United States

Keywords: bryophytes, mosses, hornworts, *Funaria, Physcomitrium, Marchantia*, liverworts

OPEN ACCESS

Edited by:
Sonia Garcia,
Instituto Botánico de Barcelona, Consejo Superior de Investigaciones Científicas (CSIC), Spain

Reviewed by:
Isabel Draper,
Universidad Autónoma de Madrid, Spain
Juan Antonio Calleja Alarcón,
Autonomous University of Madrid, Spain

***Correspondence:**
Meenu Kapoor
meenukapoor@me.com
Jeffrey G. Duckett
j.g.duckett@qmul.ac.uk
Stefan A. Rensing
stefan.rensing@biologie.uni-marburg.de
Bernard Goffinet
bernard.goffinet@uconn.edu

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *13 April 2021*
Accepted: *27 April 2021*
Published: *20 May 2021*

Citation:
Kapoor M, Duckett JG, Rensing SA and Goffinet B (2021) Editorial: Highlights of IAB IMOSS SEB 2019 Joint Conference. Front. Plant Sci. 12:694765. doi: 10.3389/fpls.2021.694765

Editorial on the Research Topic

Highlights of IAB IMOSS SEB 2019 Joint Conference

Bryophytes comprise mosses, liverworts, and hornworts that form an ecologically and evolutionarily distinct group from extant vascular plants or tracheophytes and inhabit diverse world-wide geographical locations. Unlike the vascular plants, bryophytes are characterized by a haploid dominant generation (gametophyte) that alternates with a shorter-lived diploid generation (sporophyte). In July 2019 (9th−12th) the International Association of Bryologists (IAB), the International Molecular Moss Science Society (iMOSS), and the Spanish Bryological Association (Sociedad Española de Briología, SEB) jointly held a scientific conference in Madrid, Spain in which bryophyte biologists were invited to share and update the scientific community with their findings related to different topics in bryology that included history, genetics and phylogenetics, taxonomy and systematics, biogeography, checklists, structure and physiology, ecology, climate change, pollution, and conservation. The meeting was attended by 212 researchers of 40 different nationalities. A total of 79 oral lectures were delivered and 85 posters were presented during the 3-day program. Two days prior to the conference (on 7th and 8th July 2019) the IUCN Red-Listing Workshop was conducted by the steering committee of the IUCN SSC Bryophyte Specialist Group that provided training in IUCN methodology to 30 participants of 17 nationalities. The meeting ended with an award ceremony and an excursion. The IAB presented "The Spruce" award and "The Hattori" prize, while iMOSS presented "The Golden Spore" award. "The Spruce" award recognizes an IAB member who has made important contributions to bryology within the first 25 years of his/her/their career and "The Hattori prize" for the best publication in Bryology in the previous 2 years. The winners were Matt Renner (Royal Botanic Garden, Sydney, Australia) and Juan Guerra (University of Murcia, Spain), respectively. The "Golden Spore" award was given to Mitsuyasu Hasebe (National Institute for Basic Biology, Japan) for his pioneering works on evo-devo and the genome of the moss *Physcomitrium patens*. As part of the excursion, on the last day of the meeting (July 12th) participants made a day trip to La Pedriza de Manzanares within the Sierra de Guadarrama National Park near Madrid. A second excursion organized by Modesto Luceño (Pablo Olavide University of Seville) was held after the meeting and extended over 3 days (July 13th, 14th, and 15th) in which 30 participants belonging to 17 different nationalities visited the Sierra de Gredos Natural Park in central Spain.

The oral presentations in the meeting were organized into the broad categories: **Species diversity and evolution, Genetics and functional studies, Development and methods of study**. In each research area selected contributions were presented.

Under the ***Species diversity and evolution section*** we present a collection of articles focused on understanding genetic diversity among natural populations and phylogenetic relationships among subspecies. Linde et al. present data on evolution of the *Marchantia polymorpha* complex

that clarify the phylogenetic relationships between the three subsp. *montivagans*, *ruderalis,* and *polymorpha* on the basis of whole genome analysis. This study highlights the significance of hybridization and introgression between subspecies across evolutionary timescales. Biersma et al. present a global phylogeographic reconstruction for the cosmopolitan moss, *Ceratodon purpureus*, revealing a robust biogeographic structure highlighting latitudinal (e.g., pantropical) and bipolar recurrent connectivities. Alonso-García et al. present the first comprehensive phylogeographic and population genetic study of a hornwort, revealing that Southern Appalachian populations of *Nothoceros aenigmaticus*, arose from migration from the neotropical during the Pleistocene and the allopatry of its unisexuality is thus best explained by multiple migrations. Perera-Castro et al. contribute significantly to our understanding of physiological adaptations of Antarctic mosses, by demonstrating that, these mosses have photosynthetic optima at the highest temperatures and achieve positive carbon assimilation by maintaining low respiratory rates at the lowest temperatures. Vendrell-Mir et al. present a non-biased expression analysis of transposable elements (TE) encoded in the *Physcomitrium* (previously *Physcomitrella) patens* genome at different developmental stages and under diverse stress conditions. They suggest that these elements might contribute to variability in genic regions across the genome. Their work also sheds light on TE polymorphism among four different accessions of *P. patens* from Reute (collected close to Freiburg, Germany), Kaskaskia (IL, USA), Villersexel (Haute-Saone, France), and Gransden (Huntingdonshire, UK). The study presented by Spirina et al. explores the homology of filamentous and foliose paraphyllia that develop on the stems of pleurocarpous mosses (i.e., Hypnaceae) based on responses to exogenous application of ABA, proposing a novel hypothesis, whereby not all paraphyllia are alike and that some are in fact homologous to leaves. Sousa et al. explore the effect of alternative models of evolution of chloroplast genes for a broad phylogenetic sampling of streptophytes. They argue, based on inferences from non-synonymous site nucleotide data and amino acid translation data, for the reciprocal monophyly of bryophytes and extant tracheophytes and for Zygnematophyceae that forms the sister group to land plants. Their work provides a robust framework for further understanding of the evolution of early land plants. Maul et al. assess the influence of regional-scale macroclimatic conditions and local factors on the biodiversity of epiphytes and non-epiphytic liverworts in forests in Uganda. They demonstrate that richness of the former ecological group is driven by regional climatic factors, and that of the latter by local microhabitat-specific factors. Based on their study they propose that models relying on macroclimatic factors would only perform well for epiphytic liverworts. Lang et al. present the first population genetic study of a dioicous moss, *Dicranum scoparium*, exhibiting facultative nannandry, that is developing perennial normal sized and ephemeral, epiphytic, dwarf males. They reveal extensive gene flow among female and dwarf male populations, while normal size males tend to be mostly clonal. These authors conclude that each type of males contributes differently but significantly to sexual reproduction, which may explain the persistence of facultative nannandry in this system.

Under ***Genetics and functional studies***, Haas et al. present a genome-wide atlas of Single Nucleotide Polymorphisms (SNP) identified in genic regions of five different accessions of the moss *Physcomitrium patens* and the different laboratory cultivated pedigrees derived over years from the original Gransden accession (collected by H.L.K. Whitehouse in Gransden wood in Huntingdonshire, United Kingdom in 1962). This study shows that prolonged vegetative propagation of plants in laboratories can result in accumulation of deleterious somatic mutations and that annual sexual reproduction may help to purge these mutations. On the basis of identification of characteristic SNPs in different pedigrees and accessions the authors provide a chart with signature SNPs that may facilitate accession/pedigree identification based on RFLP tests. Singh et al. provide insights into the transcriptome landscape of *P. patens* plants harboring mutation in the cytosine DNA/RNA methyltransferase *PpDNMT2* that has been previously shown to be involved in stress recovery. The authors also present quantitative proteomic analysis and shed light on the role of *PpDNMT2* in regulating oxidative stress and maintenance of ion homeostasis by affecting genes and pathways involved in signal perception and transduction.

Under the ***Development and methods section***, Kirbis et al. contrast gene expression in gametophytic and sporophytic generations and across four sporophytic developmental stages of two mosses, *P. patens* and *Funaria hygrometrica*, that exhibit extreme, i.e., reduced and complex, respectively, architecture of their sporophyte. Their analyses suggest that heterochronic expression of major developmental regulators may play a major role in shaping the contrasting complexity of the sporophyte, providing a robust basis for testing whether similar mechanisms drive convergence in sporophyte reduction across mosses. Ignatov et al. reconstruct the pattern of cell divisions in the peristome forming amphithecial layers of the sporangium in a representative of the aperistomate Gigaspermaceae, which are sister to true arthodontous mosses. Their study reveals surprising similarities with the development of the nematodontous peristome of the Polytrichopsida, a lineage that arose from an earlier split in the diversification of mosses.

King et al. describe a semi-automatic and sensitive Object-Based Image Analysis (OBIA) method for long term assessment of vegetation health particularly of plants of short stature such as mosses growing in extreme environmental locations inhospitable for field work. Using this non-destructive methodology, the authors present an assessment of moss health and an evaluation of changes in vegetation cover from 2003 to 2011. Werner et al. present a modified epiGBS protocol for studying DNA methylation changes in genomes of non-model organisms that is cost-effective and that can be useful for laboratories with limited financial resources. Using existing softwares and newly designed algorithms the authors provide proof-of-concept by analyzing methylation in DNA fragments from two almond cultivars *Prunus dulcis* (Mill.) D.A. Webb cv. "Desmayo Largeta" and cv. "Penta" at two different stages during dormancy release.

AUTHOR CONTRIBUTIONS

MK wrote the first draft. JD, SR, and BG revised the article. All authors finalized the article.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 May 2020
doi: 10.3389/fpls.2020.00495

Population Genomics and Phylogeography of a Clonal Bryophyte With Spatially Separated Sexes and Extreme Sex Ratios

Marta Alonso-García[1†], Juan Carlos Villarreal A.[1,2*†], Kenneth McFarland[3] and Bernard Goffinet[4]*

[1] Département de Biologie, Université Laval, Quebec City, QC, Canada, [2] Smithsonian Tropical Research Institute, Ancón, Panama, [3] Department of Ecology and Evolutionary Biology, The University of Tennessee, Knoxville, TN, United States, [4] Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, United States

OPEN ACCESS

Edited by:
Kathleen Pryer, Duke University, United States

Reviewed by:
Santiago Ramírez Barahona, National Autonomous University of Mexico, Mexico
Dirk Carl Albach, University of Oldenburg, Germany

***Correspondence:**
Marta Alonso-García
marta.alonso-garcia.1@ulaval.ca
Juan Carlos Villarreal A.
juan-carlos.villarreal-aguilar@bio.ulaval.ca

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *06 December 2019*
Accepted: *01 April 2020*
Published: *08 May 2020*

Citation:
Alonso-García M, Villarreal A JC, McFarland K and Goffinet B (2020) Population Genomics and Phylogeography of a Clonal Bryophyte With Spatially Separated Sexes and Extreme Sex Ratios. Front. Plant Sci. 11:495. doi: 10.3389/fpls.2020.00495

The southern Appalachian (SA) is one of the most biodiversity–rich areas in North America and has been considered a refugium for many disjunct plant species, from the last glacial period to the present. Our study focuses on the SA clonal hornwort, *Nothoceros aenigmaticus* J. C. Villarreal & K. D. McFarland. This hornwort was described from North Carolina and is widespread in the SA, growing on rocks near or submerged in streams in six and one watersheds of the Tennessee (TR) and Alabama (AR) Rivers, respectively. Males and female populations occur in different watersheds, except in the Little Tennessee (TN) River where an isolated male population exists ca. 48 km upstream from the female populations. The sex ratio of 1:0 seems extreme in each population. In this study, we use nuclear and organellar microsatellites from 250 individuals from six watersheds (seven populations) in the SA region and two populations from Mexico (23 individuals). We, then, selected 86 individuals from seven populations and used genotyping by sequencing to sample over 600 bi-allelic markers. Our results suggest that the SA *N. aenigmaticus* and Mexican plants are a nested within a clade of sexual tropical populations. In the US populations, we confirm an extreme sex ratio and only contiguous US watersheds share genotypes. The phylogenetic analysis of SNP data resolves four clusters: Mexican populations, male plants (Little Pigeon and Pigeon river watersheds) and two clusters of female plants; one from the Little Tennessee and Hiwassee Rivers (TR) and the other from the Ocoee (TR) and Coosa (AR) Rivers. All clusters are highly differentiated (Fst values over 0.9). In addition, our individual assignment analyses and PCAs reflect the phylogenetic results grouping the SA samples in three clades and recovering males and female plants with high genetic differentiation (Fst values between 0.5 and 0.9 using microsatellites and bi-allelic markers). Our results point to Pleistocene events shaping the biogeographical pattern seen in US populations. The extreme sex ratio reflects isolation and highlights the high vulnerability of the populations in the SA.

Keywords: clonality, hornwort, *Nothoceros*, population genomic, sex separation, SNPs, stacks

INTRODUCTION

Skewed sex ratios, dioicy and clonality are prominent features of plant reproductive systems with profound implications on the ecology, evolution and survival of the species (Barrett, 2015; Petry et al., 2016; Scott et al., 2018). In genetically determined sexual systems, the theoretical sex ratio is expected to be close to 1:1 (female: male), if we assume a similar cost to the production of male and female individuals (Fisher, 1930; Scott et al., 2018). However, sex ratios can be significantly skewed (Bisang and Hedenäs, 2005; Sinclair et al., 2012). For example, 61% of 234 species of angiosperms distributed among 61 families display a male-biased ratio (Field et al., 2013). In bryophytes, most of the species studied display a strong female-biased ratio. In nearly 85% of the 103 moss and liverworts species surveyed (Bisang and Hedenäs, 2005), females outnumbered males in natural populations. One striking example is the desert moss, *Syntrichia caninervis* Mitt. In the Mojave Desert only 11 males were found among 700 individuals examined (Stark et al., 2005). The causes and genetic repercussions of such skewed ratios are now being explored (Bisang and Hedenäs, 2005; Bisang et al., 2015, 2017; Baughman et al., 2017; Brzyski et al., 2018).

In contrast to flowering plants, where separate sexes (diploid dioecy) is rare (Renner and Ricklefs, 1995), most bryophytes species display separate male and female plants (haploid dioicy). Nearly 70% of mosses and liverworts and 40% of hornworts species display separate sexes (Wyatt and Anderson, 1984; Villarreal and Renner, 2013; Laenen et al., 2016). In bryophytes, the gametophyte is free-living, and the male gametes are motile, and consequently, sexual reproduction is constrained by the availability of water between the sexes. Although this constraint may be lessened by degrees of dehydration tolerances of the sperm (Shortlidge et al., 2012), or by invertebrates acting as vectors (Cronberg et al., 2006), the rate of sexual reproduction decreases with the distance between sexes, with an upper limit of several meters (Van der Velde et al., 2001), but typically within a few centimeters (Longton, 1997). Proximity of the sexes is therefore imperative for successful sexual reproduction in bryophytes. The widespread occurrence of clonality may exacerbate the distance between sexes with a preferential growth of female plants (e.g., in *Marchantia*; Brzyski et al., 2018), biasing sex ratios and accentuating the difficulty of sexual reproduction. Despite these functional impediments, spatial separation of sexes (SSS) is widespread among all groups of plants (Bierzychudek and Eckhart, 1988; Groen et al., 2010; Brzyski et al., 2018). The evolution of dioicy may be also driven by the benefits of the SSS (Bierzychudek and Eckhart, 1988), as SSS may favor niche partitioning and avoid intersexual competition for example (Bierzychudek and Eckhart, 1988; Rogers and Eppley, 2015).

The female-biased ratio, the widespread dioicy (Bisang and Hedenäs, 2005; Bisang et al., 2017) and the prominent role of clonality in bryophytes may account for the incidence of SSS with over 20 cases reported at a large geographic scale (Schuster, 1983; Stark et al., 2005; Villarreal et al., 2014; Bisang et al., 2015). In fact, the most extreme examples of SSS in plants are found in bryophytes. For instance, the leafy liverwort *Plagiochila corniculata* Dumort. is known only from female plants in North America and all European individuals bear only male sex organs (Schuster, 1983, 1992). Clonality, limited dispersal and SSS may amplify the non-random spatial distribution of genotypes, or spatial genetic structure (SGS). In fact, the type of mating system seems to contribute to SGS, with higher incidence of SGS in dioecious flowering plants (Nazareno et al., 2013; Dering et al., 2016).

Some repercussions of extreme separation of sexes are SGS or reduced sexual reproduction (Barrett, 2015). In addition, in flowering plants, the proliferation of sterility mutations (the "use-it-or-lose-it" hypothesis; Eckert, 2002) and events of local extinction (Hu et al., 2017) are other consequences of the spatial separation of sexes. In bryophytes, the occurrence of such extreme cases of SSS is intriguing and leads to a number of evolutionary questions about the occurrence of sex-specific SGS in bryophytes with SSS or the time populations can be maintained clonally. Given the limited sexual reproduction and the little genetic diversity, will these populations face local extinction? In the liverwort *Marchantia inflexa* Nees & Mont., some of these questions have already been answered: sexes are differentially clumped (males tend to be near male plants in larger patches), suggesting a fine scale trade-off between sexual reproduction and clonality (Brzyski et al., 2018).

The present study focuses on the southern Appalachian (SA) hornwort *Nothoceros aenigmaticus* J. C. Villarreal and K. D. McFarland, a perennial dioicous hornwort that is locally abundant on rocks along or in streams. Schuster (1992) described this species and considered it to be endemic to the SA, i.e., North Carolina (NC), Georgia (G), and Tennessee (TN) (**Figure 1**). Later, the species was found in Mexico, as well as in alpine neotropical regions (Villarreal et al., 2012a,b). The original description was based entirely on female plants collected in the Little TN River, which connects to the TN River (**Figure 1**). Male plants were subsequently found in watersheds of the Pigeon and Little Pigeon River system (Renzaglia and McFarland, 1999) and in the Little TN by Ken McFarland at Balsam Mt (TENN Herbarium). The latter being ca. 48 km upstream from the nearest female population in that same watershed. In Mexico, the species occurs in two populations where male and female plants grow together (Villarreal et al., 2012a,b).

The isolation of sexes in SA may account for the lack of known sexual reproduction, which would further be compromised by the premature abortion of motile male gametes (Renzaglia and McFarland, 1999). In addition, complete maturation of hornwort antheridia requires a short desiccation, which is absent in the *quasi* permanent aquatic habitat of *N. aenigmaticus*. These factors might contribute to the males' inability to develop and release functional sperm cells (Renzaglia and McFarland, 1999). In the absence of fertilization, and thus sporophyte and subsequent spore formation, dispersal is achieved only via fragmentation of the gametophyte margins, and hence reproduction is purely clonal. Headwater springs of most of the streams investigated had large colonies of *N. aenigmaticus* present. Colonies were observed long distances down these streams, thus downstream dispersal is likely the primary direction for dispersing vegetative propagules. Based on that, we used a phylogenetic and population genomic

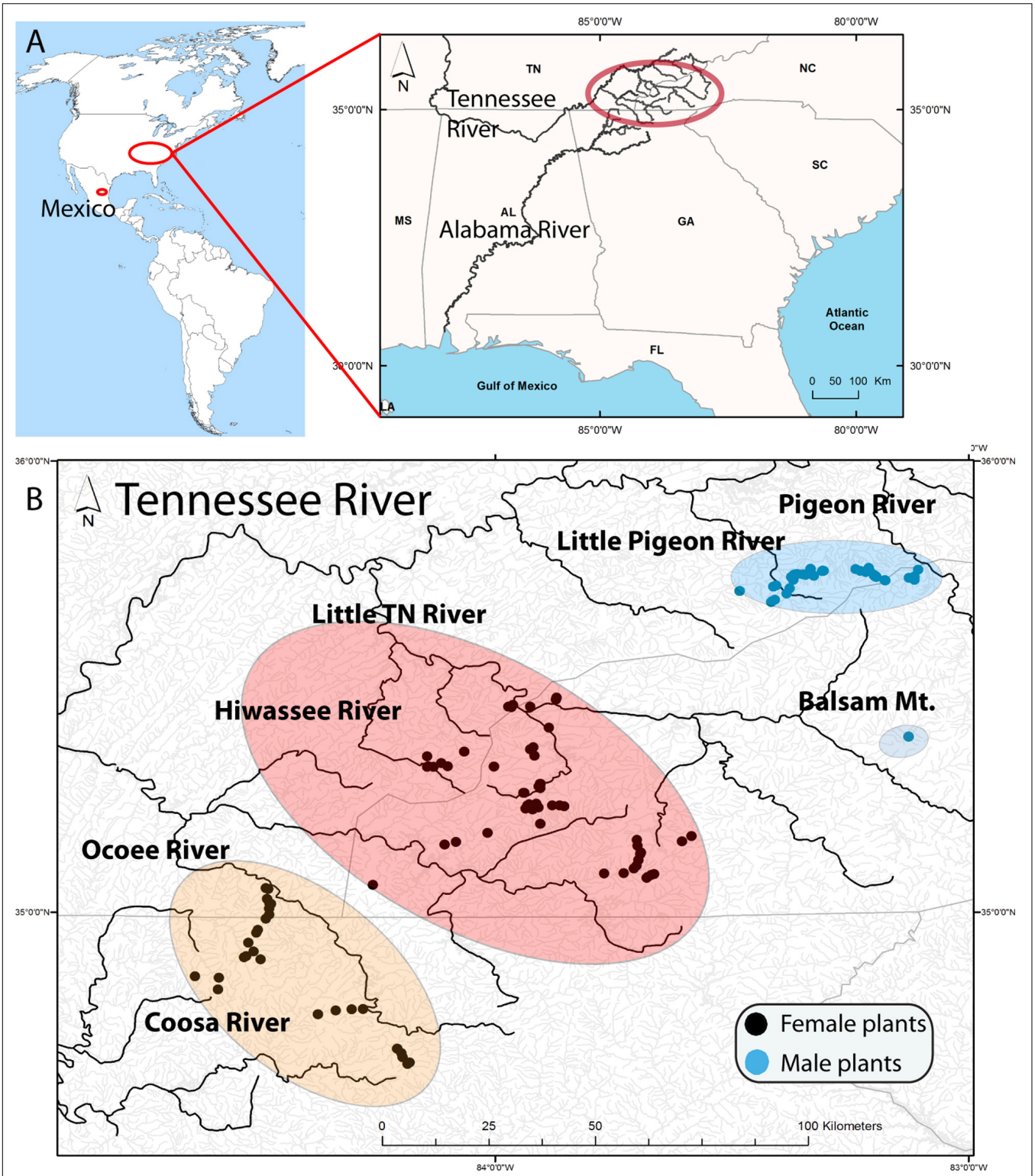


FIGURE 1 | (A) Location of samples in America with a map of the Southern United States illustrating the Tennessee River (Mississippi River) and the Coosa River (Alabama River) that drain in the Gulf of Mexico. The upper Tennessee River, where the samples were collected, is highlighted by a red oval. **(B)** Distribution of *Nothoceros aenigmaticus* in the Southern Appalachians showing the spatially segregated sexes. Most watersheds drain in the Tennessee River, except the Coosa River which drain into the Alabama River. The populations are labeled Pigeon River and Little Pigeon River (surrounded by a blue oval), Balsam Mountain Preserve (blue oval), Little TN River and Hiwassee River (surrounded by a reddish oval) and Ocoee River and Coosa River (surrounded by an orange oval). Black dots represent sampling sites for female plants and blue dots for male plants.

approach (i) to estimate the divergence time of the split between US and Mexican populations, (ii) to verify that US populations of *N. aenigmaticus* are strictly clonal and (iii) to assess and confirm the extreme sex ratio and clonal structure of the species in the SA.

We used a published dataset (Villarreal and Renner, 2014) to date the timeframe of the SA origin of *N. aenigmaticus*. Divergence time between Mexican and US populations will show when the apparent geographic isolation of the female and male gametophytes in the US started and potentially explain the causes of the extreme sex allopatry in Appalachian plants. We analyzed 10 microsatellite loci distributed across all genomic compartments (plastid, mitochondrion and nucleus) for 273 collections of the species from seven populations in the SA and two in Mexico. Additionally, we selected 86 individuals (80 from the US and six from Mexico) and used genotyping by sequencing (GBS) methods to sample SNP markers to document species' genetic diversity and contrast it to the estimates based on microsatellites.

MATERIALS AND METHODS

Dating Analyses

We used a dataset derived from Villarreal et al. (2012a) and Villarreal and Renner (2014) to provide a time frame for the divergence of the US and Mexican *N. aenigmaticus*. Dating relied on a Bayesian divergence time estimation as implemented in BEAST 1.8.3 (Drummond et al., 2012). Our inferences are based on all markers for a total of 3,398 aligned nucleotides of four plastid loci (*rbc*L, *trn*L-F intron, *rps*4-trnS spacer and *mat*K). We restricted the sampling to the genus *Nothoceros* (R. M. Schust.) J. Haseg, and three outgroup genera (*Dendroceros* Ness, *Megaceros* Campb. and *Phaeomegaceros* R. J. Duff, J. C. Villarreal, Cargill & Renzaglia) based on Villarreal et al. (2015). The ingroup included 23 accessions belonging to American species of *Nothoceros* (**Supplementary Table S1**), such as *N. canaliculatus* J. C. Villarreal, Hassel & N. Salazar (1), *N. endiviifolius* (Mont.) J. Haseg. (2), *N. fuegiensis* (Steph.) J. C. Villarreal (2), *N. minarum* (Nees.) J. C. Villarreal (2), *N. renzagliensis* J. C. Villarreal, L. V. Campos & Uribe (1), *N. schizophyllus* (2), *N. superbus* (1), *N. vincentianus* (4) and, *N. aenigmaticus* (7). For the latter, we sampled accessions from high tropical elevations (>3,000 m, Paramos) as well as samples from US and Mexico. One individual of *N. giganteus* (Lehm. & Lindenb.) J. Haseg. from New Zealand was also included. We used a single calibration on the most recent common ancestor (MRCA) of the genus *Nothoceros*. We implemented a uniform distribution with a minimum age of 16 Mya and a maximum age of 42 Mya, following the dates estimated by Villarreal et al. (2015) for the crown age of *Nothoceros*.

Two tree priors were used (birth-death and Yule priors). Runs were performed using a relaxed clock model; the latter with uncorrelated and log-normally distributed rate variation across branches. The Markov Chain Monte Carlo (MCMC) was run for 30 million generations using the relaxed clock model, with parameters sampled every 1,000 generations. Tracer v1.5 (Rambaut and Drummond, 2007) was used to assess effective sample sizes for all estimated parameters (over 200) and to judge the percentage of burn-in for tree constructions. Trees were combined in TreeAnnotator 1.6.1 (Rambaut and Drummond, 2010), and maximum clade credibility trees with mean node heights were visualized using FigTree 1.3.1 (Rambaut, 2015). The high posterior densities (HPDs) or the smaller interval that contains 95% of the sampled values were reported. We performed a maximum likelihood (ML) analysis on RAxML (Stamatakis, 2014) using the Cipres Science Gateway.[1] The substitution model GTR+ CAT was used with the unlinked partitions and the statistical support was evaluated using 500 ML bootstrap replicates using the same substitution model.

Identification of Populations and Sampling Effort

We defined nine populations, seven from the SA: Pigeon River, Little Pigeon River, Balsam Mountain Preserve, Little TN River, Hiwassee River, Ocoee River and Coosa River; and two from Mexico: Zempoala and Diamantes (**Figure 1** and **Table 1**). Our sampling of *N. aenigmaticus* comprised 273 individuals, of which 250 were from the SA and 23 from two localities in Mexico (**Table 1**). In the SA, the species is present in two main rivers, the TN River and the Coosa River (**Figure 1A**). The former is one of the largest tributaries of the Ohio River and drains into the Mississippi river, while the Coosa River drains into the Alabama River (Gulf of Mexico) (**Figure 1A**). We collected in five watersheds of the Tennessee River (Pigeon River, Little Pigeon River, Little TN River, Hiwassee River and Ocoee River Ocoee), and one watershed of the Alabama River (Coosa River). Some male populations from the Little TN River were collected in

[1]https://www.phylo.org

TABLE 1 | Main characteristics of the populations of *Nothoceros aenigmaticus* sampled in this study, including watersheds, main rivers, sex and number of individuals sampled for microsatellites and GBS analysis.

Population name (watershed)	Main river/country	Sex	No. microsatellites	No. GBS
Pigeon River	Mississippi River (United States)	Males	44	18
Little Pigeon River	Mississippi River (United States)	Males	61	23
Balsam Mountain Preserve	Mississippi River (United States)	Males	10	0
Little TN River	Mississippi River (United States)	Females	62	17
Hiwassee River	Mississippi River (United States)	Females	37	14
Ocoee River	Mississippi River (United States)	Females	28	2
Coosa River	Alabama River (United States)	Females	8	6
Zempoala	Lagunas de Zempoala (Mexico)	Females/ males	18	6
Los Diamantes	Cascada de los Diamantes (Mexico)	Females	5	0

Balsam Mountain Preserve, ca. 48 km upstream from the female populations of this watershed (**Figure 1B**).

All collections in a stream were made from colonies at least 2 m apart. Of the 250 collections, less than 10% were sterile (without sex organs), and these were predominantly aquatic forms. In these cases, samples were cultured in the lab to verify the presence of sex organs. *Nothoceros aenigmaticus* produces abundant archegonia (Schuster, 1992, his Figure 2) and the identification is straightforward. Sex verification for putative male aquatic plants were not possible. Male plants were found exclusively in Pigeon River, Little Pigeon River and Balsam Mountain Preserve. In the remaining watersheds, only female plants were observed (**Figure 1** and **Table 1**).

In Mexico, *N. aenigmaticus* had rarely been collected and seems restricted to high elevation (above 3,000 m) in the Neovolcanic Belt (Villarreal et al., 2012a,b). Despite an intensive search, the species was found just in two localities. The first locality was a dormant Pleistocene volcano in Las Lagunas de Zempoala, state of Morelos, where five different streams were sampled. The second locality was in Cascada de Los Diamantes (hereafter Diamantes), State of Mexico. In the former case, males and females inhabit the same patch and produced sporophytes; whereas in Los Diamantes the patches were exclusively composed of female plants.

Population Genetic Based on Microsatellite Data

Microsatellite Preparation and Sequencing

Three mitochondrial (mit1, mit3, mit4), three plastid (cp44, cp45, cp59) and four nuclear (nuc35, nuc38, nuc42, nuc53) microsatellite loci were selected for use in the population study (Villarreal et al., 2012b). All samples were genotyped by multiplexing and scored using GeneMarker version 1.5 (Soft Genetics, State College, PA, United States). The marker nuc35 had a hexa-nucleotide repeat (GAGCTT). Fragment analyses revealed a single peak but inconsistent changes in repeat number. In order to verify the nature of the repeats in the locus nuc35, we used Sanger sequencing of this microsatellite for all individuals and found indels (3 or 4 bp) in some of the accessions. Manual alignment was carried out using Se-Al version 2.0a11 (Rambaut, 2002); because of low quality at the beginning and ends of the reads the matrix was trimmed to an aligned length of 171 bp. Cross-validation of exemplars from most populations was done with fragment analyses to confirm the presence of a single peak for the locus.

Analyzing the Chlorotype, Mitotype, and Nuclear Microsatellites

Because of the non-recombining nature of the plastid and mitochondrial genomes, each allelic combination at the three plastids (hereafter chlorotypes) and two mitochondrial (mitotypes) loci were treated as an allele at a single locus. The locus mit4 proved to be difficult to score in the Mexico and United States (US) individuals and was excluded from downstream analyses.

For the chlorotype and mitotype microsatellites, haplotype diversity was estimated as the number of haplotypes per geographical region divided by the total number of samples (P). In addition, the number of private haplotypes (G_P); haploid genetic diversity (h) and unbiased haplotype diversity (uh) were calculated.

Φ_{PT} (analogous to F_{st}) was estimated for chlorotypes and mitotypes $\left(\Phi_{PT} = \frac{Vap}{(Vap - Vwp)}\right)$ where *Vap* is the variation among populations and *Vwp* is variation within populations, as suggested for haploid data, Φ_{PT}. All statistical analyses were done with GenAlex 6.4b5 using 1000 permutations (Peakall and Smouse, 2006) and excluding individuals with missing data. From a total of 273 individuals and after filtering for missing-data, 267 individuals were included for the plastid analysis and 263 for mitochondrial.

For the nuclear microsatellites, 255 individuals were included after filtering for missing data; then allele frequencies, the number of alleles per locus (Na) and number of effective alleles (Ne) were calculated using GenAlex 6.4b5 (Peakall and Smouse, 2006).

Clonal Structure Using All Microsatellite Data

Given the clonal nature of the species, all microsatellites were analyzed together to determine the number of unique haplotypes among the clones. Clonal assignments were made using genetic distance under the stepwise model with different thresholds between individuals (Geno Dive version 2.0b2, Meirmans and Van Tienderen, 2004). If the genetic distance between a pair of individuals is below a set threshold, they will belong to the same clone. We tried several thresholds, but all resulted in a drastic reduction of the number of clones (e.g., a threshold of 0–3 gives between 74 and 70 clones, a threshold of 4–5 give 43 clones, a threshold of 6 gives 32 clones and a threshold of 9–10 gives 26 and 27 clones, respectively). One method to choose the threshold is to use the distance to the nearest neighbor. The distances between each sequence is calculated and the histogram is generated, and the threshold is choosing between the modes. In our case, the mode is between thresholds 9 and 12. However, the number of clones is drastically reduced (27 clones) using a threshold of 9. We decided to take the most conservative approach and present no threshold (0), and a threshold of 4 and the threshold of 9 for the all microsatellites.

Clonal diversity for each threshold was quantified from all microsatellites together and values of several indexes, such as Nei's genetic diversity and evenness (eve = eff/G) were estimated. Additionally, we also calculated the ratio between genets (G) and the total number of ramets sampled (N). This ratio (G/N) is another measure of clonality where G is identified as unique multilocus genotypes by genetic markers. As G/N approaches 1/N, the signal of clonality is strong. The indexes were calculated in GenoDive version 2.0b2 (Meirmans and Van Tienderen, 2004).

Fst values were computed in GenoDive version 2.0b2 (Meirmans and Van Tienderen, 2004), and *p*-values estimated based on 1000 permutations using a stepwise mutation model. A K-Means clustering were done by calculating the allele frequencies removing the source of the individual genotypes, the K-means clustering does not assume Hardy–Weinberg equilibrium. The Bayesian information criterion and Calinksi

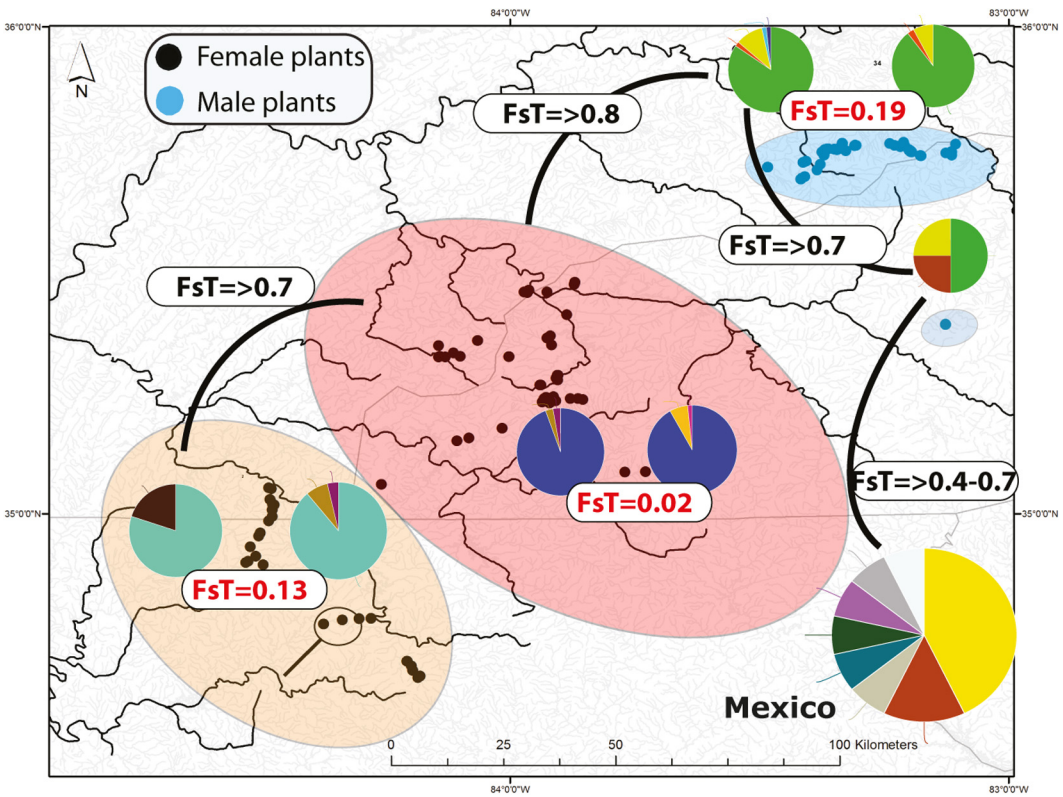


FIGURE 2 | Distribution of microsatellite multilocus nuclear genotypes (MLG) and spatially segregate sexes of *Nothoceros aenigmaticus* in the Southern Appalachians. The pies represent the proportion of nuclear MLGs. Watersheds that share at least one MLG are color-coded. The Pigeon and Little Pigeon Rivers (blue oval); Balsam Mountain Preserve (oval); Little TN and Hiwassee Rivers (reddish oval) and Ocoee and Coosa Rivers (orange oval). The two localities draining into the Coosa River are circled within the Ocoee/Coosa River cluster (orange oval). Black dots represent sampling sites for female plants and blue dots for male plants. Fst values from the analyses of all microsatellites (organellar and nuclear) show low Fst values amongst contiguous watersheds (**Table 2**).

and Harabasz pseudo –F were used to assess the optimal value of K using 50,000 simulated annealings and 20 random repeats. Additionally, an assignment analysis was conducted to assign an individual to a population based on the allele frequencies and determine potential migrants.

Population Genomics Based on GBS Data

Library Preparation and Sequencing

We included 86 collections from seven populations in the GBS approach (**Table 1**). A double digest GBS library (enzymes PstI/MspI) (Abed et al., 2019) was prepared and sequenced on an Ion Proton instrument by the Plateforme d'analyses génomiques [Institut de Biologie Intégrative et des Systèmes (IBIS), Laval University, Quebec City, QC, Canada].

GBS Data Processing

Raw data were assessed for quality with the fastQC version 0.11.3 (Gordon and Hannon, 2010). Reads demultiplexing, filtering, and trimming were accomplished using the *process_radtags* option from the pipeline software Stacks version 2.3 (Catchen et al., 2011). To uniform the reads' length, the sequence length distribution of fastQC was checked and reads were trimmed to 125 bp. The script was run as follow *–inline_null –renz_1 pstI –renz_2 mspI -c -r -t 125 -q -s 10*. Sequences reads were mapped to the nuclear transcriptome of *N. aenigmaticus* (SRA: ERS368224) (Wickett et al., 2014) using the software package BWA version 0.7.17 (Li and Durbin, 2009) and the SAMtools suite (Li, 2011) to convert the output of BWA to BAM format as required for Stacks (Rochette and Catchen, 2017). We created loci and identified SNPs with *gstacks* option (Stacks pipeline) on the full dataset. We concluded the SNPs calling filtering with *populations* (Stacks pipeline) applied to two different datasets, one including all GBS individuals (86) and a second one excluding those from Mexico (80 individuals). We set up the maximum accepted heterozygosity at 0 (*–max-obs-het*) because all samples were haploid gametophytes. Minimum minor allele frequency was 0.05 (*–min-maf*). The setting *populations* provides filtering options to only include loci that occur at certain frequencies in each population. We considered all individual a single population and eliminated loci that were missing from more than 50% of individuals (*–R*).

Phylogenetic Inference

A phylogeny of *N. aenigmaticus* was performed for both dataset (86 individuals from Mexico and US and 80 individuals

from US, with known sex) including only unlinked SNPs (*–write-single-snp*). The matrices were used for a RAxML v8.2.9 (Stamatakis, 2014) maximum likelihood analysis. A bootstrap analysis with 100 replicates was run under the GTR + Γ model. The resulting phylogenetic trees were visualized and rooted using FigTree v1.4.3.

Genetic Diversity and Analysis of Population Structure

The genetic diversity was estimated for the SNP matrix including Mexico and US individuals (86). Statistics were calculated with the Stacks program *populations* including the *–fstats* option to have the Fst values for each pair of populations, with *p*-values describing if the Fst measures were statistically significant according to Fisher's Exact Test. Each watershed was originally considered a different population, then individuals were split into seven groups. At least 50% of the individuals inside a population should share a locus to process it (*–r 0.5*). The maximum accepted heterozygosity was set up at 0 (*–max-obs-het*) and the minimum minor allele frequency was 0.05 (*–min-maf*).

To assess genetic structure and sex-ratio in SA, we excluded samples from Zempoala (Mexico). The SNP matrix for 80 US individuals (410 SNPs) was used hereafter. The SNP matrix was converted into a *genind* object from the R *adegenet* version 2.0.2 (Jombart and Ahmed, 2011) and used to investigate genetic structure among the individuals by a principal component analysis (PCA) using the *dudi.pca* function in the R package *ade4* version 1.7-13 (Dray and Dufour, 2007). Based on PCA results, the genetic diversity statistics were estimated. In this case, contiguous watersheds were considered a single population hence, the 80 individuals were grouped into three clusters: Pigeon and Little Pigeon Rivers (n = 41), Little TN and Hiwassee Rivers (n = 31), and Ocoee and Coosa Rivers (n = 8) (**Figure 1B**). Fst values between pairwise groups were calculated and *p*-values provided by Stacks. The same filters applied before were used (*–max-obs-het 0 –min-maf 0.05 -r 0.5 –fstats*).

We inferred population structure based on nearest neighbor haplotype co-ancestry using the fineRADstructure package (Malinsky et al., 2018), which is specially designed for RADseq data. Following the manual, we first converted the haplotypes output from *populations* (Stacks pipeline) to a *fineRADpainter* input file, reducing the maximum number of SNPs allowed at a locus to 10. We re-ordered the loci with the script provided (*sampleLD.R*), as suggested by authors. Afterward, the co-ancestry matrix (*RADpainter*) was calculated, the individuals were assigned to populations and the coalescence tree was built (*finestructure*). The outputs were loaded into the program Finestructure GUI for visualization.

Clonal Structure Using SNPs Data

We used a similar approach to that used for microsatellites to assess clonal diversity on GenoDive version 2.0b2 (Meirmans and Van Tienderen, 2004). We used the dataset including Mexico and US individuals (86) and set a threshold of 0 using the infinite allele model and using the setting "*clones specific to each population.*" Higher thresholds were evaluated but all resulted in a drastic reduction of the number of clones. Clonal diversity was quantified from the multilocus nuclear genotypes and values of several indexes, such as Nei's genetic diversity and evenness (eve) were estimated. Additionally, we also calculated the ratio between genets (G) and the total number of ramets sampled (N). The indexes were calculated in GenoDive version 2.0b2 (Meirmans and Van Tienderen, 2004).

Pairwise Fst values were also computed in GenoDive version 2.0b2 (Meirmans and Van Tienderen, 2004), and *p*-values estimated based on 1,000 permutations. A K-Means clustering were done by calculating the allele frequencies removing the source of the individual genotypes. The Bayesian information criterion and Calinksi and Harabasz pseudo –F were used to assess the optimal value of K using 50,000 simulated annealings and 20 random repeats. Additionally, an assignment analysis was conducted to assign an individual to a population based on the allele frequencies and determine potential migrants.

Spatial Genetic Structure (SGS) Using All Microsatellite Data

Based on the genetic groups derived from the phylogeny and the PCA analysis, we conducted a SGS test using microsatellites. Spatial autocorrelation analysis of kinship among ramets was carried out using SPAGeDi 1.5. (Hardy and Vekemans, 2002). Contiguous watersheds were examined at the ramet level (e.g., Pigeon and Little Pigeon Rivers; Little TN and Hiwassee Rivers; and Ocoee and Coosa Rivers). The average multilocus kinship coefficient (F_{ij}) was computed according to Loiselle et al. (1995) and averaged within each of the chosen eight distance classes: 0.25, 1, 5, 10, 20, 30, 40, 50 km. Average values of F_{ij} were regressed on the natural logarithm of distance $\ln(d_{ij})$ in order to obtain the regression slope, b, which quantifies the SGS. To test for SGS under the null hypothesis of no correlation of F_{ij} and $\ln(d_{ij})$, the spatial positions of individuals were permuted 10,000 times.

RESULTS

Dating Analyses

The divergence times for the two nodes of interest are shown in **Supplementary Figure S1** and **Table S2**. The MRCA to the crown group from Neotropics, Mexican and US *N. aenigmaticus* populations dates to **3.89** (1.03–7.76) Mya using a birth-death prior or **4.35** (1.16–8.4) Mya with the Yule prior. The diversification of the US/Mexican *N. aenigmaticus* crown group dates to **0.79** (0.061–1.93) Mya using a birth-death tree prior, or **0.88** (0.06–2.13) Mya with a Yule prior (**Supplementary Figure S1** and **Table S2**).

Population Genetic Based on Microsatellites Data

Genetic Diversity

The percentage of missing data for chlorotypes and mitotypes was 2.1 and 3.6%, respectively. No chlorotype or mitotype was shared between male and female populations. The highest chlorotype diversity occurred in Zempoala, Mexico (h = 0.796;

males and females), followed by SA populations in the Little Pigeon (h = 0.496), Pigeon (h = 0.328), and Little TN (h = 0.211) Rivers (**Supplementary Table S3**). The number of mitotypes (i.e., 39) was always higher than the number of chlorotypes (i.e., 24; **Supplementary Table S3** and **Figure S2**). Zempoala, Mexico harbored the highest diversity of mitotypes (h = 0.86), followed by the Little TN (n = 0.589), Little Pigeon (h = 0.517) and Ocoee (h = 0.502) rivers (**Supplementary Table S3**). Eleven chlorotypes and 16 mitotypes were found among all male populations (Pigeon, Little Pigeon and Balsam Mt.) and 10 chlorotypes and 19 mitotypes among female populations (Little TN, Hiwassee, Ocoee, and Coosa Rivers). Likewise, most mitotypes and chlorotypes were restricted to single or contiguous watersheds with limited downstream dispersal. Examples of share widespread chlorotypes are found in Little TN and Hiwassee River (**Supplementary Figure S2**). Similarly, one chlorotype is shared between the Coosa River and Ocoee River and one is restricted to the headwaters of the Ocoee River (**Supplementary Figure S2**). Regarding nuclear diversity, 20 microsatellite alleles are distributed across the four microsatellite loci (**Supplementary Table S4** and **Figure 2**). The missing data composed less than 2% of the total data. The number of alleles per marker ranged between 2 and 10. The most variable marker (nuc35) has 10 alleles (**Supplementary Table S4**).

Population Differentiation

Average Φ_{ST} values for chlorotypes are high (0.592; p. 0.01) with low levels of pairwise Φ_{PT} between Little TN and Hiwassee Rivers (0.00) and between Zempoala and Cascada de Los Diamantes (0.099) and moderate values between Pigeon and Little Pigeon Rivers (0.272). Average Φ_{PT} values for mitotypes are high (0.409; p. 0.01) with low levels of pairwise Φ_{PT} between Little TN and Hiwassee Rivers (0.013), between Zempoala and Cascada de Los Diamantes (0.035) and between Pigeon and Little Pigeon Rivers (0.004) (**Supplementary Table S5**). In general, Φ_{PT} values are low between contiguous watersheds and the differentiation between non-contiguous watersheds is high (**Supplementary Table S5**).

Clonal Structure Using All Microsatellites

Seventy-two, 43 and 26 clones (thresholds 0, 4, and 9) were detected using GenoDive (**Supplementary Table S6**). With no threshold, the number of clones is high in Little TN River (13), Ocoee River (11), Pigeon River (15) and Little Pigeon River (24). With a threshold of 4, the number of clones in Little TN River was drastically lower (five clones). There is evidence of migrants between contiguous watersheds and two cases of potential migrants between non-contiguous watersheds. One migrant shared by Little TN and Pigeon River and a second one shared by Zempoala and Ocoee Rivers. No multilocus genotype was shared between sexes. The Fst values for nuclear data show low values between Pigeon/Little Pigeon River (0.191), Little TN/Hiwassee Rivers (0.026) and Ocoee/Coosa Rivers (0.133). The Fst values are much higher between non-contiguous watersheds (**Table 2**) and a Fst value of 0.556 between Balsam Mountain Preserve and Mexican populations (**Table 2**).

The K-Means clustering recovered six groups (k = 7), namely Little Pigeon, Pigeon Balsam Mountain Preserve, Little TN/Hiwassee, Ocoee/Coosa, and Zempoala/Diamantes (**Supplementary Table S7**). Population assignment analyses recovered migrants between contiguous watersheds, bus also two potential migrants between Little TN and Little Pigeon River and between Coosa River and Zempoala (Mexico).

Population Genetic Based on GBS Data

GBS Inferences

Approximately 88 million sequencing reads were generated for the 86 samples. The total number of reads retained after demultiplexing, filtering and trimming was 64,836,726, varying between 18,885 for sample MA42 to 4,029,405 for sample MA237. Retaining only reads that mapped to the transcriptome reference, reduced the number of reads drastically, e.g., to 2,621 for MA42 and from 4,021,346 to 109,298 for the best mapped sample (MA135). An average of 43% of the reads mapped to the reference transcriptome was retained for *gstacks* (Stacks pipeline) to create 6,157 loci (at least 5X mean coverage), of which 99% had BLAST hits on the transcriptome of *N. aenigmaticus*. **Table 3A** summarizes the number of loci and SNPs for each dataset (including and excluding Mexican samples) before and after filtering with the Stacks pipeline (maximum accepted heterozygosity of 0, SNPs with minor allele frequency greater than 0.05 and more than 50% coverage).

Phylogenetic Inference

The phylogenetic relationships of the 86 samples (Mexico and US individual) were inferred from 491 SNPs (**Table 3A**) with 20.4% of missing-data. The samples are clustered in two highly supported clades (MLBS = 100) corresponding to the Mexican and US individuals of *N. aenigmaticus* (**Figure 3A**). The US samples were further resolved in three well-supported clades (MLBS > 85) (**Figure 3A**): the first includes the male plants from Pigeon and Little Pigeon Rivers, the second one consists of female plants from Little TN and Hiwassee, and the third one contains specimens from Ocoee plus Coosa Rivers (females). Within each clade, the relationships are ambiguous. The matrix of the 80 US samples had 305 SNPs (**Table 3A**) and 19.9% of missing-data, and its analysis yielded the same topology and it is not shown here.

Genetic Diversity and Population Structure

Using the SNP matrix containing 86 individuals, we obtained a total of 6,157 loci. Among them, 2,930 were present in at least 50% of the individuals inside a population (watershed) (filter $-r$ *0.5*). Broadly, levels of nucleotide diversity (π) were low, ranging from 0 (Ocoee River) to 0,05436 (Zempoala) (**Supplementary Table S8**). The number of polymorphic sites inside a population varied from 0 (Ocoee River) to 10 in US watersheds (**Supplementary Table S8**). In Zempoala (Mexico), a population with sympatric sexes, this number increased to 93. Private sites were also particularly abundant among Mexican plants with 307 unique alleles, and comparatively really low within the US populations with values between two to eight (**Supplementary Table S8**). Pairwise Fst values showed little differentiation between contiguous rivers (watersheds), in contrast to the high differentiation between non-contiguous rivers (**Table 2**).

TABLE 2 | Pairwise Fst values among populations of *Nothoceros aenigmaticus*.

	Little TN River	Hiwassee River	Ocoee River	Coosa River	Pigeon River	Little Pigeon River	Balsam Mt. Preserve	Zempoala Mexico	Diamantes Mexico
Little TN River	–	0.152	0.988	0.981	0.981	0.972	–	0.936	–
Hiwassee River	0.026	–	0.9859	0.987	0.980	0.971	–	0.933	–
Ocoee River	0.787	0.819	–	0.057	0.980	0.966	–	0.907	–
Coosa River	0.827	0.883	0.133	–	0.980	0.964	–	0.922	–
Pigeon River	0.828	0.857	0.750	0.785	–	0.120	–	0.932	–
Little Pigeon River	0.803	0.821	0.717	0.762	0.191	–	–	0.919	–
Balsam Mt. Preserve	0.802	0.871	0.697	0.827	0.775	0.725	–	–	–
Zempoala Mexico	0.704	0.743	0.534	0.558	0.589	0.566	0.440	–	–
Diamantes Mexico	0.809	0.873	0.737	0.843	0.760	0.715	0.753	0.356	–

Values below the diagonal for all microsatellites and above the diagonal for SNPs (GBS). The values from microsatellite data were significant after 1000 permutations at a p-value of 0.01.

The PCA plot (**Figure 3C**) based on the US matrix 410 SNPs from 80 samples reflected the most likely phylogenetic relationships (**Figure 3A**). Individuals of *N. aenigmaticus* from the US clustered in three main groups reflecting the contiguity of watersheds: one comprising samples from Pigeon and Little Pigeon Rivers (males), the second one samples from Little TN-Hiwassee Rivers (females) and the third samples from Ocoee/Coosa Rivers (females). Pairwise Fst values among these three groups were high and agreed with phylogenetic and PCA data. Fst value between Little TN-Hiwassee and Ocoee/Coosa Rivers was 0.98 with p-values < 0.005 for 234 of the 239 scaffolds compared; Fst value for Little TN-Hiwassee and Pigeon-Little Pigeon Rivers was 0.97 with p-values < 0.005 for 299 of the 304 scaffolds compared; and the Fst between Ocoee/Coosa and Pigeon-Little Rivers was 0.97 with p-values < 0.005 for 266 of the 274 scaffolds compared. The Fst value between sexes was also high (0.65) (p-values < 0.005 for all the scaffolds compared). In terms of genetic diversity considering three populations (contiguous watersheds), 2,571 loci (of 6,157) were shared by at least 50% of the individuals within each population. The loci were composed of 320,884 sites; 412 variants. Levels of nucleotide diversity (π) were low (<0.006 for variant positions and < 0.00001 for variant and fixed positions) (**Table 3B**). Following the same trend, variation within each population never exceeded 10 polymorphic sites. However, the number of private or unique alleles within a population highly increased when individuals were grouped by contiguous watersheds (**Table 3B**). In this case, it ranged from 99 to 168, compared to two and eight in **Supplementary Table S8** (individuals grouped by watersheds).

According to the fineRADstructure results (**Figure 3B**), the three geographic clusters shared more co-ancestry within each other than between them. Comparison between clusters shows that individuals of *N. aenigmaticus* from Ocoee plus Coosa Rivers (females) share more co-ancestry with individuals from Little TN plus Hiwassee Rivers (females) than with those from Pigeon and Little Pigeon Rivers (males) (**Figure 3B**).

Clonal Structure

The total number of clones (Num) was 31 (when the clones were made specific to each of population) or nine (clones not specific to each population). A single clone occurs in each US river, except Hiwassee and Coosa River (two clones each), which suggests prominent clonality in the species in SA. In Mexico, the population of Zempoala had five clones. The G/N ratio was low (G/N < 0.1) in Pigeon, Little Pigeon, Little TN and Hiwassee Rivers. The collections from Ocoee and Coosa Rivers show higher G/N ratio (i.e., 0.5 and 0.33, respectively; **Supplementary Table S9**), and Zempoala presents the highest ratio (G/N = 0.83). Pairwise Fst values were lower between contiguous watersheds. For example, 0.363 between Little TN and Hiwassee River (**Supplementary Table S10**) and over 0.9 between the non-contiguous watersheds (**Supplementary Table S10**). Population assignment recovered migrants between contiguous watersheds, especially between Little TN and Hiwassee and Pigeon/Little Pigeon River (**Supplementary Tables S11, S12**). The K-Means clustering recovered five groups ($k = 5$), namely Little TN/Hiwassee Rivers, Ocoee River, Coosa River, Little Pigeon/Pigeon Rivers and Zempoala (**Supplementary Tables S11, S12**).

Spatial Genetic Structure Using All Microsatellite Data

At the ramet level, the regression of F_{ij} over the natural logarithm of the geographic distance produced a negative slope for the three contiguous watersheds (**Supplementary Table S13**). A significant average F_{ij} was estimated up to 10 km in Pigeon and Little Pigeon Rivers; in 1 km in Little TN and Hiwassee Rivers, whereas in Ocoee and Coosa Rivers, it was significant in 0.25 km and further in third, fourth and fifth distance classes (5, 10, and 20 km, respectively) (**Supplementary Figure S3**). At the genet level, the kinship-distance regression slopes were also negative for the three contiguous watersheds (**Supplementary Table S13**). The average significant values of F_{ij} in genets were found in the two first distance classes (up to 1 km) in Pigeon and Little Pigeon Rivers; in km 1 and 10 in Little TN and Hiwassee Rivers; while in Ocoee and Coosa Rivers, it was significant in the third (5 km) and the fifth class (20 km) (**Supplementary Figure S3**).

DISCUSSION

The evolution of separate sexes (dioicy) has intrigued biologists for centuries. The genetic advantages of separate sexes are

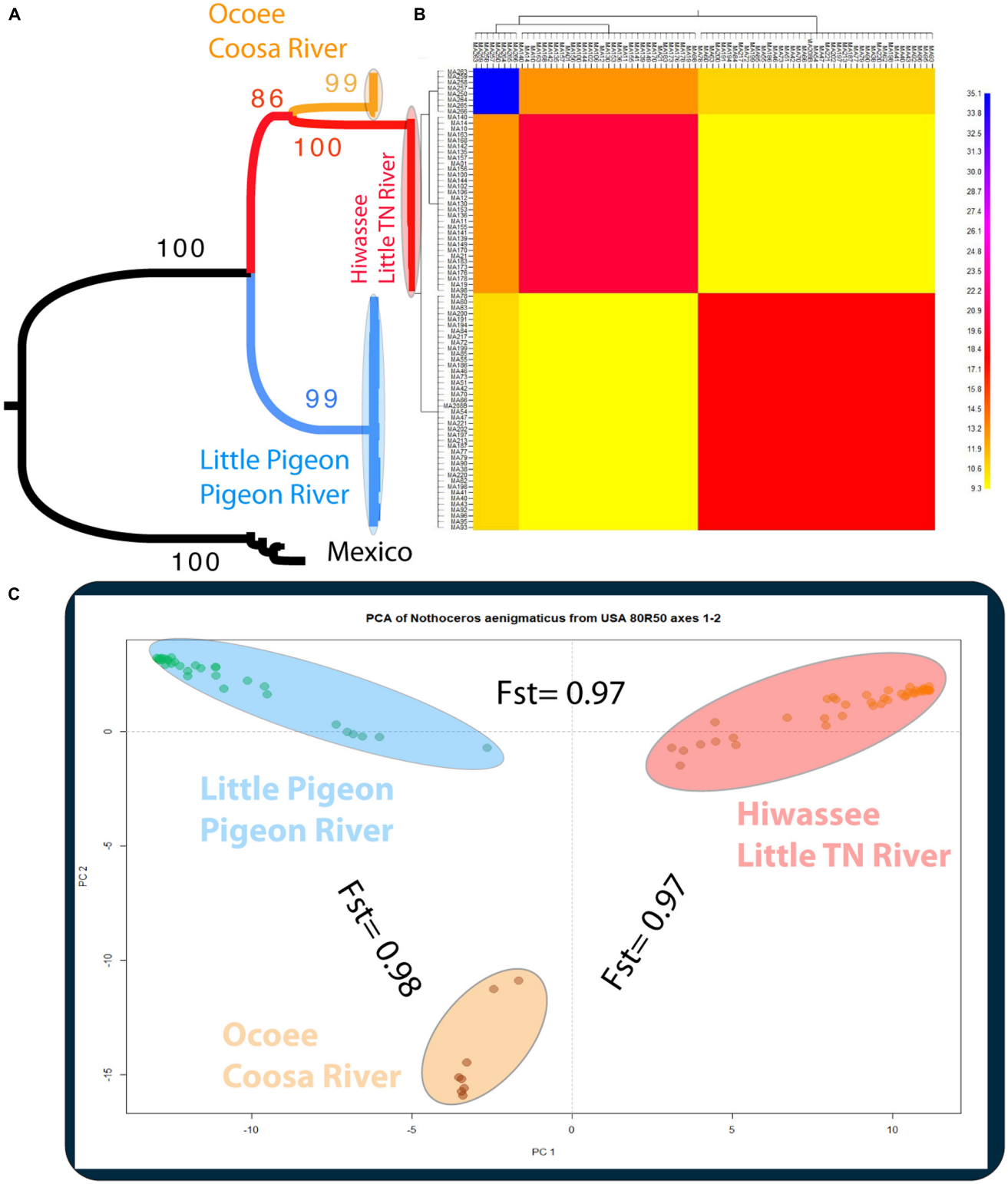


FIGURE 3 | (A) Phylogenetic tree inferred from the Mexican and United States GBS data of *Nothoceros aenigmaticus* (86 samples SNPs matrix). The clade of each genetic group is highlighted. Bootstrap values are indicated at the branches. **(B)** Clustered fineRADstructure co-ancestry matrix of *N. aenigmaticus* from the United States estimated from the unlinked SNPs matrix (80 samples). The heatmap shows pairwise co-ancestry between individuals, with blue representing the highest level, red and orange indicating intermediate levels, and yellow representing the lowest levels of shared co-ancestry. **(C)** Principal Components Analysis (PCA) of 80 samples of *N. aenigmaticus* from the United States (SNPs matrix) indicating Fst values between groups. Colors correspond to those in **Figure 1**.

TABLE 3A | Number of loci and SNPs of *Nothoceros aenigmaticus* for both dataset (exclusively United States and US plus Mexico), before and after filtering with the option populations of the Stacks pipeline *(–max-het-obs 0 –min-maf 0.05 –R 0.5*).

Matrices	Prefilter		Post-filter		
	No. loci	No. SNPs	No. loci	No. SNPs	No. unlinked SNPs
Mexico and US individuals	6,157	1,399	2,402	669	491
US individuals	6,157	823	2,368	410	305

evident (e.g., avoid inbreeding, providing opportunities for outcrossing and recombination enhancing haplotype diversity), however, the costs of dioicy outnumber the theoretical benefits (e.g., Obeso, 2002). This is particularly evident in plants with spatially separated sexes, which in the most extreme cases may lead to a lack of sexual reproduction. In this study, we analyzed the population structure of a haploid plant with extreme sex ratio (either male or female populations), degenerate male gametes and extended clonality. We here show and confirm (1) the divergence times between Mexican and US plants dating back to the Pleistocene, (2) the strictly clonal nature of the US populations and (3) US populations consisting of few clones with extreme sex ratios (0:1) and high genetic structure (Fst values over 0.9 using nuclear markers). The spatial segregation of sexes within the SA is strongly supported by SNP markers and microsatellite data with shared genotypes mostly between contiguous watersheds. The SA drainage system is thought to have been remodeled by geomorphological processes during the Pleistocene, which could have led to a mixing or isolation of alleles in aquatic and rheophilous taxa (Ross, 1971; Mayden, 1988; Bossu et al., 2013) in contiguous watersheds.

Dispersal Origin of US *N. aenigmaticus* and Extreme Sex Ratios

The crown group of alpine neotropical *N. aenigmaticus* dates to 3–4 Mya (**Supplementary Figure S1** and **Table S2**). SA populations of *N. aenigmaticus*, a taxon formerly considered endemic to this region, share a MRCA with Mexican plants, no later than ca. 600–800,000 years (3,200 years –1.7 Mya), suggesting that the US plants experienced the last Ice age. The high differentiation between Mexican (sexual) and US (clonal) plants points to a long and on-going isolation of US populations with highly restricted contemporary migration between these two regions.

The Mexican/SA populations were probably founded following dispersal from sexually reproducing, and hence spore producing, alpine tropical populations. Although the spores do not exhibit the syndrome necessary for effective wind-mediated long distance dispersal (e.g., a thick wall; van Zanten and Gradstein, 1988), they could have been transported by lower altitude winds or even by a biotic vector, such as migratory birds (Lewis et al., 2014) as has been invoked for leafy liverworts that exhibit disjunct distributions and have similar spores with large plastids ("green-looking") with a rather thin exine (van Zanten and Gradstein, 1988). *Nothoceros aenigmaticus* is dioicous and hence the spores are "unisexual." Several scenarios are possible. (1) The occurrence of both sexes in the SA thus requires at least two independent dispersal events: one for male plants (Pigeon/Little Pigeon River) and the other for female plants (Little TN/Hiwassee River, Ocoee/Coosa River) and a third potentially to Balsam Mountain Gap (male plants). (2) It is possible that the current species distribution is product of a co-dispersal of male and female spores and the species was widespread in the SA region after a dispersal from Mexico in the Plio-Pleistocene. (3) Or, local unknown disturbances fragmented and isolated the populations in the US. In this case, we envision a scenario of dispersal from Paramos to Mexico/US, following local vicariance of US watersheds from Mexico.

The spatial separation of sexes (sex ratios 0:1) shown in *N. aenigmaticus* from US suggests that SA populations survived through the Pleistocene by clonal growth. The reasons for a preferential expression, survival and establishment of either sex is not yet clear in bryophytes (Bisang and Hedenäs, 2005; Baughman et al., 2017). Male spore mortality is one of the most commonly suggested factors attributed to the striking female biased ratio in bryophytes (Baughman et al., 2017). Using microsatellite data, we found three potential migrants between the Little TN watershed (female) and Little Pigeon River (male) and one between Zempoala (Mexico) and Coosa River. We cannot dismiss whether these are true migrants or reflect homoplasy in

TABLE 3B | Population statistics for polymorphic loci and for both polymorphic and fixed loci. Filters applied were *–max-het-obs 0 –min-maf 0.05 –r 0.5*.

Watersheds	Variant positions			Variant and fixed positions						
	N	Private	π	*N*	Private	Sites	Variant sites	Polymorphic sites	PPL %	π
Pigeon and Little Pigeon Rivers	26	128	0.0035	26	128	309,062	409	7	0.0023	0.00000
Little TN and Hiwassee Rivers	32	168	0.0061	32	168	298,544	405	10	0.0034	0.00001
Ocoee and Coosa Rivers	7	99	0.0006	7	99	250,941	409	1	0.0004	0.00000

Statistics include the total average number of individuals genotyped at each locus (N), the number of alleles unique to each population (private), mean nucleotide diversity (π), the total number of sites for each populations (sites), the number of sites across all populations that are variant (variant sites), the number of sites inside each population that are variant (polymorphic sites) and the percentage of polymorphic sites (PPL %).

the microsatellite data, but based on SNP data the latter explanation is more likely.

Spatial Segregate Sexes (SSS) and Spatial Genetic Structure (SGS)

The evolution of SSS in dioicous species is intriguing and rather counter-intuitive. A reduced reproductive success seems to be compensated by niche specialization (Bierzychudek and Eckhart, 1988) or high intersexual competition (Rogers and Eppley, 2015). Niche specialization lacks experimental evidence in bryophytes (Bisang et al., 2015; Brzyski et al., 2018) or empirical support in flowering plants (Rogers and Eppley, 2015). In some cases, female plants tend to live in more benign habitats such as moisture slopes in the conifer species *Austrocedrus chilensis* (D. Don) Pic. Serm. & Bizzarri (Nuñez et al., 2008) or high phosphorous substrates in the grass *Distichlis spicata* (L.) Green (Eppley et al., 2009). In the desert moss, *Syntrichia caninervis*, however, males were never found in harsh environments (Stark et al., 2005). No studies have been attempted to measure the micro-environmental conditions or nutrient concentration of the streams where male and female *N. aenigmaticus* grow to assess niche specialization. Regarding intersexual competition, Rogers and Eppley (2015) found evidence for such competition in the grass *Distichlis spicata*, suggesting that competition rather than niche specialization is driving the SSS in the species. Equally, the males from the Pigeon and Little Pigeon rivers are separated from the closest female plants by nearly 30 km (**Figures 1**, **2**). The male population from the Balsam Mountain Preserve occurs ca. 48 km upstream from the closest known female population (Little TN River) remains a puzzle. The strict sex allopatry across watersheds suggests intersexual competition. In addition, the lack of sexual selection, combined with exclusively clonal growth likely explains the loss of the ability to produce viable sperm cells (Renzaglia and McFarland, 1999).

Our expanded sampling confirmed the SSS, especially among distant watersheds. The observed fine-scale SGS indicate that within contiguous watersheds, alleles of *N. aenigmaticus* are genetically closer between nearby individuals than between distant individuals. Gene flow is limited in the species, and it seems to occur mostly between contiguous watersheds. Usually, SGS is the result of limited gene dispersal and can be influenced by a broad array of life-history traits including clonality and selfing. In our case, clonality is a crucial determining factor because clonality increases kinship at a small spatial scale. Therefore, clonality, SGS and SSS may have an impact on the complete absence of sexual reproduction in US *N. aenigmaticus*.

Eckert (2002) summarized the ecological implications of the loss of sex for clonal plants, invoking a "use-it-or-lose-it" hypothesis for degeneration of sexual structures in apomictic plants. Potential sterility mutations may occur in clonal plants that have completely abandoned sexual reproduction and are particularly likely in populations at the margins of their range (Eckert, 2002; Barrett, 2015) and especially in dioicous taxa (Hu et al., 2017). The "use-it-or-lose-it" certainly applies to *N. aenigmaticus* and it is more evident in male gametophytes. In Mexican *N. aenigmaticus*, sexually functional male and female plants grow intermixed. Mexican male plants are small and inconspicuous (less than 4 mm wide) with abundant (>20) antheridial chambers per gametophyte, typical of dioicous hornworts (Proskauer, 1948). By contrast, US male plants have larger thalli, exuberant growth and few antheridial chambers per thallus (up to 6), which might not open and if opened, the antheridia bear abortive sperm cells (Renzaglia and McFarland, 1999). Our observations suggest that male plants are no longer constrained by sexual selection; therefore, they do not have the selective pressure to produce viable sperm cells. We could not determine whether archegonia (in female plants) degenerate and we do not know of a single study capable of assessing a degeneration of the female function in sporic plants. A detailed developmental analysis using expression data is needed to test whether the male and female function are under relaxed selection.

Estimates of divergence times suggests that clones have persisted since the Pleistocene (**Supplementary Figure S1**) and hence that their higher levels of genotypic diversity in plastid and mitochondrial loci, high number of nuclear private alleles (e.g., 168 for Little TN together with Hiwassee river, **Table 3**) may result from a high number of somatic mutations (Schuster, 1992). In contrast to many organisms in the SA (Soltis et al., 2006), a lack of northward expansion of *N. aenigmaticus* may be due for the limited dispersal capabilities (lack of spores) and lack of suitable habitats.

Patterns of Clonal Diversity

Population genetic theory predicts that in asexual taxa, clonal growth may amplify the effect of genetic drift by reducing the effective size of local populations (Chung and Kang, 1996; Jones and Gliddon, 1999) and that genetic polymorphism will decrease over time. Using microsatellite and SNP markers, the best-sampled SA watersheds (Little TN River and Little Pigeon River) have few multilocus genotypes and little diversity (**Figure 2** and **Supplementary Tables S3, S4, S6**), consistent with clonal growth. The high Φ_{PT} and Fst values (**Table 2** and **Supplementary Tables S5, S10**) between non-contiguous populations are consistent with little genotype exchange between them.

The moss *Pleurochaete squarrosa* (Brid.) Lindb. has genetically highly structured populations (Grundmann et al., 2008). Its central European populations are mostly clonal, have a low G/N ratio and are well-differentiated (i.e., Gst values of 0.89) from putatively sexually reproducing Mediterranean populations. The low levels of genotypic diversity in glaciated areas seem to be a widespread pattern across European bryophytes (Kyrkjeeide et al., 2014). In lacustrine bryophytes, flow among populations is moderate (Fst values from 0.14 to 0.280), such that the genetic structure among these populations is likely best explained by clonal growth and episodic spore and diaspore dispersal (Korpelainen et al., 2012). The low values of genotypic diversity and high levels of clonal differentiation among bryophyte populations are in line with studies from dioecious island-inhabiting angiosperms with clonal propagation (Meloni et al., 2013). In bryophytes, high population differentiation (Fst values over 0.9) is rare and typically linked to transoceanic distribution (Patiño and Vanderpoorten, 2018).

Genetic diversity in aquatic plants is expected to increase downstream by increasing gene flow between isolated populations (Nilsson et al., 2010). In addition to the migration bias that would result in higher genetic diversity among downstream populations, the upstream populations may further see their genotypic diversity decline due to downstream drift (i.e., the "drift paradox"; Müller, 1974), unless the regular loss of individuals drifting downstream and resulting in the extinction of genotypes is compensated by some other means of upstream dispersal (Pollux et al., 2007; Nilsson et al., 2010). Most populations of *N. aenigmaticus* are found in the headwaters or medium-size streams, rarely in high volume rivers, suggesting that stranding and successful establishment downstream is rare. One prime example is the lack of male plants downstream in the Little TN River (just ca. 48 km upstream of the first colony of female populations) suggesting a highly limited dispersal capability of *N. aenigmaticus*. Almost every single river is separated by a mountain divide, suggesting that dispersal may be highly limited without having spores. Intercatchment dispersal mainly via water birds has repeatedly been reported in plants, particularly angiosperms (Pollux et al., 2007; Smulders et al., 2008). In the aquatic and sexually reproducing moss *Platyhypnidium riparioides* (Hedw.) Dixon, episodic dispersal of vegetative diaspores and occasional from spores contribute to the structured population dynamic of the species with very little downstream dispersal (Hutsemékers et al., 2003).

One potential explanation, and extremely challenging to test at this scale, is that *N. aenigmaticus* dispersed between catchments following changes in river geomorphology and potential post-glaciation floods (Ross, 1971; Mayden, 1988; Willet et al., 2014). We postulate that in the case of lack of sexual reproduction (and therefore spores) the dispersal via fragments could be limited and only accomplished via Pleistocene river piracy. This process has been established for the Little TN River Basin (Ross, 1971; Prince et al., 2010; Willet et al., 2014), and Ocoee and Coosa Rivers, the latter rivers draining in different watersheds (Mills et al., 2005). Recent modeling studies of the Southern US river systems have confirmed the historical dynamics of the drainages due to constant erosion and drainage changes, including stream captures (Willet et al., 2014). Testing the hypothesis of river piracy in the watersheds in our study is very challenging because of the localized spatial scale of the present study.

CONCLUSION

The SA *N. aenigmaticus* is a recent immigrant from sexual tropical populations with an extreme case of spatial separation of sexes (1:0 ratio). Based on microsatellite and SNP data, some genotypes are shared between contiguous watersheds, and hence the current genetic diversity of this species is not partitioned by watersheds. The data show sex-specific SGS with watersheds displaying a single sex. The presence of only clonal *N. aenigmaticus* in the SA suggests that the species is "locked" in the SA. Given the low extant genetic diversity, which can only be increased by somatic mutations in the absence of sexual reproduction, any slight alteration of the habitat may cause the local extinction of the species. *Nothoceros aenigmaticus* may surge as a model to study the impact of clonality on the deterioration and lack of sexual function (e.g., genes involved in flagellum formation), the proliferation of somatic mutations (in organellar and nuclear genomes) and mutational meltdown in haploid plants.

DATA AVAILABILITY STATEMENT

The sequence data are already published and found in **Supplementary Table S1**. Microsatellite raw data used for the dating and population genetic analyses are available upon request. Raw sequences GBS data are deposited in NCBI BioProject ID PRJNA562471.

AUTHOR CONTRIBUTIONS

MA-G performed and analyzed GBS data, performed population genomic analyses. JV conceived the study, conducted fieldwork, developed and analyzed microsatellite data, performed phylogenetic analyses. KM planned and conducted fieldwork, provided research context, natural history and geographic data on the species. BG conceived and planned the study, provided funding. All authors read and approved the manuscript.

FUNDING

Financial support came from the NSF DDIG-0910258 award to JV and the Ronald Bamford Endowment (UCONN), the UCONN president Fellowship and the NSERC Discovery Grant (CRNSG- RGPIN/05967-2016). MA-G benefited from financial support from "Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia, España" (project 20369/PD/17).

ACKNOWLEDGMENTS

JV thank Karen Renzaglia (SIUC) for insightful discussions on the biology of *Nothoceros aenigmaticus*, Patrick Meirmans, Laura L. Forrest, Pablo Arroyo, O. Pérez, and K. Holsinger for suggestions on population genetic analyses and maps. The Smoky Mountain National Park, Nantahala National Forest, Cherokee National Forest, Panama (A.N.A.M.), Costa Rican (MINAET INBIO), and Dominican Republic authorities provide permits for plant collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00495/full#supplementary-material

REFERENCES

Abed, A., Légaré, G., Pomerleau, S., St-Cyr, J., Boyle, B., and Belzile, F. J. (2019). "Genotyping-by-sequencing on the ion torrent platform in barley," in *Methods in Molecular Biology*, Vol. 1900, ed. W. Harwood (New York, NY: Humana Press).

Barrett, S. C. H. (2015). Influence of clonality on plant sexual reproduction. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8859–8866. doi: 10.1073/pnas.1501712112

Baughman, J. T., Payton, A. C., Paasch, A. E., Fisher, K. M., and McDaniel, S. F. (2017). Multiple factors influence population sex ratios in the Mojave desert moss *Syntrichia caninervis*. *Am. J. Bot.* 104, 733–742. doi: 10.3732/ajb.1700045

Bierzychudek, P., and Eckhart, V. (1988). Spatial segregation of the sexes in dioecious plants. *Am. Nat.* 132, 34–43. doi: 10.1086/284836

Bisang, I., Ehrlén, J., Korpelainen, H., and Hedenäs, L. (2015). No evidence of sexual nice partitioning in a dioecious moss with rare sexual reproduction. *Ann. Bot.* 116, 771–779. doi: 10.1093/aob/mcv133

Bisang, I., and Hedenäs, L. (2005). Sex ratio patterns in dioicous bryophytes re-visited. *J. Bryol.* 27, 207–219. doi: 10.1179/174328205X69959

Bisang, I., Hedenäs, L., and Nils, C. (2017). Can the meiotic sex ratio explain the sex ratio bias in adult populations in the dioicous moss *Drepanocladus lycopodioides*? *J. Bryol.* 39, 115–120. doi: 10.1080/03736687.2016.1264662

Bossu, C. M., Beaulieu, J. M., Ceas, P. A., and Near, T. J. (2013). Explicit tests of palaeodrainage connections of southeastern North America and the historical biogeography of Orangethroat darters (Percidae: *Etheostoma: Ceasia*). *Mol. Ecol.* 22, 5397–5417. doi: 10.1111/mec.12485

Brzyski, J. R., Stieha, C. R., and McLetchie, D. N. (2018). The impact of asexual and sexual reproduction in spatial genetic structure within and between populations of the dioecious plant *Marchantia inflexa* (Marchantiaceae). *Ann. Bot.* 122, 993–1003. doi: 10.1093/aob/mcy106

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3-Genes Genom Genet*. 1, 171–182. doi: 10.1534/g3.111.000240

Chung, M. G., and Kang, S. S. (1996). Allozyme genetic and clonal diversity within populations of *Chimaphila japonica* and *Pyrola japonica* (Pyrolaceae). *Isr. J. Plant Sci.* 44, 259–271. doi: 10.1080/07929978.1996.10676650

Cronberg, N., Natcheva, R., and Hedlund, K. (2006). Microarthropods mediate sperm transfer in mosses. *Science* 313, 1255–1255. doi: 10.1126/science.1128707

Dering, M., Raczka, G., and Szmyt, J. (2016). Sex-specific pattern of spatial genetic structure in dioecious and clonal tree species, *Populus alba* L. *Tree Genet. Genomes* 12:70. doi: 10.1007/s11295-016-1028-5

Dray, S., and Dufour, A. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1–20. doi: 10.18637/jss.v022.i04

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Eckert, C. G. (2002). The loss of sex in clonal plants. *Evol. Ecol.* 15, 501–520. doi: 10.1023/A:101600551

Eppley, S. M., Mercer, C. A., Haaning, C., and Graves, C. B. (2009). Sex-specific variation in the interaction between *Distichlis spicata* (Poaceae) and mycorrhizal fungi. *Am. J. Bot.* 96, 1967–1973. doi: 10.3732/ajb.0900076

Field, D. L., Pickup, M., and Barrett, S. C. H. (2013). Comparative analyses of sex-ratio variation in dioecious flowering plants. *Evolution* 67, 661–672. doi: 10.1111/evo.12001

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Gordon, A., and Hannon, G. J. (2010). *Fastx-Toolkit. Computer Program Distributed by the Author*. Available online at: http://hannonlab.cshl.edu/fastx_toolkit/index.html (accessed October, 2019).

Groen, K. E., Stieha, C. R., Crowley, P. H., and McLetchie, D. N. (2010). Sex-specific plant responses to two light levels in the liverwort *Marchantia inflexa* (Marchantiaceae). *Bryologist* 113, 81–89. doi: 10.1639/0007-2745-113.1.81

Grundmann, M., Ansell, S. W., Russell, S. J., Koch, M. A., and Vogel, J. C. (2008). Hotspots of diversity in a clonal world – the Mediterranean moss *Pleurochaete squarrosa* in central Europe. *Mol. Ecol.* 17, 825–838. doi: 10.1111/j.1365-294X.2007.03634.x

Hardy, O. J., and Vekemans, X. (2002). Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x

Hu, A. Q., Gale, S. W., Kumar, P., Saunders, R. M. K., Sun, M., and Fischer, G. A. (2017). Preponderance of clonality triggers loss of sex in *Bulbophyllum bicolor*, an obligately outcrossing epiphytic orchid. *Mol. Ecol.* 26, 3358–3372. doi: 10.1111/mec.14139

Hutsemékers, V., Hardy, O. J., and Vanderpoorten, A. (2003). Does water facilitate gene flow in spore-producing plants? Insights from the fine-scale genetic structure of the aquatic moss *Rhynchostegium riparioides*. *Aquat. Bot.* 108, 1–6. doi: 10.1016/j.aquabot.2013.02.001

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Jones, B., and Gliddon, C. (1999). Reproductive biology and genetic structure in *Lloydia serotina*. *Plant Ecol.* 141, 151–161.

Korpelainen, H., von Cräutlein, M., Kostamo, M., and Virtanen, N. (2012). Spatial genetic structure of aquatic bryophytes in a connected lake system. *Plant Biol.* 15, 514–521. doi: 10.1111/j.1438-8677.2012.00660.x

Kyrkjeeide, M. O., Stenøien, H. K., Flatberg, K. I., and Hassel, K. (2014). Glacial refugia and post-glacial colonization patterns in European bryophytes. *Lindbergia* 37, 47–59. doi: 10.25227/linbg.01046

Laenen, B., Machac, A., Gradstein, S. R., Shaw, B., Patiño, J., Désamoré, A., et al. (2016). Increased diversification rates follow shifts to bisexuality in liverworts. *New Phytol.* 210, 1121–1129. doi: 10.1111/nph.13835

Lewis, L. R., Behling, E., Gousse, H., Qian, E., Elphick, C., Lamarre, J. F., et al. (2014). First evidence of bryophyte diaspores in the plumage of transequatorial migrant birds. *PeerJ* 2, e424. doi: 10.7717/peerj.424

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Loiselle, B. A., Sork, V. L., Nason, J., and Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am. J. Bot.* 82, 1420–1425. doi: 10.1002/j.1537-2197.1995.tb12679.x

Longton, R. E. (1997). Reproductive biology and life-history strategies. *Adv. Bryol.* 6, 65–101.

Malinsky, M., Trucchi, E., Lawson, D. J., and Falush, D. (2018). RADpainter and fineRADstructure: population inference from RADseq data. *Mol. Biol. Evol.* 35, 1284–1290. doi: 10.1093/molbev/msy023

Mayden, R. L. (1988). Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. *Syst. Zool.* 37, 329–355. doi: 10.1093/sysbio/37.4.329

Meirmans, P. G., and Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* 4, 792–794. doi: 10.1111/j.1471-8286.2004.00770.x

Meloni, M., Reid, A., Caujapé-Castells, J., Marrero, Á, Fernández-Palacios, J. M., Mesa-Coelo, R. A., et al. (2013). Effects of clonality on the genetic variability of rare, insular species: the case of *Ruta microcarpa* from the Canary islands. *Ecol. Evol.* 3, 1569–1579. doi: 10.1002/ece3.571

Mills, H. H., Sumners, D. N., Hart, E. A., and Li, P. (2005). Distribution of high-level alluvial deposits in the Valley and Ridge of Polk County, southeastern Tennessee: implications for river history and drainage evolution. *Southeastern Geol.* 44, 37–44.

Müller, K. (1974). Stream drift as a chronobiological phenomenon in running water ecosystems. *Annu. Rev. Ecol. Evol. Syst.* 5, 309–323. doi: 10.1146/annurev.es.05.110174.001521

Nazareno, A. G., Alzate-Marin, A. L., and Pereira, R. A. S. (2013). Dioicy, more than monoecy, affects plant spatial genetic structure: the case study of *Ficus*. *Ecol. Evol.* 3, 3495–3508. doi: 10.1002/ece3.739

Nilsson, C., Brown, R. L., Jansson, R., and Merritt, D. M. (2010). The role of hydrochory in structuring riparian and wetland vegetation. *Biol. Rev.* 85, 837–858. doi: 10.1111/j.1469-185X.2010.00129.x

Nuñez, C. I, Nuñez, M. A., and Kitzberger, T. (2008). Sex-related spatial segregation and growth in a dioecious conifer along environmental gradients in northwestern Patagonia. *Ecoscience* 15, 73–80. doi: 10.2980/1195-6860(2008)15[73:sssagi]2.0.co;2

Obeso, J. R. (2002). The costs of reproduction in plants. *New Phytol.* 155, 321–348. doi: 10.1046/j.1469-8137.2002.00477.x

Patiño, J., and Vanderpoorten, A. (2018). Bryophytes biogeography. *Crit. Rev. Plant Sci.* 37, 175–209. doi: 10.1080/07352689.2018.1482444

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*. 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x

Petry, W. K., Soule, J. D., Iler, A. M., Chicas-Mosier, A., Inouye, D. W., Miller, T. E. X., et al. (2016). Sex-specific responses to climate change in plants alter population sex ratio and performance. *Science* 353, 69–71. doi: 10.1126/science.aaf2588

Pollux, B. J. A., Jong, M. D. E., Steegh, A., Verbruggen, E., van Groenendael, J. M., and Ouborg, N. J. (2007). Reproductive strategy, clonal structure and genetic diversity in populations of the aquatic macrophyte *Sparganium emersum* in river systems. *Mol. Ecol.* 16, 313–325. doi: 10.1111/j.1365-294X.2006.03146.x

Prince, P. S., Spotila, J. A., and Henika, W. S. (2010). New physical evidence of the role of stream capture in active retreat of the Blue Ridge escarpment, southern Appalachians. *Geomorphology* 123, 305–319. doi: 10.1016/j.geomorph.2010.07.023

Proskauer, J. (1948). Studies on the morphology of *Anthoceros*. I. *Ann. Bot.* 12, 237–265. doi: 10.1093/oxfordjournals.aob.a083187

Rambaut, A. (2002). *Se-Al v2.0a11: Sequence Alignment Editor*. Available online at: http://tree.bio.ed.ac.uk/software/seal/ (accessed October, 2019).

Rambaut, A. (2015). *Figtree: A Graphical Viewer of Phylogenetic Trees*. Available online at: http://tree.bio.ed.ac.uk/software/ (accessed October, 2019).

Rambaut, A., and Drummond, A. J. (2007). *Tracer v. 1.5*. Available online at: http://beast.bio.ed.ac.uk/Tracer (accessed October, 2019).

Rambaut, A., and Drummond, A. J. (2010). *TreeAnnotator version 1.6.1*. Available online at: http://beast.bio.ed.ac.uk (accessed October, 2019).

Renner, S. S., and Ricklefs, R. E. (1995). Dioicy and its correlates in the flowering plants. *Am. J. Bot.* 82, 595–606.

Renzaglia, K. S., and McFarland, K. D. (1999). Antheridial plants of *Megaceros aenigmaticus* in the Southern Appalachians: anatomy, ultrastructure and population distribution. *Haussknechtia Beiheft* 9, 307–316.

Rochette, N. C., and Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* 12, 2640–2659. doi: 10.1038/nprot.2017.123

Rogers, S. R., and Eppley, S. M. (2015). Testing the interaction between intersexual competition and phosphorous availability in a dioecious grass. *Botany* 90, 704–710. doi: 10.1139/b2012-042

Ross, R. D. (1971). "The drainage history of the Tennessee river," in *The Distributional History of the Biota of the Southern Appalachians, Part III, Vertebrates*, ed. P. C. Holt (Blacksburg, VA: Virginia Polytechnic Institute), 11–42.

Schuster, R. M. (1983). "Phytogeography of bryophytes," in *New Manual of Bryology*, Vol. 1, ed. R. M. Schuster (Nichinan: The Hattori Botanical Laboratory), 463–626.

Schuster, R. M. (1992). On *Megaceros aenigmaticus*. *Bryologist* 95, 305–315.

Scott, M. F., Osmond, M. M., and Otto, S. P. (2018). Haploid selection, sex ratio bias, and transitions between sex-determining systems. *PLoS One* 16:e2005609. doi: 10.1371/journal.pbio.2005609

Shortlidge, E. N., Rosenstiel, T. N., and Eppley, S. M. (2012). Tolerance to environmental desiccation in moss sperm. *New Phytol.* 194, 741–750. doi: 10.1111/j.1469-8137.2012.04106.x

Sinclair, J. P., Emlen, J., and Freeman, D. C. (2012). Biased sex ratios in plants: theory and trends. *Bot. Rev.* 78, 63–86. doi: 10.1007/s12229-011-9065-0

Smulders, M. J. M., Cottrell, J. E., Lefèvre, F., van der Schoot, J., Arens, P., Vosman, B., et al. (2008). Structure of the genetic diversity in Black poplar (*Populus nigra* L.) populations across European river systems: consequences for conservation and restoration. *Forest Ecol. Manag.* 255, 1388–1399. doi: 10.1016/j.foreco.2007.10.063

Soltis, D. E., Morris, A. B., McLachian, J. S., Manos, P. S., and Soltis, P. S. (2006). Comparative phylogeography of unglaciated Eastern North America. *Mol. Ecol.* 15, 4261–4293. doi: 10.1111/j.1365-294x.2006.03061.x

Stamatakis, A. (2014). RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stark, L. R., McLetchie, D. N., and Mishler, B. D. (2005). Sex expression, plant size, and spatial segregation of the sexes across a stress gradient in the desert moss *Syntrichia caninervis*. *Bryologist* 108, 183–193. doi: 10.1639/0007-2745(2005)108%5B0183:sepsas%5D2.0.co;2

Van der Velde, M., During, H. J., van de Zande, L., and Bijlsma, R. (2001). The reproductive biology of *Polytrichum formosum*: clonal structure and paternity revealed by microsatellites. *Mol. Ecol.* 10, 2423–2434. doi: 10.1046/j.0962-1083.2001.01385.x

van Zanten, B. O., and Gradstein, S. R. (1988). Experimental dispersal geography of neotropical liverworts. *Beihefte Nova Hedwigia* 90, 41–94.

Villarreal, J. C., Campos, L., Uribe, J., and Goffinet, B. (2012a). Parallel evolution of endospory in hornworts: *Nothoceros renzagliensis*, sp. nov. *Syst. Bot.* 37, 31–37. doi: 10.1600/036364412X616594

Villarreal, J. C., Cusimano, N., and Renner, S. S. (2015). Biogeography and diversification rates in hornworts – the limitations of diversification modeling. *Taxon* 64, 229–238. doi: 10.12705/642.7

Villarreal, J. C., Forrest, L., McFarland, K., and Goffinet, B. (2012b). Chloroplast, mitochondrial and nuclear microsatellites from the Southern Appalachian hornwort, *Nothoceros aenigmaticus*. *Am. J. Bot.* 99, e88–e90. doi: 10.3732/ajb.1100392

Villarreal, J. C., Forrest, L. L., Cooper, E. D., and Long, D. G. (2014). Phylogenetic affinities and conservation status of the liverwort *Telaranea murphyae* in Britain. *J. Bryol.* 36, 191–199. doi: 10.1179/1743282014Y.0000000106

Villarreal, J. C., and Renner, S. S. (2013). Correlates of monoicy and dioicy in hornworts, the apparent sister group to vascular plants. *BMC Evol. Biol.* 13:239. doi: 10.1186/1471-2148-13-239

Villarreal, J. C., and Renner, S. S. (2014). A review of molecular-clock calibrations and substitution rates in liverworts, mosses, and hornworts, and a timeframe for a taxonomically cleaned-up genus *Nothoceros*. *Mol. Phylogenet. Evol.* 78, 25–35. doi: 10.1016/j.ympev.2014.04.014

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868. doi: 10.1073/pnas.1323926111

Willet, S. D., McCoy, S. W., Perron, J. T., Goren, L., and Chen, C. Y. (2014). Dynamic reorganization of river basins. *Science* 343:1248765. doi: 10.1126/science.1248765

Wyatt, R., and Anderson, L. E. (1984). "Breeding systems in bryophytes," in *The Experimental Biology of Bryophytes*, eds A. F. Dyer, and J. G. Duckett (London: Academic press), 39–64.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

METHODS
published: 28 May 2020
doi: 10.3389/fpls.2020.00694

A Cost Reduced Variant of Epi-Genotyping by Sequencing for Studying DNA Methylation in Non-model Organisms

Olaf Werner[1], Ángela S. Prudencio[2], Elena de la Cruz-Martínez[1], Marta Nieto-Lugilde[1], Pedro Martínez-Gómez[2] and Rosa M. Ros[1]*

[1] Laboratory of Molecular Systematics, Phylogeography and Conservation in Bryophytes, Department of Plant Biology, Faculty of Biology, University of Murcia, Murcia, Spain, [2] Laboratory of Fruit Tree Breeding, Department of Plant Breeding, CEBAS-CSIC, Murcia, Spain

Reference-free reduced representation bisulfite sequencing uses enzymatic digestion for reducing genome complexity and allows detection of markers to study DNA methylation of a high number of individuals in natural populations of non-model organisms. Current methods like epiGBS enquire the use of a higher number of methylated DNA oligos with a significant cost (especially for small labs and first pilot studies). In this paper, we present a modification of this epiGBS protocol that requires the use of only one hemimethylated P2 (common) adapter, which is combined with unmethylated barcoded adapters. The unmethylated cytosines of one chain of the barcoded adapter are replaced by methylated cytosines using nick translation with methylated cytosines in dNTP solution. The basic version of our technique uses only one restriction enzyme, and as a result, genomic fragments are integrated into two orientations with respect to the adapter sequences. Comparing the sequences of two chain orientations makes it possible to reconstruct the original sequence before bisulfite treatment with the help of standard software and newly developed software written in C and described here. We provide a proof of concept via data obtained from almond (*Prunus dulcis*). Example data and a detailed description of the complete software pipeline starting from the raw reads up until the final differentially methylated cytosines are given in **Supplementary Material** making this technique accessible to non-expert computer users. The adapter design showed in this paper should allow the use of a two restriction enzyme approach with minor changes in software parameters.

Keywords: DNA methylation, epi genotyping by sequencing, population genetics, reduced representation bisulfite sequencing, non-model organisms, *Prunus dulcis*

OPEN ACCESS

Edited by:
Stefan A. Rensing, University of Marburg, Germany

Reviewed by:
Luis Valledor, University of Oviedo, Spain
Conchita Alonso, Estación Biológica de Doñana (EBD), Spain

****Correspondence:***
Olaf Werner
werner@um.es

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *09 January 2020*
Accepted: *01 May 2020*
Published: *28 May 2020*

Citation:
Werner O, Prudencio ÁS, de la Cruz-Martínez E, Nieto-Lugilde M, Martínez-Gómez P and Ros RM (2020) A Cost Reduced Variant of Epi-Genotyping by Sequencing for Studying DNA Methylation in Non-model Organisms. Front. Plant Sci. 11:694. doi: 10.3389/fpls.2020.00694

INTRODUCTION

Contemporary understanding of epigenetics encompasses "the study of changes in gene function that are heritable and that do not entail a change in DNA sequence" (Wu and Morris, 2001). These changes comprise histone variants, posttranslational modifications of amino acids on the amino-terminal tail of histones, and covalent modifications of DNA bases

(Dupont et al., 2009). Most research on epigenetics focuses on DNA methylation, because the covalent changes in DNA bases are relatively easy to investigate with modern sequencing technologies. As a result, the term "epigenetics" is sometimes used to refer exclusively to DNA methylation (Seymour and Becker, 2017).

DNA methylation is under genetic control via a complex network of DNA methyltransferase and DNA glycosylase genes (reviewed in Pikaard and Scheid, 2014), although the extent to which epigenetic variation is under direct genetic control is not clear at this moment (Richards et al., 2017). Epigenetic variation can be the result of ordinary developmental processes that are triggered by internal signals (constitutive), such as those that occur during seed development or fruit ripening (reviewed in Li et al., 2018), or that are the result of external factors (facultative) like biotic or abiotic stress (reviewed in Bräutigam and Cronk, 2018). Additionally, spontaneous epimutations occur and change the DNA methylation pattern in unpredictable ways (reviewed by Richards et al., 2017; Johannes and Schmitz, 2019). Changes in the methylation pattern can be associated with gene expression levels. Generally, DNA methylation is linked to gene silencing, which is especially important in the control of the activity of transposable elements (reviewed by Hosaka and Kakutani, 2018). While the methylation of transposable elements, promotors and transcriptional start sites results in lower gene activity, gene body methylation is typical for housekeeping genes, which are expressed constitutively (Zemach et al., 2010; Bewick and Schmitz, 2017). The role of gene body methylation is not clear, although its conservation in plant evolution (at least 400 Myr) (Zilberman, 2017) and its apparently universal occurrence in the animal kingdom (Zemach and Zilberman, 2010) suggest its relevance (Bräutigam and Cronk, 2018).

In some cases, changes in methylation can be directly linked to distinct phenotypes. A naturally occurring form of *Linaria vulgaris* Mill. with radial flower symmetry instead of the bilateral symmetry of the wild type is characterized by an extensively methylated Lcyc gene, which is transcriptionally inactive; the demethylation of this gene activates the gene leading to the wild-type phenotype (Cubas et al., 1999). Other phenotypes that could be directly related to the methylation state of epigenetic alleles are the late flowering phenotype of fwa mutants in *A. thaliana* (L.) Heynh (Soppe et al., 2000); inhibited tomato fruit ripening (Manning et al., 2006); and sex determination in melon (Martin et al., 2009).

In natural plant populations DNA methylation is highly variable in different species (Richards et al., 2017). However, the rate and evolutionary significance of epimutations in these natural populations is at present largely unknown (Richards et al., 2017). On the other hand, there are several studies that document a correlation of epigenetic marks and environmental factors. For example, Herrera et al. (2016) concluded that in *Helleborus foetidus* L. the epigenetic spatial structure is driven by a moderate to high heritability and responsiveness to local environments. In addition, Alvarez et al. (2019) found differential methylation in oil-exposed and unexposed populations of *Spartina alterniflora* Loisel.

One elegant way to study differences in DNA methylation between samples is based on bisulfite sequencing. This technique takes advantage of the fact that sodium bisulfite causes the deamination of cytosines, unless they are protected by methylation (Frommer et al., 1992). This results in an uracil residue, which is later converted into thymine by a PCR reaction using a compatible polymerase. Sites where a thymine is identified after a bisulfite treatment, but a cytosine is found in the untreated reference indicate an unmethylated cytosine, while sites with a cytosine in the bisulfite-treated DNA indicate a methylated cytosine. This method was first applied to individual genes, but with the introduction of Next Generation Sequencing (NGS) platforms, scientists became aware of the possibility of obtaining the methylation pattern of all cytosines of a given genome (e.g., Cokus et al., 2008; Lister et al., 2008). While whole genome bisulfite sequencing (WGBS) has many advantages when studying model organisms with a known genome sequence, it cannot be applied to non-model organisms without considerable effort to create a *de novo* whole genome sequence. Even if a high-quality reference genome is available, in the case of experimental designs that require a large amount of samples like those encountered frequently in ecological research, the cost of WGBS can reach amounts that are prohibitive, especially in species with a medium to large genome size (Paun et al., 2019). When trying to obtain a genome scan in the search for differential methylation in natural populations of non-model organisms, researchers therefore used other techniques based on the fact that there are isoschizomer pairs of restriction enzymes, one methylation-sensitive and the other methylation-insensitive using a variant of AFLP (MS-AFLP; McClelland et al., 1994).

Several NGS-based protocols like RADseq (Baird et al., 2008), GBS (Elshire et al., 2011) and derived versions [e.g., double digest RADseq (ddRADseq), Peterson et al., 2012] are used to reduce the complexity of genomes by using restriction enzymes in order to obtain well-defined fragments. Barcoded adaptors make it possible to mix many specimens after the initial restriction-ligation steps into one sequencing lane, drastically reducing the cost per sample. The software pipelines developed to be used with this type of data like Stacks (Catchen et al., 2013) make it possible to work with species with known genome sequences but also with non-model taxa with no reference genome. Researchers interested in the bisulfite sequencing of non-model species became aware of the possibility of adapting the RADseq and GBS protocols in order to obtain the methylation data of reduced genome libraries in the absence of a reference genome. As a result of these efforts, three reduced representation bisulfite sequencing (RRBS) protocols were presented in 2016: epiRADseq (Schield et al., 2016), bsRADseq (Trucchi et al., 2016), and epiGBS (van Gurp et al., 2016). EpiRADseq uses a methylation-sensitive restriction enzyme (*Hpa*II, recognition site C↓CGG) together with an unsensitive restriction enzyme (*Pst*I, recognition site CTGCA↓G). Methylated *Hpa*II recognition sites are not cleaved and the corresponding fragments are absent from the resulting RRBS genomic library. The lab procedure follows essentially the standard ddRADseq protocol of Peterson et al. (2012) and only the computational analysis is adapted. While epiRADseq is not more expensive than ddRADseq, the disadvantage of this

method is the fact that it only gives information about the methylation state of the *Hpa*II cut site, but not of the cytosines of the remainder sequenced fragments. The remaining two methods gain this information, but require the use of methylated adapters, which are much more expensive than unmethylated adapters. In the protocol of Trucchi et al. (2016), adapters are fully methylated. In the case of a project with 96 samples prepared in a library to be sequenced on one Illumina lane, 40 oligos with a total of approximately 436 methylated cytosines (depending on the barcode sequences) are needed. Additionally, in the absence of a reference genome, the protocol requires the sequencing of an aliquot of the library prior to the bisulfite treatment in order to build a "reference genome" with the aid of standard RADseq markers.

Although in their original epiGBS publication van Gurp et al. (2016) described the use of fully methylated adapters, hemimethylated adapters can be used (van Moorsel et al., 2019). In this case, the adapter strand whose 3'-end is ligated to the 5'-end of the genomic fragment is methylated while the protocol includes a nick translation step, which is used to repair the nicks between the 3'-end of the genomic fragments and the 5'-end of the unphosphorylated adapter sequences. The dNTP mix contains 5m-cytosine, which is used by the DNA polymerase I as an alternative substrate. As a result of the nick translation, the nick is repaired and the unmethylated cytosines in the adapter strands that are not ligated to the genomic DNA are replaced by 5 m-cytosine. But even so, 20 oligonucleotides with approximately 15 5-mC positions each (depending on the barcode sequence) are still needed, and the cost of the adapters can be higher than the Illumina sequencing of a paired-end library.

Protocols like GBS, RADseq, epiGBS, and bsRADseq use custom-made adapters instead of standard adapters supplied with kits. This is due to several restrictions given the specific conditions of these experiments. One major problem is that all these methods work with restriction enzymes and not randomly sheared DNA. As a result, all sequences start with identical base calls. But an equal per cycle composition of the first forward read bases is important in order to prevent phasing and pre-phasing detection errors (Kircher et al., 2011). In order to filter out PCR duplicates, adapters may be designed to integrate wobble positions. In the case of epiGBS it is convenient to introduce an unmethylated cytosine that can be used to calculate bisulfite conversion rates. But on the contrary, barcode indices must be methylated for epiGBS and similar protocols.

Here we present a variation of the original epiGBS protocol that uses unmethylated standard P1 GBS adapters presented by Poland et al. (2012) and requires only one hemimethylated P2 (common) adapter. This is a highly economical solution if standard GBS adapters are already available in a lab. If no standard GBS adapters are available, it is also possible to combine a high number of barcoded unmethylated P1 adapters with a low number of barcoded hemimethylated P2 adapters.

We describe the necessary software tools – a combination of existing programs like Stacks (Catchen et al., 2013) or USEARCH (Edgar, 2010) and newly designed software for reconstructing the original sequence of the bisulfite-treated fragments. The reconstructed fragments are then joined into a mock genome. The mock genome can then be used with standard software tools like Bismark (Krueger and Andrews, 2011) and methylKit (Akalin et al., 2012) in order to extract methylation information and identify differentially methylated cytosines. **Figure 1** explains the rationale behind our method. Detailed instructions on the use of the new programs together with preexisting software are given in the supplementary attached document. The instructions are presented in a way that is accessible for non-expert users with short shell-scripts that can easily be adapted to the specific conditions of different projects. Precompiled versions of the newly written programs (Linux operating system) and example data files are available for download. The instructions include detailed comments on the use of the different components of the software pipeline and how to change parameters if interested scientists want to use adapters different from those shown here or if other sequencing parameters (for example changed barcodes or enzymes) and/or read lengths are used.

MATERIALS AND METHODS

Plant Material and DNA Extraction

We analyzed DNA from two almond [*Prunus dulcis* (Mill.) D.A. Webb] cultivars (cv. "Desmayo Largeta" and cv. "Penta") at early and late stages during dormancy release (Prudencio et al., 2018). The DNA was extracted from a pool of 10 flower buds according to the protocol of Doyle and Doyle (1987). We performed the DNA extractions independently in two consecutive years. The DNA concentrations of the samples were measured in a Qubit 2 fluorometer and then adjusted to 20 ng/μl. The DNA extractions were stored at −80°C until use.

Adapter Design

The design of the adapters is the essential difference of our protocol in comparison with other variants of epiGBS. The sequences of the barcoded P1 adapters correspond to standard GBS adapters and were taken from Poland et al. (2012). Their sequences are given in **Table 1**. The P1 adapters are completely unmethylated. The P2 adapter (see **Table 2**) was designed for this study. The upper strand is completely unmethylated. The P2 adapter carries five wobble positions, which can be used to eliminate PCR clones (Kebschull and Zador, 2015). If the raw data show an abnormally high number of duplicates, they can be filtered out with the help of the clone_filter module of Stacks, for example (Rochette and Catchen, 2017). Additionally, there is a 5 bp stretch that can be replaced by a barcode if necessary. All cytosines of the bottom strand of the P2 adapter are methylated with the exception of the cytosine in the *Pst*I overlap. As a result of the bisulfite treatment and the final PCR amplification, this position should be converted to thymine. The efficiency of the bisulfite treatment can be calculated as the number of converted cytosines/total number of cytosines at this position. If using other enzymes and/or adapters, an unmethylated cytosine should be integrated in the bottom P2 adapter strand outside the barcode region to guaranty the possibility to calculate the cytosine conversion rate. Both P1 and P2 adapters are designed to avoid the reconstitution of the restriction enzyme cut site

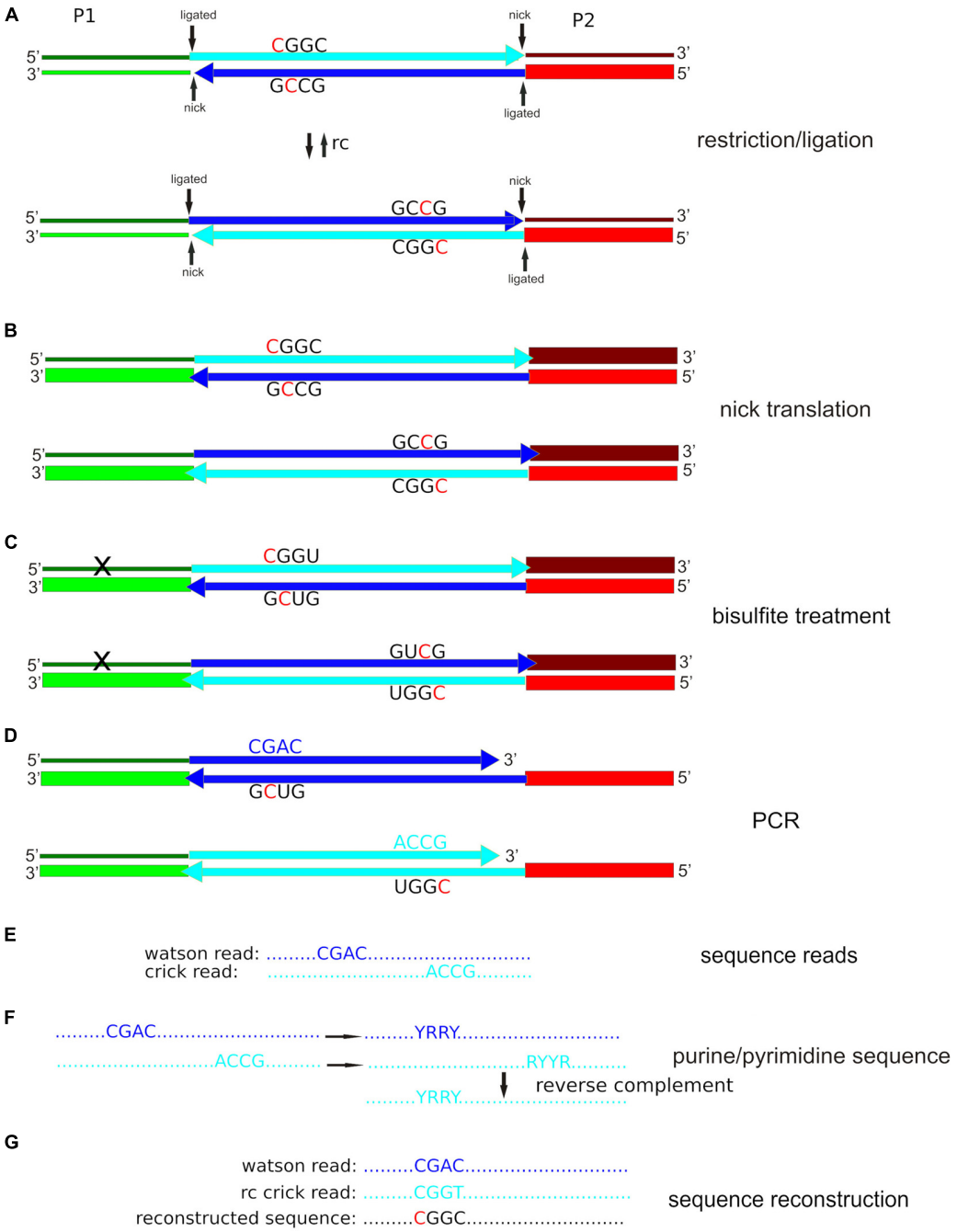


FIGURE 1 | Simplified epiGBS scheme using our protocol with *PstI* as an example. **(A)** Adapter P1 is a barcoded standard GBS adapter; adapter P2 is hemimethylated (only the lower strand has 5-methyl cytosine incorporated; indicated by a thick line). Both adapters are unphosphorylated at the 5′-termini. As a result, after the ligation reaction two nicks remain. The genomic DNA fragment is incorporated into two different orientations with respect to the adapter sequences, which are the reverse complement (rc) of each other. **(B)** After nick translation, the top chain of adapter P1 keeps unmethylated cytosines (thin line). The adapter sequences of the bottom chains are completely methylated (thick lines). **(C)** The bisulfite treatment converts cytosine to uracil unless the cytosines are protected by methylation. The top chain of adapter P1 contains a high number of converted unmethylated cytosines. **(D)** During the PCR step, uracil is read as thymine by a specially engineered polymerase. **(E)** Illumina sequence reads correspond to the complement of the bottom chain. **(F)** The software codifies DNA bases as either purines (R) or pyrimidines (Y). The program takes one arbitrarily defined Watson purine/pyrimidine sequence and tries to find the corresponding Crick sequence with an identical reverse complement purine/pyrimidine sequence. **(G)** If the software finds a Watson/Crick sequence pair, it compares the original Watson sequence with the reverse complement of the original Crick sequence. A cytosine in one sequence and a thymine in the other sequence indicate that there was an unmethylated cytosine in the original sequence. Two cytosines indicate a methylated cytosine in the original sequence, a guanine, and an adenine indicate a guanine with an unmethylated cytosine in the opposite strand in the original sequence and two guanines indicate a guanine with a methylated cytosine in the opposite strand of the original sequence.

TABLE 1 | Sample identification, barcode, and adapter sequences for the top and bottom strand of the barcoded P1 adapters.

Sample	Barcode	Adapter sequence top 5′ ->3′	Adapter sequence bottom 5′ ->3′
AlDA1	CATCTGCCG	cacgacgctcttccgatctCATCTGCCGtgca	CGGCAGATGagatcggaagagcgtcgtg
AlDA2	GGACAG	cacgacgctcttccgatctGGACAGtgca	CTGTCCagatcggaagagcgtcgtg
AlDB1	ATCTGT	cacgacgctcttccgatctATCTGTtgca	ACAGATagatcggaagagcgtcgtg
AlDB2	AAGACGCT	cacgacgctcttccgatctAAGACGCTtgca	AGCGTCTTagatcggaagagcgtcgtg
AlPA1	GAATGCAATA	cacgacgctcttccgatctGAATGCAATAtgca	TATTGCATTCagatcggaagagcgtcgtg
AlPA2	TAGCAG	cacgacgctcttccgatctTAGCAGtgca	CTGCTAagatcggaagagcgtcgtg
AlPB1	ATCCG	cacgacgctcttccgatctATCCGtgca	CGGATagatcggaagagcgtcgtg
AlPB2	CTTAG	cacgacgctcttccgatctCTTAGtgca	CTAAGagatcggaagagcgtcgtg

The sample code consists of the following elements: Al, Almond; Prunus dulcis; "D" or "P," cultivar "Desmayo Largueta" or "Penta"; "A" or "B," early or late in flower bud dormancy breaking, respectively; and "1" or "2," year one or two of the experiment. The barcodes and adapter sequences are taken from Poland et al. (2012). The barcode part of the adapter sequences is given in upper case letters. Sample identification, barcode, and adapter sequences for the top and bottom strand of the barcoded P1 adapters.

TABLE 2 | Sequences of the P2 (common) adapter.

Adapter	Sequence 5′->3′
cre-epiGBS P2 top strand	CAGTTHHHHHagatcggaagagcggttcagcaggaatgccgag
cre-epiGBS P2 bottom strand	t5gg5att55tg5tgaa55g5t5tt55gat5tDDDDDAA5TGTG**C**A

The top strand sequence does not contain methylated cytosines. In the bottom strand, all cytosines with the exception of the last one (corresponding to the PstI overhang) are methylated (given as "5" instead of "C"). The unmethylated cytosine near the end of the bottom strand can be used to calculate the conversion rate achieved with the bisulfite treatment. Lower case letters indicate sequence parts corresponding to Illumina specifications. HHHHH and DDDDD are wobble positions that make it possible to filter out sequencing PCR replicates. The first five bases of the top strand and the AA5TG stretch in the lower strand can be replaced by a barcode sequence if combinatorial barcodes are used. Underlined parts belong to the enzyme-specific recognition site. In the case of PstI, the first base (in italics, here C in the top strand and G in the bottom strand) should not be set to G in the top strand and C in the bottom strand in order to avoid the reconstitution of the PstI cut site. The 3′ most C (in bold) of the bottom strand is unmethylated. This C should be converted into T after the bisulfite treatment and amplification with Kapa Uracil + polymerase. The bisulfite conversion rate can be calculated based on this position.

after the ligation of the genomic DNA fragment to the adapters. The adapter sequences can be changed without any problems to adjust them to other enzymes or to implement specific desired characteristics like wobble bases in the P1 region or other barcodes. **Supplementary Figure 1** shows a general scheme for this purpose. Special care should be taken when designing new barcodes. There are several points that require attention and the use of a GBS barcode generator like the GBSX barcode generator (Herten et al., 2015) is advisable. Recommendations on the design of new P1 and P2 adapters are given in **Supplementary Figure 1**.

Restriction, Ligation, Bisulfite Treatment, and PCR Amplification

All these steps essentially followed the protocol previously described by van Gurp et al. (2016). Boquete et al. (2020) give a detailed description of the protocol with many useful hints for scientist aiming to implement these methods. The first step consists in the restriction of the genomic DNA and adapter ligation (**Figure 1A**). The important difference is that in our protocol the genomic fragments are integrated necessarily in two orientations with respect to the P1 and P2 adapters (**Figure 1A**) whereas in van Gurp et al. (2016) this is not the case. For this step 10 units of *Pst*I-HF (NEB, Ipswich, MA, United States) were added to cut 200 ng (20 µl) of genomic DNA in a 30 µl final volume in 1× CutSmart buffer. Reactions took place overnight at 37°C. The next morning, a mix of 1 µl barcoded P1 adapter (1 mM), 1 µl P2 adapter (1 mM), 1 µl T4 DNA Ligase (NEB, Ipswich, MA, United States; 400 units/µl), 0.4 µl ATP (Thermo Scientific, Alcobendas, Madrid, Spain; 100 mM), 1 µl of CutSmart buffer and 5.6 µl of water were added to each sample to reach a volume of 40 µl. The samples were then incubated for an additional 3 h at 22°C. After adapter ligation, the DNA samples were pooled, which was followed by a cleanup and concentrating step with the help of GeneJet Gel Extraction and DNA Cleanup columns (Thermo Fisher Scientific, Alcobendas, Madrid, Spain). The final elution volume was adjusted to 23 µl.

Because adapters are not phosphorylated, a nick remains between the 3' terminus of the genomic fragment and the 5' terminus of the adapters (**Figure 1B**). This nick is closed with the help of DNA polymerase I. Due to the 5'-3' exonuclease activity of DNA polymerase I the nick repair not only closes the nick between the 3' terminus of the genomic fragment and the unphosphorylated 5' terminus of the adapter, but the complete adapter strand is replaced (van Gurp et al., 2016). This fact is used by the improved version of the epiGBS protocol (van Moorsel et al., 2019) to incorporate 5 methyl-cytosine into the adapter. To this aim, 19.25 µl of the cleaned digestion/ligation mix were incubated for one h at 15°C with 2.5 µl 5-mC-dNTP mix (10 mM, Zymo Research, Irvine, CA, United States), 2.5 µl NEB buffer 2 and 0.75 µl of DNA polymerase I (NEB, Ipswich, MA, United States; 10 units/µl). As a result of this step, three of the four adapter strands are methylated (**Figure 1B**).

The nick translation is followed by the bisulfite treatment (**Figure 1C**). We used the EZ DNA Methylation-Lightning Kit (Zymo Research, Irvine, CA, United States) following the protocol provided with the kit. At the end of the treatment, DNA was eluted in a volume of 10 µl and used directly for PCR amplification. At the end of this step, all unmethylated cytosines are converted to uracil. It is important to note that this is the case of the adapter sequence that was ordered unmethylated

and not replaced in the course of the nick translation (upper left in **Figure 1A**).

The next step is the PCR amplification (**Figure 1D**). Four independent reactions of 25 µl each were set up. Each reaction included 2 µl of template DNA, 12.5 µl of Kapa HiFi HotStart Uracil + ReadyMix (Kapa Biosystems, Wilmington, MA, United States), 1 µl of Illumina PE-PCR primer 1 (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT-3′; 10 µM), 1 µl of Illumina PE-PCR primer 2 (5′-CAAGCAGAAGACGGCATACGAGATC GGTCTCGGCATTCCTGCTGAA-3′; 10 µM) and 8.5 µl H_2O. Cycling conditions were set to an initial denaturation at 95°C for 3 min, followed by 20 cycles of 98°C for 10 s, 65°C for 10 s, 72°C for 30 s, and a final extension at 72°C for 5 min. Because the upper left adapter sequence (according to **Figure 1**) is changed by the bisulfite treatment, only the lower strands are amplified exponentially (**Figure 1D**). But because the genomic fragments are integrated in the two possible orientations with respect to the adapters (**Figure 1A**), it is possible to obtain the sequence information for both strands (**Figures 1E–G**; see section "Data Analysis").

Before submitting the genomic libraries to the sequencing service, it is necessary to eliminate very small (primer dimers, if present, or very short genomic fragments) and too large DNA fragments. We used the MagJet NGS Cleanup and Size Selection Kit (Thermo Fisher Scientific, Alcobendas, Madrid, Spain) with an initial binding mix volume of 400 µl for an average desired DNA fragment length of 300 bp (=approx. 200 bp insert). Due to budget restrictions the almond library was mixed as 1/12 part with samples of another independent project. As a consequence, coverage of the almond sequencing is the same as would be expected in a 96-plex experimental design. The library was sequenced by Macrogen on an Illumina 2500 machine (2 × 100 PE option).

Data Analysis

The data analysis is designed to build a catalog of the genomic fragments with the help of Stacks v2.4 (Catchen et al., 2013). In this catalog, the sequences that correspond to the same fragment but in the opposite orientation are separated in independent entries as Stacks is not designed to identify reverse complements. Furthermore, the sequences obtained from the both strands of the genomic DNA are not identical after reverse complementation because of the effect of the bisulfite treatment. Therefore, after catalog construction with the help of Stacks, custom designed software converts the original sequences to purine-pyrimidine sequences (**Figure 1F**). Because bisulfite treatment converts a pyrimidine (cytosine) to another pyrimidine (thymine), the reverse complements of reads with origin from opposite strands are identical when purines and pyrimidines are considered. Once identical reverse complements of the purine/pyrimidine sequences have been identified, we go back to the original sequences (**Figure 1G**). If one of the reads shows a thymine where the other shows a cytosine, the original state was an unmethylated cytosine, if both are cytosines, the original cytosine was protected from bisulfite action by methylation. In the supplement to this article we give detailed instructions on the use of the software pipeline. The provided material also contains shell scripts that can easily be adapted to user cases and then pasted into terminal windows for direct use.

In detail, the library was demultiplexed using the Stacks v2.4 component "process_radtags" (Catchen et al., 2013). It is important to use the "disable_rad_check" option, because the bisulfite treatment affects the *Pst*I recognition site. The sequences were then shortened by first eliminating the *Pst*I overhang of forward and reverse reads and truncating the sequences to 86 bases. As a consequence, all sequences across all samples are of the same length, independently of the length of the used barcode sequence, which simplifies the design of our own software. This was done using the "–fastqfilter" function of USEARCH v10 (Edgar, 2010) with the options "-fastq_trunclen 86" and "fastq_stripleft 4" for the forward reads and "-fastq_trunclen 86" and "fastq_stripleft 14" for the reverse reads. For other enzymes and/or read lengths, these parameters should be changed (see detailed information in the **Supplementary Material**). In both cases, it is mandatory to set the "–threads" option to one, because the default setting of "–threads" changes the order of reads in the output in an unpredictable manner on multicore systems, and as a result, forward and reverse reads no longer match if more than one thread is used. The resulting sequence pairs were then joined using "usearch –fastq_join –join_padgap ATATATAT – join_padgapq IIIIIIII" options. The resulting combined sequence consists of the two original reads separated by an artificial ATATATAT sequence, which is assigned a quality score of "IIIIIIII." The joined sequences were then quality-filtered by "usearch –fastq_filter –fastq_maxee_rate 0.01." This step eliminates sequences with a ≥ 0.01 probability of errors per base. The reads were aligned per sample into exactly matching stacks with "ustacks" with the default settings, and a catalog was built with "cstacks" ("-n 4" to allow 4 mismatches between sample loci). It is important to note that the output files of "cstacks" from Stacks v1 are organized in a different manner than those of Stacks v2. As a consequence, catalogs obtained with Stacks v1 are not compatible with our pipeline. The sequences of the catalog were read by the newly designed "creepi" program, which reconstructs the original sequences of each GBS locus before the bisulfite treatment and stiches them together to form a mock genome, which unifies the potential thousands of fragment sequences in one file for easier handling in the following steps. This rationale is similar to the one used in the GBS-SNP-CROP pipeline (Melo et al., 2016) and the bsRADseq software pipeline (Trucchi et al., 2016). Additionally, "creepi" eliminates the padgap part of the joined sequences, and if the forward and reverse reads overlap, the overlapping region is removed as well (merging).

"Creepi" also outputs a file with the individual sequences included in the mock genome together with the position of their boundaries in the mock genome and a plain fasta file with the individual sequences. The names given to the individual sequences are the line number corresponding to the first sequence that allowed the reconstruction of the original

sequence. This feature allows tracing back the reconstructed sequence to the catalog.

The original sequence reads were mapped to the mock genome, and the cytosine methylation states were determined with the help of Bismark v0.19.0 (Krueger and Andrews, 2011) with the default settings, with the exception that the "–non-_directional" option was used. The correlation between samples and differentially methylated sites were identified by methylKit 1.4.1 (Akalin et al., 2012). The difference of the "getMethylDiff()" function was set to 25 and the qvalue to 0.01. The second column of the output contains the positions of the differentially methylated cytosines. This information can easily be extracted to a file with the help of an R script given in the **Supplementary Material**, which can be used together with the fragment file produced by cre-epiGBS to identify the original fragments where these positions are located. We present a program (seek_fragments) that is designed to extract this information.

Figure 2 shows a flowchart of the software pipeline with the individual programs and their basic function.

We used SimRAD 0.96 (Lepais and Weir, 2014) to calculate the expected number of *Pst*I fragments of the *P. dulcis* Texas genome v2.0 (available at https://www.rosaceae.org/analysis/295) in given size ranges and to obtain the sequences of these fragments. We then searched a custom BLAST database constructed with the mock genomes obtained with our pipeline for homologous sequences.

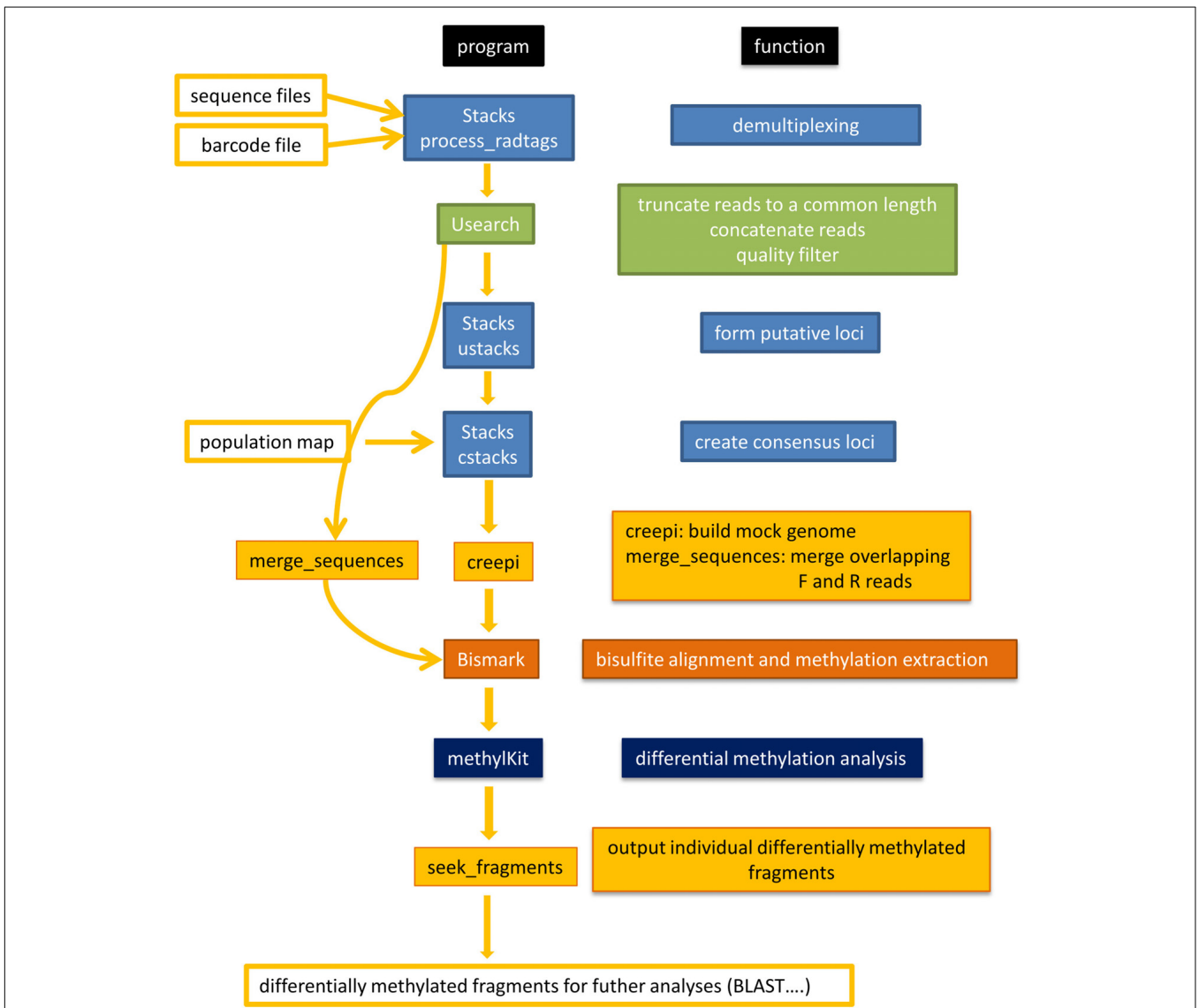


FIGURE 2 | Flowchart to illustrate the software pipeline. A selection of freely available software (Stacks, USEARCH, Bismark, methylKit) and own software (creepi, merge_sequences, seek_fragments) was used. The freely available software was chosen for its ease of installation and good documentation. The function is specified for each program.

RESULTS

With the modified adapters, we obtained 5,252,208–8,365,052 reads for the individual samples. After quality filtering, 2,055,858–3,494,318 joined sequences were retained, which means that 77.8–85.8% of the reads passed the quality filter. An initial test showed a low number of PCR clones (<1%), so filtering them out was therefore deemed unnecessary. The final mock genome consisted of 3,109 fragments. 2,467 of them showed homology with the *P. dulcis* reference genome. Of the homologous fragments, 1,813 produced one hit and the remaining fragments up to a maximum of five hits against the *P. dulcis* genome. Fragments that did not show homology with the *P. dulcis* genome were not filtered out, because 15% of them showed homology with other Rosaceae sequences in public databases, and no match was found for 45% of them (*E*-value cut-off: 0.001). The vast majority (36%) of the remaining non-*P. dulcis* fragments belonged to fungi, mainly the yeast-like *Pseudomicrostroma glucosiphilum* T. Kij. & Aime (Basidiomycota) and *Aureobasidium* ssp (Ascomycota). Under most scenarios fragments that do clearly not belong to the target organism can be filtered out easily at the end of the pipeline. The resulting mock genome had a length of 662,459 bp, which means 0.28% of the 246 Mbp *P. dulcis* genome (Sánchez-Pérez et al., 2019).

With SimRAD and the available *P. dulcis* reference genome, we calculated that 1,438 *Pst*I-*Pst*I fragments are expected to be in the range of 150–250 bp, 2,925 in the range of 100–300 bp and 4,179 in the range of 100–400 bp. We built a BLAST database of the mock genome with the makeblastdb order (Altschul et al., 1997) and searched the database against the expected sequences of the *in silico Pst*I digestion of the almond genome. Within the mock genomes, we found homologous sequences to these expected sequences in 1,193 fragments in the 150–250 bp range (83.0%), in 2,526 of the expected 2,925 fragments in the 100–300 bp range (86.4%) and in 3,266 of the expected 4,179 fragments in the 100–400 bp range (78.2%). The fragments created *in silico* produced always exactly one hit in the mock genome. The two BLAST searches together indicate that our pipeline merges on occasion different genetic loci with identical or nearly identical sequences as one locus of the mock genome. This fact also explains why the number of *in silico* fragments that hit the mock genome is higher than the number of fragments of which the mock genome consists. Increasing the parameter *-n* when building the catalog with "cstacks" might lead to more merged loci in the mock genome but lowering *-n* has the effect of considering fragments with different methylation states as distinct loci.

We then ran the Bismark alignment step with the obtained mock genome as genome file and extracted the methylation information. Of a total of 21,425,884 fragments 13,209,275 (61.65%) gave unique best hits and 115,465 did not map uniquely. We achieved a high coverage with mean values ranging from 211 to 342 reads per base with a minimum of 10 reads. The percentage of cytosines methylated in the CpG context was 31.23% (±0.53% SD) in the "Desmayo Largueta" cultivar and slightly lower at 30.00% (±0.77% SD) in the "Penta" cultivar. In the CHG context, 1.28% of the cytosines were methylated in both cultivars. The methylation state in the CpG context between biological replicates (same cultivar, same flowering stage, but different year) was highly correlated in all four cases (Pearson's $r = 0.99$), estimated by methylKit (Akalin et al., 2012). At this stage, 98.3% of the fragments with differentially methylated cytosines could be mapped against the *P. dulcis* genome in a local BLAST 2.8.1 (Altschul et al., 1997) search, while the remaining 1.7% could be matched against other Rosaceae sequences deposited in the GenBank. No other fragments with differential methylation were detected. Details on the biological importance of our results are published elsewhere (Prudencio et al., 2018). In summary, most of the observed differential methylation corresponded to differences between cultivars, but in ten fragments it was correlated with flowering stage.

With a small dataset like the one presented here the whole pipeline can be run in one day. Our developed programs require very limited computer resources. The creepi program, which makes the most extensive calculations of our own software occupied 56.2 MB of computer memory and needed 20.3 ± 0.7 s ($n = 5$) execution time on an Intel Xeon E5-2630 v4 (2.2 GHz) machine with 125.8 GB RAM. However, it should be noted that the execution time of creepi grows with the square of entries in the catalog produced by Stacks.

DISCUSSION

The recent papers of van Gurp et al. (2016) and Trucchi et al. (2016) make it possible to use NGS in studies of DNA methylation in non-model organisms. Based on these protocols, studies that involve a high number of individuals, like population genetics studies, are feasible at a reduced cost in the absence of available genome sequences in public databases. Nevertheless, the costs for hemimethylated adapters remain high and can be greater than the costs for Illumina sequencing in small-scale projects that use only a few sequencing lanes. In the absence of a reference genome, the protocol of Trucchi et al. (2016) also requires the parallel sequencing of non-bisulfite treated samples in order to reconstruct a reference for methylation calls. This is due to the fact that the software pipeline compares the bisulfite treated fragments to an untreated reference and when it encounters a thymine in the treated fragment where there is cytosine in the reference, it concludes that there was an unmethylated cytosine in the genome. Cytosines in the sequence of treated fragments correspond to methylated cytosines in the genomic DNA. The need of a reference was eliminated by van Gurp et al. (2016) using the information available in the complementary strands of the genomic DNA, looking for G-T and G-C base pairs.

Our protocol requires only one hemimethylated P2 adapter, while the barcoded P1 adapters are unmethylated. The unmethylated P1 adapters are hemimethylated by a nick translation reaction. As shown in **Figure 1**, it is possible to reconstruct the original sequence of a bisulfite-treated GBS fragment in the absence of a reference genome if the forward sequence and the reverse complement of the bisulfite treated DNA are available in a way similar to the original epiGBS

protocol. The necessity of additional sequencing of untreated DNA required in other protocols like bsRADseq is therefore eliminated. If standard GBS adapters are available, the two DNA oligonucleotides of the P2 adapter (one methylated) are the only ones that need to be newly synthesized. If there are no GBS adapters already available in the lab, the least expensive solution is to combine a high number of barcoded unmethylated P1 adapters and a low number of barcoded hemimethylated P2 adapters and calculate the cost for the adapter combination that reaches the minimum price for the desired number of samples. In the **Supplementary Table 1**, we show the calculation of the total cost of our method in comparison with published protocols for 96 samples. The savings are in the range of $2,000–$4,500 compared to the epiGBS protocol and $5,900–$8,300 compared to the bsRADseq protocol. The highest savings can be achieved when the lab already uses standard GBS barcoded P1 adapters. Another important factor is the price the manufacturers charge for each methylated cytosine, because there are huge differences between the different companies. The quantity of the oligos delivered by the manufacturers is sufficient for a high number of assays, so the cost advantage of our method per sample will be diluted in high throughput labs. As a result, our protocol is especially interesting for small labs or pilot studies with a low number of libraries to be sequenced.

We could reconstruct over 80% of the fragments in the 100–300 bp range. The percentage of fragments that can be reconstructed in other settings depends on several factors. The most important are genome size, the restriction enzyme used and the number of samples per sequencing lane. The *P. dulcis* genome is relatively small (aprox. 246 Mb, roughly double the size of the *A. thaliana* genome), and around 2,925 *Pst*I fragments are expected in the 100–300 bp size range. *Pst*I is a six-cutter restriction enzyme, and as a result, using for example a 4.5-cutter restriction enzyme like *Ape*KI frequently used in GBS would probably drastically reduce the mean coverage per base if the number of samples was not adjusted accordingly. If whole genome data of organisms close to the species in question are available, the use of bioinformatic instruments like SimRAD (Lepais and Weir, 2014) can help to find good starting points for the design of a project and fix an appropriate number of samples per sequencing lane. In the absence of genome data, SimRAD can also be used to generate a random genome of a given length and a fixed GC content. *C*-values for many plant species, which can easily be converted into genome length in bp, are available, for example, at the Plant DNA *C*-values Database of the Royal Botanical Gardens at Kew (data.kew.org/cvalues/).

The choice of the restriction enzyme also has consequences with respect to the genomic regions that are of special interest. For example, *Pst*I has the recognition site 5′-CTGCAG-3′. This site is not affected by CpG methylation but by CHG (CTG; CAG) methylation. Nevertheless, CHG or CHH methylation can be detected in the fragments obtained with *Pst*I. But in *A. thaliana* at least, CHG methylation is spatially autocorrelated (Cokus et al., 2008; Becker et al., 2011), and methylation rates in these fragments could be underestimated if partially methylated *Pst*I recognition sites are present (van Gurp et al., 2016). Available data (Becker et al., 2011; van Gurp et al., 2016) suggest that this behavior might steer the obtained fragments away from repetitive regions like transposable elements, which show a higher incidence of CHG methylation and favor the targeting of coding regions and their vicinity, which are less prone to CHG methylation. This might explain in part the considerable differences found between CpG methylation and CHG methylation in our results, which are similar to those found by van Gurp et al. (2016) for several plant species using *Pst*I. Nevertheless, Alvarez et al. (2019), also using *Pst*I as restriction enzyme, found an only slightly lower CHG than CG methylation in *Spartina alternifolia*. The data of van Gurp et al. (2016) also show that the coverage of chromosomal regions (1 MB window) with a high methylation rate is low in *A. thaliana* with *Pst*I epiGBS data. Depending of the aim of a project, this might be an advantage (if coding regions are of major interest) or a disadvantage (if an equal representation of the whole genome is the aim of the project). Restriction enzymes combinations like *Csp*6I-*Nsi*I (recognition sites G↓TAC and ATGCA↓T) represent all three methylation contexts (CG, CHG, and CHH) equally well and might therefore be better suited for certain experimental setups (van Moorsel et al., 2019).

Another important factor to be considered with respect to the choice of the restriction enzyme(s) to use is the number of expected fragments in the targeted size range. For example, Sonah et al. (2013) calculated that in soybean, *Mse*I produces 9.5 million fragments, *Ape*KI 800,000 fragments and *Pst*I 100,000 fragments, but in the case of *Mse*I, many fragments are below 100 bp, and in the case of *Pst*I, a high percentage of fragments has a length of over 500 bp. As a result, the number of usable fragments varies largely as a function of the restriction enzyme used. Enzymes that produce a low number of fragments in the size range used by NGS sequencing (like *Pst*I) are appropriate for genotyping a moderate number of markers with a high multiplexing level and large genomes (Hamblin and Rabbi, 2014), while frequent cutters can be used if the study aims to produce a high number of markers with a low multiplexing level and in organisms with small genomes. Schmidt et al. (2017), for example, showed that combinations of *Msp*I with *Dpn*II or *Ape*KI resulted in a high genome coverage and high cytosine coverage. Although these authors do not specify the number of sequencing lanes they used on an Illumina HiSeq 2500 machine, from the available data it is evident that the multiplexing level in their study was low. It should also be taken into account that high coverage is desirable for the calculations necessary to identify differentially methylated cytosines. Software like SimRAD (Lepais and Weir, 2014) may help to make the best possible decision regarding the choice of restriction enzyme, coverage, and multiplexing level. Most real-world scenarios where epiGBS is applied will depend on a high number of samples in order to find significant signals, especially in the field of molecular ecology and evolutionary biology. In our case, although we used only eight samples, they occupied 1/12 of the entire library sent for sequencing. Therefore, the coverage we found is expected to correspond to a 96-plex experiment. The high coverage in our experiment with mean values clearly above 200 reads indicates that using *Pst*I in an organism with a relatively small genome like *P. dulcis* allows for high multiplexing. Alvarez et al. (2019), who used *Pst*I in *Spartina alterniflora* with a genome

roughly seven times the size of the *P. dulcis* genome (Baisakh et al., 2009), processed 48 samples together.

If a reference genome is available, the fragments obtained by our lab protocol can be directly used as input for methylation extraction after the trimming of technical sequences (barcodes, wobble, etc.) and quality filtering. Therefore, epiGBS in its different variants might be an interesting option in organisms with known genome if a high number of samples is used and whole genome bisulfite sequencing is not cost-effective.

In comparison standard epiGBS protocol (van Gurp et al., 2016) there is no theoretical reason why our method should produce significantly different results when the same restriction enzymes are used. Our mock genome covers 0.28% of the 246 Mb almond genome, while van Gurp et al. (2016) covered 0.37% of the 135 Mb genome of *A. thaliana*. We recovered 86.6% of the theoretically expected *Pst*I fragments in the range of 100 – 300 bp, while in *Arabidopsis* 89% of the fragments in the range of 11 – 300 bp were found. We calculated a Pearson's R^2 of 0.98 between replicates while in *A. thaliana* this value was 0.95. Although other quality related indicators like less than 1% PCR clones, a high coverage of reads (mean value 211 – 342) similar read number for different samples (5,252,208–8,365,052 reads) and 80% of sequences passing the quality filter are adequate.

The method as presented here is limited to the use of only one restriction enzyme, but combinations of enzymes like *Pst*I-*Msp*I can be used if two sets of adapters are used. In the case of *Pst*I-*Msp*I, for example, this means that a set of unmethylated P1 adapters compatible with *Pst*I and another set compatible with *Msp*I are needed. The unmethylated P1 *Pst*I adapters are then combined in the restriction-ligation steps with hemimethylated P2 adapters with *Msp*I ends and the unmethylated P1 adapters with *Msp*I ends with hemimethylated P2 adapters with *Pst*I ends. In this case, the number of necessary adapters is double that of the one enzyme only case, but the cost should still be lower than that of the original epiGBS protocol.

The software is expected to work in the two-enzyme case as well, if adapters are adjusted and the software parameters are set accordingly (explained in the **Supplementary Material**). Nevertheless, the runtime is higher than in the case of a specially developed software, because in its present state the software does not take into account the strand information available under the two-enzyme scenario.

At the end of the pipeline presented here, information on differentially methylated cytosines is obtained. The fragments are exported in a fasta formatted file that can be used as input for other software packages like Blast2Go[1] for the functional analysis of the datasets.

Computer programs that are complicated to use may have a deterrent effect on scientists who are interested in a biological problem but are no computer experts. We have made a considerable effort in trying to make the bioinformatic pipeline as straightforward to use as possible. We found that the third-party computer programs we use in our pipeline are very well documented with detailed manuals and easy to install. The newly designed software is explained in the supplement and some common pitfalls for less experienced computer users are mentioned. The supplement includes all the orders (shell scripts) that are needed to get the programs to work (third-party and new). We gave the data and scripts mentioned in the supplement to Ph.D. students without prior experience in epigenetics and average knowledge in other fields of bioinformatics and they were able to follow the steps with only the help of the instructions given in the supplement.

Summarizing, the most important differences of our method in comparison with other RRBS protocols are that epiRADseq uses only the information of one restriction enzyme cut site, while we use the sequence of the whole fragment. On the other hand, BsRADseq requires the parallel sequencing of an untreated and a bisulfite treated library and epiGBS needs hemimethylated adapters on both sites of the genomic fragments created by the restriction enzymes while with our method a hemimethylated common adapter is sufficient. As a consequence, our variation of the epiGBS protocol of van Gurp et al. (2016) has the advantage of a much lower cost associated with the purchase of methylated adapter oligos. Existing GBS barcoded adapters can be used in combination with a hemimethylated P2 adapter. These advantages are especially important for smaller laboratories with limited financial resources. The high correlation of the methylation data of the biological replicates (r = 0.99) shows the reliability of the data sets created by our method.

DATA AVAILABILITY STATEMENT

DNA sequence reads generated for this study can be found in the NCBI SRA (accession numbers SRX7526585–SRX7526592). All newly designed software and shell scripts available at https://github.com/olafumes/creepiGBS.

AUTHOR CONTRIBUTIONS

OW conceived the method, performed the lab work, developed the software, analyzed the data, and wrote the manuscript. ÁP supplied the samples, performed the lab work, analyzed the data, and contributed to writing the manuscript. EC-M and MN-L tested the software and contributed to writing the manuscript. PM-G and RR analyzed the data and contributed to writing the manuscript.

FUNDING

This study was funded by the Spanish "Fundación Séneca" of the Region of Murcia (Grants 19308/PI/and 19879/GERM/15) and the "Ministerio de Economía y Competitividad" (Projects CGL2014-52579-R and RTI2018-095556-B-I00), co-financed by ERDF of the European Union.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00694/full#supplementary-material

[1]https://www.blast2go.com

REFERENCES

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., et al. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13:R87. doi: 10.1186/gb-2012-13-10-r87

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Alvarez, M., Robertson, M., van Gurp, T., Wagemaker, N., Giraud, D., Ainouche, M. L., et al. (2019). Reduced representation characterization of genetic and epigenetic differentiation to oil pollution in the foundation plant *Spartina alterniflora*. *bioRxiv* [preprint] doi: 10.1101/426569

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376. doi: 10.1371/journal.pone.0003376

Baisakh, N., Subudhi, P. K., Arumuganathan, K., Parco, A. P., Harrison, S. A., Knott, C. A., et al. (2009). Development and interspecific transferability of genic microsatellite markers in *Spartina* spp. with different genome size. *Aquat. Bot.* 91, 262–266. doi: 10.1016/j.aquabot.2009.07.007

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., et al. (2011). Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480, 245–249. doi: 10.1038/nature10555

Bewick, A. J., and Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Curr. Opin. Chem. Biol.* 36, 103–110. doi: 10.1016/j.pbi.2016.12.007

Boquete, M. T., Wagemaker, N. C. A. M., Vergeer, P., Mounger, J., and Richards, C. L. (2020). "Epigenetic approaches in non-model plants," in *Plant Epigenetics and Epigenomics. Methods in Molecular Biology*, Vol. 2093, eds C. Spillane and P. McKeown (New York, NY: Humana), doi: 10.1007/978-1-0716-0179-2_14

Bräutigam, K., and Cronk, Q. (2018). DNA methylation and the evolution of developmental complexity in plants. *Front. Plant Sci.* 9:1447. doi: 10.3389/fpls.2018.01447

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., and Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219. doi: 10.1038/nature06745

Cubas, P., Vincent, C., and Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401, 157–161. doi: 10.1038/43657

Doyle, J. J., and Doyle, M. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Dupont, C., Armant, D. R., and Brenner, C. A. (2009). Epigenetics: definition, mechanisms and clinical perspective. *Semin. Reprod. Med.* 27, 351–357. doi: 10.1055/s-0029-1237423

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple Genotyping by sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., et al. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1827–1831. doi: 10.1073/pnas.89.5.1827

Hamblin, M., and Rabbi, I. Y. (2014). The effects of restriction enzyme choice on properties of genotyping-by-sequencing libraries: a study in cassava (*Manihot esculenta*). *Crop Sci.* 54, 2603–2608. doi: 10.2135/cropsci2014.02.0160

Herrera, C. M., Medrano, M., and Bazaga, P. (2016). Comparative spatial genetics and epigenetics of plant populations: heuristic value and a proof of concept. *Mol. Ecol.* 25, 1653–1664. doi: 10.1111/mec.13576

Herten, K., Hestand, M. S., Vermeesch, J. R., and Van Houdt, J. K. J. (2015). GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics* 16:73. doi: 10.1186/s12859-015-0514-3

Hosaka, A., and Kakutani, T. (2018). Transposable elements, genome evolution and transgenerational epigenetic variation. *Curr. Opin. Genet. Dev.* 49, 43–48. doi: 10.1016/j.gde.2018.02.012

Johannes, F., and Schmitz, R. J. (2019). Spontaneous epimutations in plants. *New Phytol.* 221, 1253–1259. doi: 0.1111/nph.15434

Kebschull, J. M., and Zador, A. M. (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res.* 43:e143. doi: 10.1093/nar/gkv717

Kircher, M., Heyn, P., and Kelso, J. (2011). Addressing challenges in the production and analysis of Illumina sequencing data. *BMC Genomic* 12:382. doi: 10.1186/1471-2164-12-382

Krueger, F., and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. doi: 10.1093/bioinformatics/btr167

Lepais, O., and Weir, J. T. (2014). SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Mol. Ecol. Resour.* 14, 1314–1321. doi: 10.1111/1755-0998.12273

Li, Y., Kumar, S., and Qian, W. (2018). Active DNA demethylation: mechanism and role in plant development. *Plant Cell Rep.* 37, 77–85. doi: 10.1007/s00299-017-2215-z

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., et al. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536. doi: 0.1016/j.cell.2008.03.029

Manning, K., Tor, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., et al. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* 38, 948–952. doi: 10.1038/ng1841

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., et al. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461, 1135–1138. doi: 10.1038/nature08498

McClelland, M., Nelson, M., and Raschke, E. (1994). Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.* 22, 3640–3659. doi: 10.1093/nar/22.17.3640

Melo, A. T. O., Bartaula, R., and Hale, I. (2016). GBS_SNP_CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping by sequencing data. *BMC Bioinformatics* 17:29. doi: 10.1186/s12859-016-0879-y

Paun, O., Verhoeven, K. J. F., and Richards, C. L. (2019). Opportunities and limitations of reduced representation bisulfite sequencing in plant ecological genomics. *New Phytol.* 221, 738–742. doi: 10.1111/nph.15388

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *De Novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7:e37135. doi: 10.1371/journal.pone.0037135

Pikaard, C. S., and Scheid, O. M. (2014). Epigenetic regulation in plants. *Cold Spring Harb. Perspect. Biol.* 6:a019315. doi: 10.1101/cshperspect.a019315

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253. doi: 10.1371/journal.pone.0032253

Prudencio, ÁS., Werner, O., Martínez-García, P. J., Dicenta, F., Ros, R. M., and Martínez-Gómez, P. (2018). DNA methylation analysis of dormancy release in almond (*Prunus dulcis*) Flower buds using epi-genotyping by sequencing. *Int. J. Mol. Sci.* 19, 3542. doi: 10.3390/ijms19113542

Richards, C., Alonso, C., Becker, C., Bossdorf, O., Bucher, E., Colomé-Tatché, C., et al. (2017). Ecological plant epigenetics: evidence from model and non-model species, and the way forward. *Ecol. Lett.* 20, 1576–1590. doi: 10.1111/ele.12858

Rochette, N. C., and Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* 12, 2640–2659. doi: 10.1038/nprot.2017.123

Sánchez-Pérez, R., Pavan, S., Mazzeo, R., Moldovan, R., Aiese, C., Del Cueto, J., et al. (2019). Mutation of a bHLH transcription factor allowed almond domestication. *Science* 364, 1095–1098. doi: 10.1126/science.aav8197

Schield, D. R., Walsh, M. R., Card, D. C., Andrew, A. L., Adams, R. H., and Castoe, T. A. (2016). EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. *Methods Ecol. Evol.* 7, 60–69. doi: 10.1111/2041-210X.12435

Schmidt, M., Van Bel, M., Woloszynska, M., Slabbinck, B., Martens, C., De Block, M., et al. (2017). Plant-RRBS, a bisulfite and next-generation sequencing-based methylome profiling method enriching for coverage of cytosine positions. *BMC Plant Biol.* 17:115. doi: 10.1186/s12870-017-1070-y

Seymour, D. K., and Becker, C. (2017). The causes and consequences of DNA methylome variation in plants. *Curr. Opin. Plant Biol.* 36, 56–63. doi: 10.1016/j.pbi.2017.01.005

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* 8:e54603. doi: 10.1371/journal.pone.0054603

Soppe, W. J. J., Jacobsen, S. E., Alonso-Blanco, C., Jackson, J. P., Kakutani, T., Koornneef, M., et al. (2000). The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* 6, 791–802. doi: 10.1016/S1097-2765(05)00090-0

Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönswetter, P., and Paun, O. (2016). BsRADseq: screening DNA methylation in natural populations of non-model species. *Mol. Ecol.* 25, 1697–1713. doi: 10.1111/mec.13550

van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., and Verhoeven, K. J. F. (2016). epiGBS: reference-free reduced representation bisulfite sequencing. *Nat. Methods* 13, 322–324. doi: 10.1038/nmeth.3763

van Moorsel, S. J., Schmid, M. W., Wagemaker, N. C. A. M., van Gurp, T., Schmid, B., and Vergeer, P. (2019). Evidence for rapid evolution in a grassland biodiversity experiment. *Mol. Ecol.* 28, 4097–4117. doi: 10.1111/mec.15191

Wu, C.-T., and Morris, J. R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science* 293, 1103–1105. 10.1126/science.293.5532.1103

Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919. doi: 10.1126/science.1186366

Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of saver sex. *Curr. Biol.* 20, R780–R785. doi: 0.1016/j.cub.2010.07.00

Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* 18, 87. doi: 10.1186/s13059-017-1230-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

METHODS
published: 09 June 2020
doi: 10.3389/fpls.2020.00766

Semi-Automated Analysis of Digital Photographs for Monitoring East Antarctic Vegetation

Diana H. King[1,2], Jane Wasley[1,3], Michael B. Ashcroft[1], Ellen Ryan-Colton[1,4], Arko Lucieer[5], Laurie A. Chisholm[1] and Sharon A. Robinson[1,2]*

[1] Centre for Sustainable Ecosystem Solutions, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, NSW, Australia, [2] Global Challenges Program, University of Wollongong, Wollongong, NSW, Australia, [3] Antarctic Conservation and Management, Australian Antarctic Division, Kingston, TAS, Australia, [4] Research Institute for the Environment and Livelihoods, Charles Darwin University, Alice Springs, NT, Australia, [5] School of Technology, Environments and Design, College of Sciences and Engineering, University of Tasmania, Hobart, TAS, Australia

OPEN ACCESS

Edited by:
Bernard Goffinet, University of Connecticut, United States

Reviewed by:
Amber Churchill, Western Sydney University, Australia
Misao Itouga, RIKEN Center for Sustainable Resource Science (CSRS), Japan

****Correspondence:***
Diana H. King
dhk442@uowmail.edu.au

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *20 December 2019*
Accepted: *14 May 2020*
Published: *09 June 2020*

Citation:
King DH, Wasley J, Ashcroft MB, Ryan-Colton E, Lucieer A, Chisholm LA and Robinson SA (2020) Semi-Automated Analysis of Digital Photographs for Monitoring East Antarctic Vegetation. Front. Plant Sci. 11:766. doi: 10.3389/fpls.2020.00766

Climate change is affecting Antarctica and minimally destructive long-term monitoring of its unique ecosystems is vital to detect biodiversity trends, and to understand how change is affecting these communities. The use of automated or semi-automated methods is especially valuable in harsh polar environments, as access is limited and conditions extreme. We assessed moss health and cover at six time points between 2003 and 2014 at two East Antarctic sites. Semi-automatic object-based image analysis (OBIA) was used to classify digital photographs using a set of rules based on digital red, green, blue (RGB) and hue-saturation-intensity (HSI) value thresholds, assigning vegetation to categories of healthy, stressed or moribund moss and lichens. Comparison with traditional visual estimates showed that estimates of percent cover using semi-automated OBIA classification fell within the range of variation determined by visual methods. Overall moss health, as assessed using the mean percentages of healthy, stressed and moribund mosses within quadrats, changed over the 11 years at both sites. A marked increase in stress and decline in health was observed across both sites in 2008, followed by recovery to baseline levels of health by 2014 at one site, but with significantly more stressed or moribund moss remaining within the two communities at the other site. Our results confirm that vegetation cover can be reliably estimated using semi-automated OBIA, providing similar accuracy to visual estimation by experts. The resulting vegetation cover estimates provide a sensitive measure to assess change in vegetation health over time and have informed a conceptual framework for the changing condition of Antarctic mosses. In demonstrating that this method can be used to monitor ground cover vegetation at small scales, we suggest it may also be suitable for other extreme environments where repeat monitoring via images is required.

Keywords: Antarctica, moss, vegetation cover, vegetation health, climate change, monitoring, OBIA, RGB

INTRODUCTION

Climate change has caused range shifts in vegetation communities toward the poles and alpine regions, as well as changes in phenology and growth worldwide (IPCC, 2014; Pecl et al., 2017). In order to inform management of ecosystems, determine community dynamics and detect change in vegetation communities over time, vegetation must be monitored over the long-term (Howard-Williams et al., 2006; Barrett et al., 2008; Michaels and Power, 2011). Where fieldwork is difficult, dangerous, remote or expensive, it is much more challenging to ensure adequate monitoring. One such example is long-term monitoring of vegetation in Antarctica, where there are logistical difficulties relating to the expense of sending researchers to Antarctica, frequency and duration of visits, weather-related delays, field surveys conducted in the extreme cold and mechanical difficulties with equipment. In addition, protective regulations may limit certain methodologies e.g., the Antarctic Treaty System principles necessitate minimal destructive sampling of vegetation.

Plant productivity and growth is limited in the harsh Antarctic environment, due to the related subzero temperatures and lack of free water (Johansson and Thor, 2008; Wasley et al., 2012). Plant life on the Antarctic continent is dominated by lichens and mosses and restricted to ice-free regions, which comprise less than one percent of the continent (Lee et al., 2017; Robinson et al., 2018). Given that Antarctic moss turfs (see **Supplementary Figure 1** for examples) span areas of centimeters to meters and are comprised of mosses of small size (<15 cm depth) and high shoot density (550–900 shoots per cm^2) (Wasley et al., 2006), some vegetation monitoring techniques typically applied at larger scales for remote sites may not be suitable. For example, remote sensing techniques can repeatedly monitor the state of Antarctic vegetation at distinct time points, which reduces time spent in the field and minimizes destructive impacts upon the vegetation communities, in accordance with Antarctic Treaty System principles. However, due to the small size of the moss turfs, even the highest resolution satellite imagery available (2.2 m pixel resolution WorldView-2) is unsuitable for vegetation studies in these areas (Malenovský et al., 2017). Digital photography with very high (1–4 m) and ultra-high (<1 m) spatial resolution, taken from the air (via unmanned aircraft systems, aka UASs) is ideal (Turner et al., 2012, 2014, 2018; Bollard-Breen et al., 2015; Malenovský et al., 2017), however, it is expensive to obtain and operate suitable UASs in Antarctica, they require a licensed pilot (and thus increased logistics support and funding), and weather conditions make their use difficult in certain locations. The small size of the mosses, in addition to the often prohibitive costs of obtaining suitably scaled aerial/UAS photography in Antarctica, means that digital photography using handheld cameras may still be the most suitable low-cost option for long-term monitoring of many Antarctic vegetation communities. Handheld digital photography is very easy to accomplish, and can be done by anyone with a camera and GPS locations of quadrats, making repeat monitoring much more feasible. The challenge lies in determining the balance between monitoring parameters and aligning these with current sensor technology, whilst future-proofing methods to enable the spatio-temporal monitoring necessary to assess the impacts of global climate change.

Standard red, green, blue (RGB) digital photography has been successfully utilized in vegetation studies to determine vegetation cover (Bennett et al., 2000; McCarthy and Zaniewski, 2001; Booth et al., 2005a; Laliberte et al., 2007b; Greenwood and Weisberg, 2009; Ko et al., 2009; Haywood and Stone, 2011; Kim et al., 2011), vegetation type (Ehlers et al., 2006; Lathrop et al., 2006; Luscier et al., 2006; Yu et al., 2006; Hájek, 2008; Greenwood and Weisberg, 2009; Laliberte et al., 2010; Michel et al., 2010; Cserhalmi et al., 2011; Whiteside et al., 2011) and vegetation changes over time (Bennett et al., 2000; Ehlers et al., 2006; Cserhalmi et al., 2011). As vegetation communities have complex characteristics, with patches varying in size, internal homogeneity and discreteness, it makes sense to analyze these communities based on combinations of their spatial and spectral patterns (Blaschke and Strobl, 2001). Object-based image analysis (OBIA) is a useful technique to analyze such communities, with images being segmented into relatively homogeneous areas to create meaningful objects for analysis (Blaschke and Strobl, 2001; Liu and Xia, 2010), with rules developed to isolate elements of interest. These rules are objective, and not prone to the errors associated with subjective human perception and interpretation of vegetation patterns (Elzinga et al., 1998; Van Coillie et al., 2014), making them ideal for use for long-term monitoring applications.

Numerous studies have used OBIA methodology for the evaluation of vegetation community characteristics in airborne and space-borne imagery, with advances in analytical software and technology leading to a marked increase in the usage of effective image segmentation methods (Hay et al., 2005; Bunting and Lucas, 2006; Conchedda et al., 2007; Laliberte et al., 2007a; Jobin et al., 2008; Pringle et al., 2009; Berberoglu et al., 2010; Blaschke, 2010; Laliberte and Rango, 2011; Torres-Sánchez et al., 2015). However, few studies have used OBIA for percent cover analysis of vegetation in near-surface field photographs (Luscier et al., 2006; Laliberte et al., 2007b; Zhang et al., 2007; Michel et al., 2010), with only one study using OBIA to measure vegetation percent cover in a polar region (Chen et al., 2010). Classification methods using fuzzy logic based on neighboring pixel values have been utilized in vegetation classification analyses, leading to high classification accuracy (Luscier et al., 2006; Laliberte et al., 2007b), however, these methods can lead to confusion between vegetation classes (Chen et al., 2010). Rule-based classification of vegetation in an Arctic study was successfully performed with high accuracy, using both RGB bands and HSI (hue, saturation, and intensity) bands (transformed from the original RGB images) threshold values (Chen et al., 2010), as HSI transformations can be very effective for vegetation classifications (Laliberte et al., 2007b).

For sensitive and remote sites, and particularly for short stature vegetation, the use of OBIA to assess digital RGB photographs of quadrats enables vegetation cover analyses to be performed in a timely and non-destructive manner, saving on field costs and labor. Digital photographs can also be easily stored and provide an invaluable source of data for other monitoring applications. Here we introduce an object-based image analysis

methodology to assess short stature vegetation percent cover and compare its results to those of two other methods: manual digitization and expert visual estimates. We then assess temporal change using Antarctic moss turfs as a model short stature plant community.

MATERIALS AND METHODS

A long-term vegetation monitoring system was established in 2003 at two sites in the Windmill Islands, East Antarctica: Antarctic Specially Protected Area (ASPA) 135 (site A2) and Robinson Ridge (site RR) (as per Wasley et al., 2012; Malenovský et al., 2017; Robinson et al., 2018) (see **Supplementary Figure 1** for site maps). These sites exhibit some of the most extensive moss beds in the region, and have been extensively described in Lovelock and Robinson (2002), Dunn and Robinson (2006), Wasley et al. (2012), and Turner et al. (2014). Sixty permanent 25 × 25 cm quadrat locations have been repeatedly monitored in three vegetation communities at these sites (10 quadrats each in Bryophyte, Transitional and Lichen communities at each site) between 2003 and 2014 (Robinson et al., 2018). These communities occur along a moisture gradient, the wettest dominated by bryophytes (Bryophyte community, approximately 90% live bryophyte cover); the driest dominated by lichens overgrowing moribund moss (Lichen community, <5% live bryophyte cover); and the community in between comprised of both types of vegetation (Transitional community, approximately 40% live bryophyte cover) (Ryan-Colton, 2007). Digital RGB photographs of each quadrat were taken in Jan–Feb five times between 2003 and 2013, as well as the Bryophyte community quadrats in 2014 (**Supplementary Table 1**). Unfortunately, the Lichen community could not be analyzed digitally for vegetation cover, as macrolichens, which dominated cover in this community, were not suitable for this analysis method. This was due to difficulties differentiating between vegetation cover types in this community using only RGB spectral characteristics, as the range of different black, gray and white lichens are visually very similar to black, gray and white rocks.

As the moss beds are sensitive vegetation, easily damaged by trampling and disturbance, care had to be taken to prevent standing on the moss, thus handheld photographs had to be taken from above the quadrat whilst standing on the nearest rocks to the quadrat location. All care was taken to stand in the same position each subsequent year for photography, however, differences in camera height and position were inevitable between different quadrats and years. Different cameras were also used in different field seasons, as technology improved over time (**Supplementary Table 1**). All photography acquired over time kept settings in mind so to best replicate consistency across different cameras (composition, adjustments for exposure, etc.), but exact duplication of the camera setup is not necessary in order to detect change over time (Rogers et al., 1983; Hall, 2002). Reference photos from the baseline 2003 field season were used to correctly reposition the physical metal quadrat each subsequent field season, however, as frost heave causes moss to move within a quadrat over time, the physical positions of the quadrats do change slightly between years, thus direct pixel to pixel change detection over time is problematic. Determining the change in percent cover of moss within each quadrat was therefore the most efficient method to assess vegetation health over time.

Image Analysis

To assess the feasibility of using object-based image analysis for estimation of percent cover of short-stature vegetation communities, first a test image mosaic was created from nine digital quadrat photographs (**Figure 1**). Photographs taken at 5-year intervals (2003, 2008, and 2013) were included for each of three representative quadrats included in the test mosaic, out of the total 40 quadrats (excluding Lichen community quadrats) in the monitoring system (**Figure 1**). These images included moss in the full range of health states from bright green/olive green healthy moss, through to red/brown stressed moss and black/gray moribund moss. Moribund moss is moss that appears dead (gray to black in color) and has no evidence of chlorophyll presence in the leaves when examined microscopically (see inset images in **Figure 8**). Although appearing dead, moribund moss can sometimes, under ideal conditions, regenerate providing some healthy cells are present (Robinson et al., 2020).

The quadrat images were georeferenced using ArcGIS v 9.3 (ESRI, Redlands, CA, United States). The spatial resolution of the quadrat images changed with the varied positioning of quadrats within the images and improved camera quality over time, from 2048 × 1536 pixels in 2003 to 6000 × 4000 pixels in 2014 (**Supplementary Table 1**). All images were rectified to a pixel size of 0.0005 m, using nearest neighbor resampling to retain original pixel values for vegetation classification (Verbyla, 2002), allowing direct comparison of results between quadrats and across years.

As the area within each quadrat was the focus for vegetation classification, the physical quadrat itself, as well as the area outside the quadrat in each image, was masked out ("Quadrat polygons," **Supplementary Figure 2**). As these images were simple RGB digital photographs, it was difficult to distinguish between some of the ground cover classes, which could typically be differentiated by spectral characteristics, for example: gray rock vs. gray moribund moss, black rock vs. black moribund moss vs. black macrolichens and white rock vs. white macrolichens. As health status of moss was the principal focus of monitoring, manual digitization was undertaken to mask out all ground cover classes in the quadrats except moss turf (e.g., rock, crustose and macro-lichens, snow and water) ("Rocks polygons," "Crustose polygons," "Macrolichen polygons," "Snow polygons," and "Water polygons," respectively, **Supplementary Figure 2**). Some quadrats contained areas of very wet moss, which was so dark in color that it tended to be incorrectly classified as moribund moss, so these areas were also manually digitized and classified ("WetMoss polygons," **Supplementary Figure 2**).

Many different methods exist for object-based image analysis, however, *eCognition* (Trimble, München, Germany) is the most common commercial software package capable of performing a

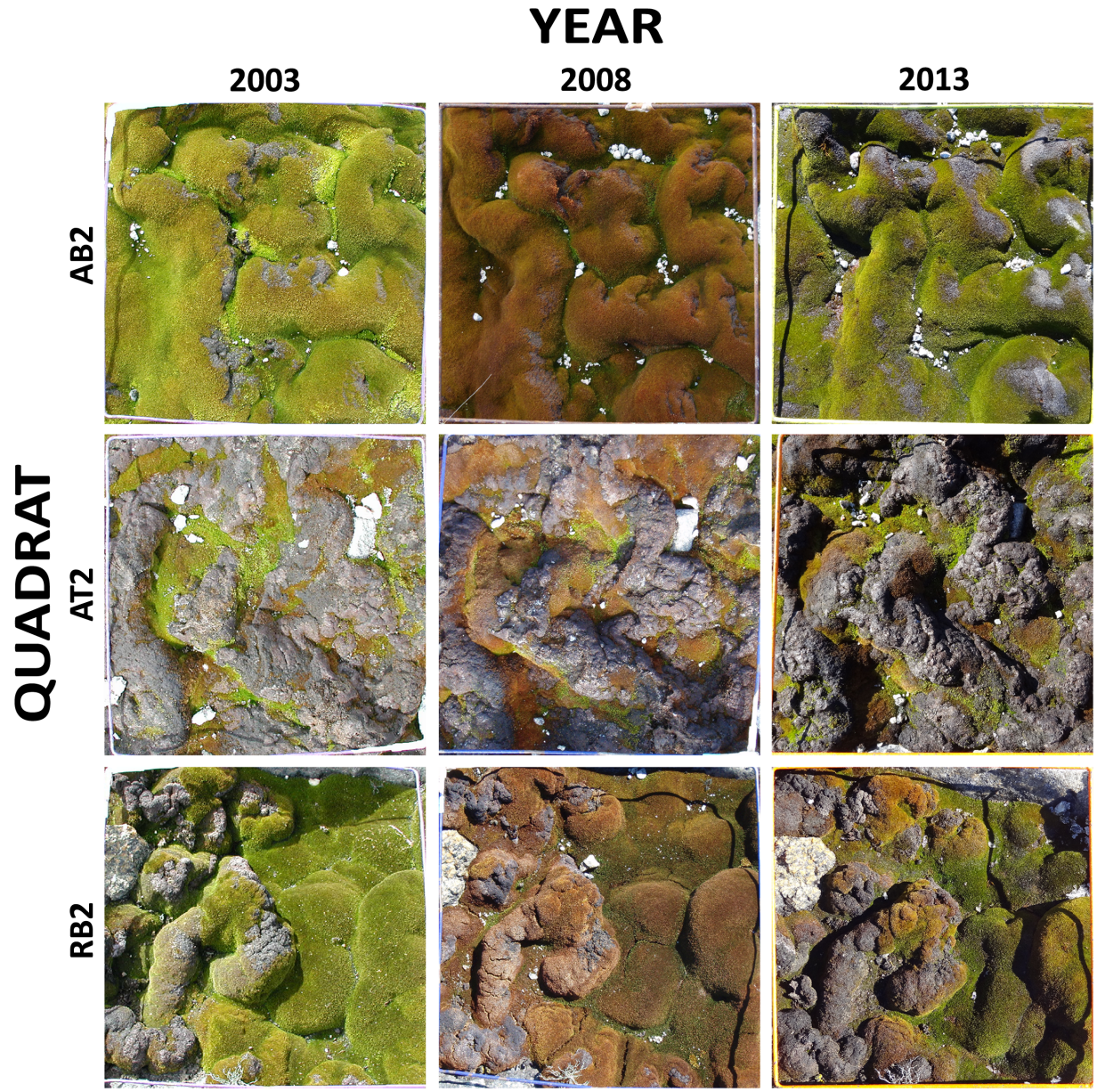


FIGURE 1 | Test mosaic of nine quadrat images from sites in the Windmill Islands, East Antarctica. The mosaic comprises photos of three quadrats, taken at 5 year intervals (2003, 2008, and 2013). AB2: ASPA 135 Bryophyte community quadrat 2, AT2: ASPA 135 Transitional community quadrat 2, RB2: Robinson Ridge Bryophyte community quadrat 2.

multiscale image segmentation approach, ideal for land cover analysis (Blaschke, 2004). Therefore, to assess which parameters and rules would be most successful in classifying vegetation into categories of healthy, stressed or moribund, the image mosaic for the quadrats, as well as the associated masks, were imported into the eCognition OBIA software for analysis (**Supplementary Figure 2**).

Segmentation

Optimal settings for image object segmentation were determined using the Estimation of Scale Parameter (ESP) tool (Drăguţ et al., 2010). This tool allows the user to improve object segmentation parameterization in eCognition by showing changes in local variance (LV, the mean of all standard deviations of a moving 3 × 3 pixel area across the image), and rate of change (ROC) across a range of scale parameters (Drăguţ et al., 2010). Peaks in the ROC-LV graph indicate the most appropriate object levels for segmentation, and segmentation at each of these optimal levels can be visually assessed to determine which scale level is necessary to segment meaningful objects of interest for subsequent image classification. Multiresolution image segmentation was performed at different optimal scales,

and the segmentation results were visually compared. Once optimal segmentation parameters were determined using the ESP tool, multi-resolution image segmentation was performed on the image mosaic, with scale parameter 27, shape 0.1 and compactness 0.5, to create image objects for classification. The manually digitized polygons created during pre-processing were included as thematic layers during image segmentation, in order to be used as masks for ground cover classification (**Supplementary Figure 2**).

Classification

Various object features were visually assessed to determine those best suited to delineating the three different health states of the moss turf (healthy, stressed, and moribund). Upon evaluation, the OBIA features found to be most useful for object classification included: pixel-based ratios of red, green and blue (RGB); Intensity and Saturation from the hue, saturation and intensity (HSI) color space (converted from RGB within the eCognition software); relative border to another class; and area of objects. Red, green and blue pixel ratios (to overall brightness) were used instead of mean R, G and B values, as ratios were less affected by different lighting conditions and different cameras (**Supplementary Table 1**) (Jensen, 2013).

To aid the visual assessment of thresholds between classes, object features were separated into 15 Jenks Natural Breaks categories, classes being determined by natural breaks between clusters in the data, maximizing the differences between classes and grouping similar values. The threshold was assessed visually and the value was determined as the digital value separating the most appropriate Jenks Natural Breaks categories. This value was then used as the threshold value for object classification within the eCognition ruleset.

The different red, green and blue pixel ratios alone were not enough to definitively separate the ground cover classes, however, when combined into Death index1 (Eq. 1) they were useful for classifying the majority of moribund moss, with high values indicating moribund moss. Death index2 (Eq. 2) was useful for distinguishing the remaining moribund moss, with lower values indicating moribund moss. A Stress index (Eq. 3) indicated stressed moss with high values indicating most stress, i.e., a higher contribution of red to overall object brightness. Conversely, a low Stress index combined with high pixel ratios of green indicated healthy moss. Both HSI Intensity and Luminance (Eq. 4, where PbR denotes "Pixel based Ratio of," and R, G, and B are the red, green and blue bands, respectively), were found to be useful in distinguishing shadows (black, low Intensity and low Luminance) and small wind-blown rocks (white, high Intensity and high Luminance) (Trussell and Vrhel, 2008).

$$Death\ index1 = \frac{\text{PbR B}}{(\text{PbR G} * \text{PbR R})} \quad (1)$$

$$Death\ index2 = \frac{\text{HSI Intensity}}{(\text{HSI Saturation} * \text{PbR R} * \text{PbR G})} \quad (2)$$

$$Stress\ index = \frac{\text{PbR R}}{\text{PbR G}} \quad (3)$$

$$Luminance = (0.2126 * PbR\ R) + (0.7152 * PbR\ G) + (0.0722 * PbR\ B) \quad (4)$$

The classification process involved a series of manually refined rules determining which class an object should be assigned to, based on digital value thresholds, followed by further refinement of certain classes based on position and size of an object (**Supplementary Figure 2**). For example, if an object was surrounded by healthy moss, and the object was very small, it was likely to be a small wind-blown rock. Similarly, areas of shadow were reassessed depending on whether the border of an object was predominantly shared with moss of a particular health state, for example if an object was classified as shadow and shared more than 50% of its border with objects classified as healthy moss, it is likely that that area of shadow was actually healthy moss.

The results of the classification were assessed visually for each quadrat. Post-processing involved manual correction of some small misclassification issues, for example areas of wet healthy moss (very dark green, too dark for the classification algorithms to detect as live moss), not digitized in the pre-processing stage, and misclassified as moribund, were corrected upon expert visual assessment. Object statistics were exported, in order to obtain a final pixel area of each class, for later conversion to a final percent cover. A tif image was also exported to show the classification results for each quadrat. Finally, a polygon shapefile was exported with a field for class name, for later visualization of the classification.

Accuracy Assessment

This monitoring methodology is designed to replace field estimation techniques of percent cover of vegetation, and it is therefore imperative to ascertain the accuracy of the automated classification process, including both the thematic accuracy of the classification as well as how the classification results compared with the most commonly used current field technique, the vegetation cover accuracy.

Thematic Accuracy

The subset of nine quadrat images used for data exploration and assessment was subsequently used for accuracy assessment of the rule-based classification. The map document created for the original image pre-processing was further used for manual classification of the mosaic by manual digitization of the various vegetation classes. Two of the 9 quadrat images in the mosaic were also independently digitized by a further three experts, in order to achieve a majority consensus of the classification of the vegetation types in the quadrats that were most difficult to assess. This majority consensus was used in the final manual classification shapefile, in addition to the other seven manually classified quadrats.

Following semi-automated classification, the classification result was imported into the pre-processing map document and compared with the manual classification. Manual and semi-automated classification results were both spatially joined to a grid of points spaced 0.005 m apart across the entire

mosaic (one point per pixel in the mosaic, a total of 21608 points) to compare vegetation cover classifications, and the resulting table was exported as a database file for further analysis.

The statistics program SPSS v. 21 (IBM, New York, NY, United States) was used to compare the manual and rule-based classification results database files, and calculate the Conditional Kappa coefficient as a measure for thematic accuracy, to test individual category agreement between two different classifications (Congalton and Green, 2009). The Kappa analysis is a standard component of every accuracy assessment of image analysis, and is a measure of how well the image classification matches with the reference (ground truth) data (Congalton and Green, 2009). Values greater than 0.80 represent strong agreement (more than 80% of the image classification is the same as the ground truth data), 0.4 to 0.8 represent moderate agreement, and less than 0.4 represent poor agreement (Congalton and Green, 2009).

Vegetation Cover Accuracy

To compare with traditional visual percent cover estimates, three Antarctic moss experts were given the same three images to analyze, one from each year of sampling for a single quadrat (bottom row of mosaic **Figure 1**, quadrat RB2 from 2003, 2008, and 2013). The quadrat was overlaid with a 5 × 5 cm grid, mimicking the grid used in field estimates of percent cover throughout the Antarctic monitoring program. The experts assessed the percent cover of the various classes for each square in the grid, and the results were tallied to calculate total percent cover for each quadrat.

Vegetation Change Assessment

Semi-automatic classification of digital photography was used to assess the change in vegetation health of continental Antarctic vegetation monitoring quadrats between 2003 and 2014. A total of 210 quadrat images from the Bryophyte and Transitional communities at both sites were acquired over the six field seasons. Some quadrats in some years were snow-covered, and could not be used for vegetation classification (10 photographs out of 220 over 6 years). The Transitional community was not photographed in 2014 due to field time constraints.

Image analysis occurred as above, including georeferencing, preprocessing, and semi-automated object-based image analysis for classification of vegetation health. A small number of photographs had unforeseen classification issues, such as areas of wet moss which were not immediately apparent in the preprocessing stage. Following manual correction, the output was saved, and a note made that this photograph had been manually corrected. This process was required for 47 photographs out of the entire set of 210 photographs of quadrats classified from the six field seasons, particularly those from 2011 and 2014 due to snow and wet moss.

Vegetation cover change over time was analyzed using a non-parametric Friedman Test in IBM SPSS v25 (IBM Corp., Armonk, NY, United States) for each vegetation cover category. Pairwise comparisons were performed (IBM SPSS v25) with a Bonferroni correction for multiple comparisons.

RESULTS

Image Analysis

Segmentation

The Estimation of Scale Parameter (ESP) tool determined optimal segmentation scales for the test mosaic of 27 and 33 for peaks in the rate of change (**Figure 2**). These values were therefore identified as optimal scale parameter settings for segmentation of the mosaic, with a scale of 27 ultimately selected as optimal for image segmentation, as there were more objects containing mixed vegetation categories when segmented with a scale of 33, which were separate objects when segmented with a scale of 27 (**Figure 3**). Scales higher than 33 were found to create objects that were too large for appropriate separation of vegetation types.

Classification

Classification of the image mosaic was an iterative process, with rules using thresholds to classify objects into intermediary classes, reclassifying objects, merging objects, looping classifications and final classification into final classes (**Supplementary Figure 3**).

Using image segmentation and classification of the test mosaic of nine quadrat images, the total estimated percent cover of each category across the entire mosaic was: 45.3% healthy moss, 23.7% stressed moss, 26.5% moribund moss, 4.0% rock, 0.2% lichens, 0.1% snow and 0.3% shadow. There were no water, wet moss or unclassified pixels in the test mosaic. Percent cover was also calculated for each of the nine quadrats (**Figure 4**).

Accuracy Assessment

Thematic Accuracy

When comparing the semi-automated classification results with the manually digitized classification, the total classification accuracy was 84%, with an overall Kappa value of 0.76 (**Table 1**). The Conditional Kappa values ranged from 0.67 for stressed moss to 0.95 for rock. The greatest errors of omission (i.e., areas incorrectly omitted from a class) were 25% moribund moss and 14% stressed moss. The greatest errors of commission (i.e., areas incorrectly classified in a class) were 26% stressed moss and 17% healthy moss.

Vegetation Cover Accuracy

When comparing the semi-automated classification with that of three experts using traditional percent cover estimation techniques, the semi-automated classification results fell within the range of expert percent cover estimates for the majority of classes across the three quadrats (**Figure 5**). The average ranges between expert observers across the three quadrats were 7% for healthy moss, 12% for stressed moss, 2% for moribund moss, 1% for macrolichen, 2% for rock, and 3% for shadow. The semi-automated classification results were not significantly different from the experts for healthy moss, moribund moss, rock or lichens, across all three quadrats. Stressed moss was not significantly different between experts and semi-automated

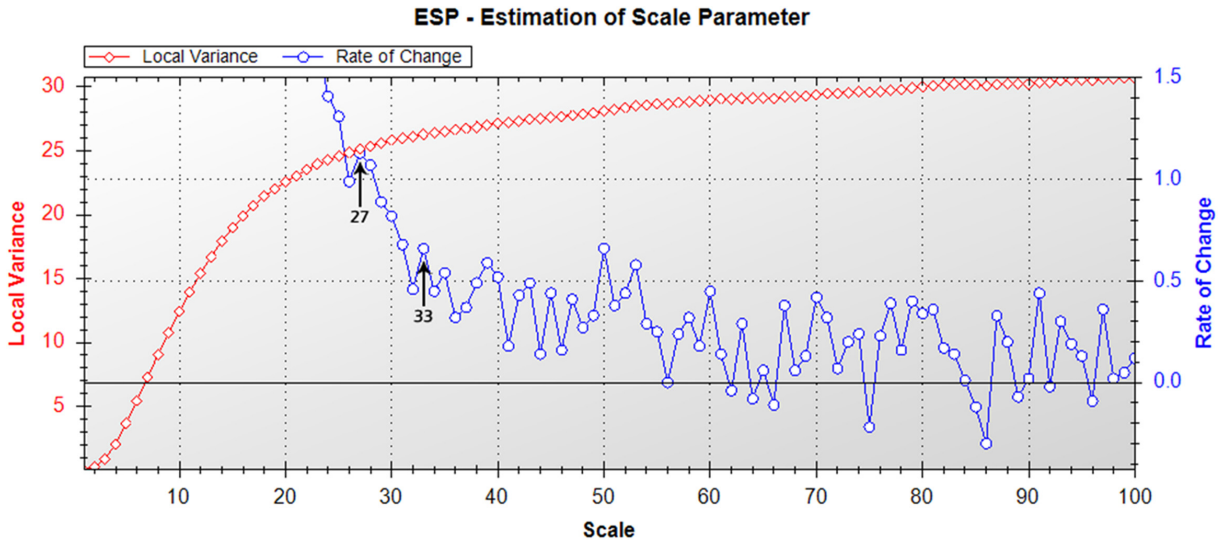


FIGURE 2 | Estimation of scale parameter tool output (Drăguţ et al., 2010) for the nine quadrat image mosaic. Graph depicts changes in local variance (LV, red) and rate of change (ROC, blue) with increasing scale parameter for segmentation of the moss quadrat test mosaic. Peaks in the ROC-LV graph indicate the most appropriate object levels for segmentation. Suitable scales for image segmentation for the mosaic were 27 or 33, as indicated by arrows.

FIGURE 3 | Comparison of scale parameters 33 (top) and 27 (bottom). Scale 27 was chosen as optimal for image segmentation, as it separated out the smaller areas of healthy, stressed and moribund moss, which formed mixed objects when the image was segmented with scale 33, as seen in the three circled areas.

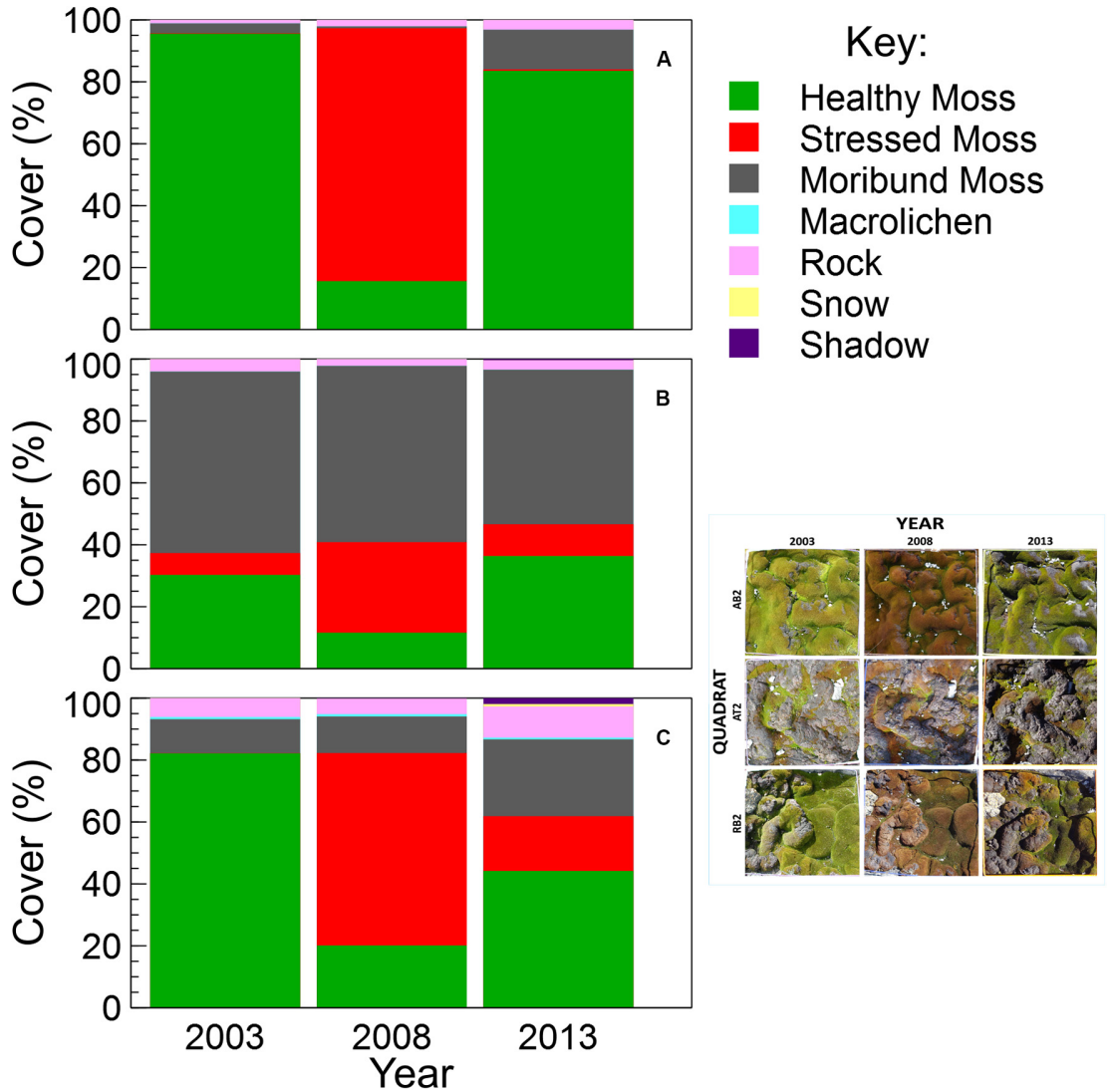


FIGURE 4 | Comparison of vegetation percent cover within each quadrat in the test mosaic (**Figure 1**; inset at right): **(A)** AB2, **(B)** AT2, and **(C)** RB2, at 5-year intervals from 2003 to 2013, assessed using semi-automated object based image analysis.

classification when there was more than 15% present, however, in quadrat R2B2003 the percent cover estimates varied significantly [$X^2_{(4,\ N\ =\ 24)}$ = 31.19, $p \leq 0.0001$], with semi-automated classification estimating 0% stressed moss, and the experts ranging from 4 to 14%. Shadow was significantly underestimated in all quadrats in the semi-automated classification [R2B2003: $X^2_{(4,\ N\ =\ 10)}$ = 13.32, p = 0.01; R2B2008: $X^2_{(4,\ N\ =\ 13)}$ = 14.34, p = 0.006; R2B2013: $X^2_{(4,\ N\ =\ 36)}$ = 30.79, $p \leq 0.0001$].

Vegetation Change Assessment

This method was successfully applied to all bryophyte (Robinson et al., 2018) and transitional community quadrats at six time points between 2003 and 2014 (**Figures 6**, **7**). Vegetation health in both communities at ASPA 135 and Robinson Ridge experienced a general decline in 2008, followed by recovery by 2014 (**Figures 6**, **7**). Healthy moss cover declined in 2008 to less than half of the 2003 cover at both sites in both communities (**Figures 6**, **7**). This decline in healthy moss was associated with an equivalent significant increase in stressed moss cover from 2003 to 2008, in both communities at each site (**Figures 6**, **7**). The amount of subsequent recovery differed between sites. ASPA 135 quadrats completely recovered to 2003 baseline levels of health by 2014 in the Bryophyte community (**Figure 6**), whilst no significant change over time was observed in the Transitional community (**Figure 7**). However, Robinson Ridge quadrats still appeared to be recovering in 2014, with a 10-fold increase (3% in 2003 to 35% in 2014) of stressed moss in the Bryophyte community (**Figure 6**), and an increase of moribund moss from 42% to 62% between 2003 and 2014, respectively, in the Transitional community (**Figure 7**). The cover of other abiotic factors, particularly snow, was variable between years (**Figures 6**, **7**).

TABLE 1 | Comparison of test mosaic classification by eCognition and manual classification by expert.

(a)		Manual classification							
		Healthy moss	**Stressed moss**	**Moribund moss**	**Lichens**	**Rock**	**Snow**	**Shadow**	**Total**
eCognition Classification	Healthy moss	38.03	2.13	5.16	0.00	0.25	0.00	0.00	**45.57**
	Stressed moss	3.24	17.54	2.89	0.00	0.06	0.00	0.00	**23.74**
	Moribund moss	1.14	0.67	24.27	0.00	0.12	0.00	0.00	**26.20**
	Lichens	0.01	0.00	0.01	0.15	0.00	0.00	0.00	**0.17**
	Rock	0.12	0.03	0.06	0.00	3.78	0.00	0.00	**3.99**
	Snow	0.01	0.00	0.00	0.00	0.00	0.08	0.00	**0.09**
	Shadow	0.12	0.05	0.08	0.00	0.02	0.00	0.00	**0.26**
	Total	**42.66**	**20.42**	**32.46**	**0.16**	**4.21**	**0.08**	**0**	**100**

(b)	Errors of omission	Errors of commission	Conditional kappa
Healthy moss	11%	17%	0.71
Stressed moss	14%	26%	0.67
Moribund moss	25%	7%	0.89
Lichens	6%	11%	0.89
Rock	10%	5%	0.95
Snow	6%	11%	0.89
Shadow	–	100%	0.0
Total accuracy			**84%**
Total kappa			**0.76**

(a) *Error matrix (%),* ***(b)*** *Accuracy assessment (%). Conditional Kappa calculated as per Congalton and Green (2009).*

DISCUSSION

Our results suggest that semi-automated classification using object-based image analysis (OBIA) is a useful tool for quantifying percent cover estimates of vegetation in sites where fieldwork can be difficult, particularly for sensitive ground cover vegetation at the small scale where either consistent photographs have been taken over time or marked photopoints are available, so that new images can be obtained. The results confirm that semi-automated OBIA estimates of vegetation percent cover are within the range of visual estimation of cover by experts, and OBIA vegetation cover estimates can be used to objectively assess moss health changes over time.

Comparing the accuracy of methods, semi-automated classification and manual classification reference data had a thematic accuracy of 84% and a kappa value of 0.76. The thematic accuracy here is a conservative estimate, as thematic errors determined on a pixel-by-pixel basis may be present but still have little effect on overall percent cover estimates. For visual estimates, the experts, who were all familiar with the Windmill Islands vegetation and its various health states, all varied in their vegetation percent cover estimates. In general, between observer variability is known to be quite high for visual estimates of percent cover, regardless of the ecosystem being studied (Elzinga et al., 1998; Van Coillie et al., 2014). The semi-automated OBIA percent cover estimates fell within the range of expert estimates for the majority of classes across all three quadrats, with the only differences being significantly less shadow and significantly less stressed moss detected by the OBIA when it was present in less than 15% of a quadrat. Semi-automated classification significantly strengthens the analysis, as the differences in shadow estimates were due to the ability of the object-based image analysis software to use the red and green pixel ratios to assess whether a shadow was likely to contain healthy, stressed or moribund moss in areas of the image that were too dark for the human observers to judge in photographs. Some shadows remained too dark for determination of vegetation health within the shadow, with this typically occurring in frost-heave crevices. Even by eye in the field it is very difficult to identify with any level of accuracy the composition of vegetation within these crevices. Some issues with shadow in images may be reduced by further standardizing lighting conditions in the field (e.g., only take photos when cloudy), however, this is impractical to achieve in this remote environment, where field time is limited and opportunistic. The differences in stressed moss appeared to occur in areas where stressed moss mixed with healthy moss. In this case, there was sufficient stressed moss for human observers to determine a certain percentage stressed in a grid estimate, but not enough to override the predominance of healthy moss within an object in the object-based image analysis.

Standard field methods of percent cover estimation can overcome problems related to lighting conditions. In the field, it is possible to change viewing position to be able to see areas that may be obscured or difficult to see in a photo, and it is also

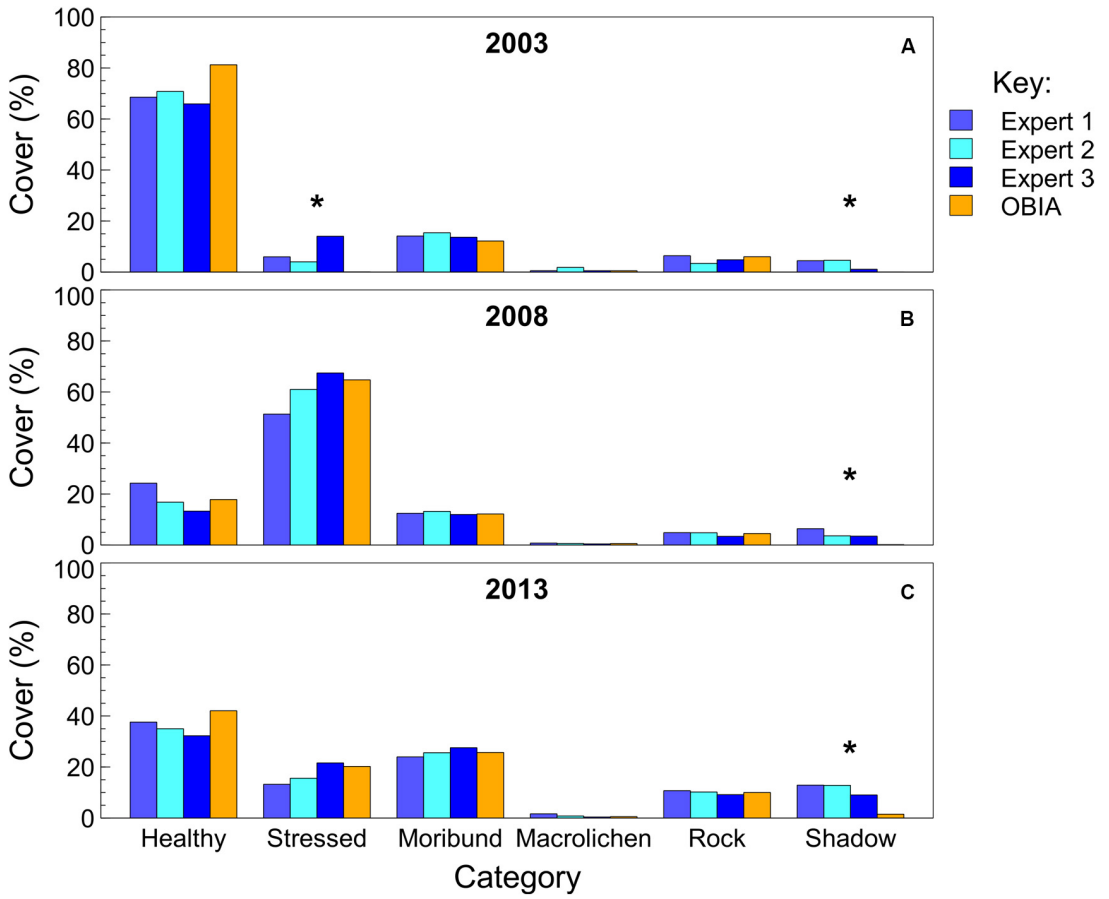


FIGURE 5 | Comparison of percent cover estimates from three experts and the semi-automated object-based image analysis (OBIA). Quadrat images assessed were the bottom row of the mosaic in **Figure 1**, from **(A)** 2003, **(B)** 2008, and **(C)** 2013. Asterisks indicate significant differences between observers within each class in each year.

possible to add extra light if needed in order to see into shadowy areas of a quadrat. However, the estimate of vegetation percent cover is subjective among human observers, and the accuracy of observers' estimates is rarely estimated (Booth et al., 2005a; Gorrod and Keith, 2009). Field estimates of percent cover can be affected by adverse environmental conditions, such as the extreme cold in Antarctica, and the longer a researcher spends in the field, the less accurate observations may be, due to observer fatigue, the effects of changes in weather, time constraints and time of day (Bennett et al., 2000; Gorrod and Keith, 2009). Digital photography of quadrats required only one third of the field time required for traditional estimates of vegetation percent cover in the field. Due to the proposed length of this long-term monitoring study (25 years, with sampling originally planned at 5 year intervals), it is also likely that different researchers would perform the percent cover estimates each sampling period, thus each season would have a different observer bias. Thus, while our semi-automatic classification has only been assessed against expert manual classification of the same images and not field based estimates, it is likely that field based classification would be more variable and hence less reliable.

Based on the expert assessments in **Figure 5**, the estimation of healthy versus stressed moss cover appears to be the most subjective between observers. As this study involves classification using vegetation color to determine vegetation category (healthy = green, stressed = red and moribund = white/brown/black), it is important to note that color perception varies between observers, particularly between genders and different age groups, and this could affect vegetation percent cover assessments (Booth et al., 2005b; Abramov et al., 2012). Color perception is also highly dependent upon the variance and mean of surrounding colors (Brown and MacLeod, 1997), thus the assessment of whether or not a particular area of vegetation is healthy, stressed or moribund can also change for a single observer. It is difficult to maintain the same vegetation health categorization when the dominant color changes between quadrats, as well as being difficult to determine boundaries between categories when the vegetation varies along a health/color gradient. Using semi-automated classification based on RGB digital values helps to prevent these problems, as the classification involves a rule set using algorithms to make objective assessments of vegetation health,

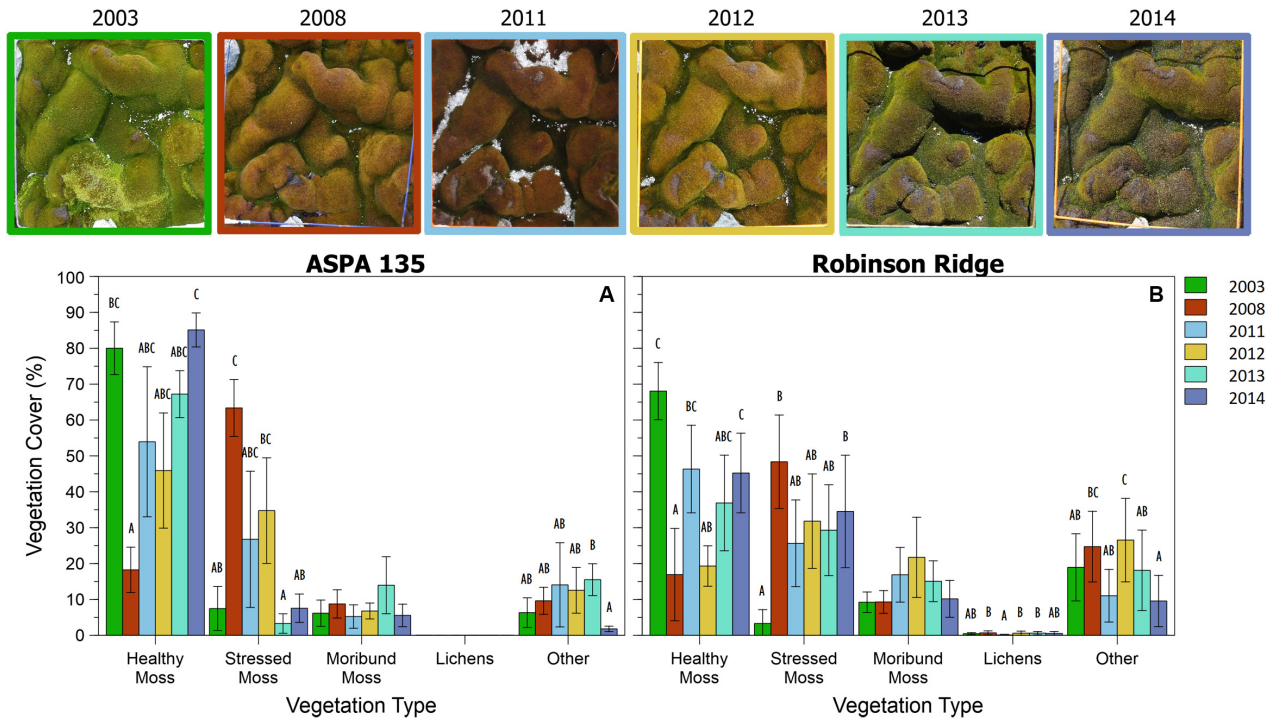


FIGURE 6 | Trends in Bryophyte community vegetation cover from 2003 to 2014 at two sites in the Windmill Islands, A2 **(A)** and RR **(B)**. Photographs depict a representative quadrat (site A2, quadrat B3) at each of the six time points. Data are mean percent cover of vegetation ± 95% CI. Different letters denote significant differences between years per vegetation type.

and the accuracy is sufficient to assess vegetation health changes over time.

To ensure long-term monitoring was able to include all of the digital photographs acquired in earlier seasons, it was imperative that the methodology developed be as robust as possible in order to be able to analyze digital RGB photographs from different years taken in different weather and acquisition conditions and at different resolutions. To achieve a conservative estimate of change over time, the rule set was designed to prevent overestimates of stressed and moribund moss, thus causing some errors with misclassification of stressed and moribund moss as healthy. Reference data introduces additional error and uncertainty in vegetation studies, thus is never perfect (Congalton and Green, 2009), and some error in the accuracy assessment was expected, due to the difficulty in defining the boundaries between vegetation health categories for such small and thin individual moss shoots. It can be difficult to define boundaries in nature and the vegetation health categories are usually found as a gradient between healthy and stressed, stressed and moribund or even healthy and moribund. The use of manual digitization as reference data is also only an approximation, due to the limitations of delineation using geographical information systems (GIS), so some geometric errors are inevitable and subject to edge effects (Bennett et al., 2000; Fensham and Fairfax, 2002; Aksoy et al., 2010). We attempted to increase the accuracy of the manual reference data by having multiple experts digitize the same portion (two quadrat images) of the mosaic to obtain a majority agreement, however, manual digitization is time consuming and due to time constraints it was not possible to have more than one researcher digitize the entire mosaic. The highest percent cover error in the error matrix was 5.16% of manually classified moribund moss being classified as healthy moss in the semi-automated classification, and this is likely due to areas of mixed vegetation health, where healthy moss shoots grow up through areas of moribund moss. It could also have been caused by the similarity in colors between some light-colored areas of moribund moss and certain areas of healthy moss which were moist enough to reflect sunlight, thus appearing light in color.

Once the methodology protocol was developed, it was applied to the full data set to assess vegetation health change over time (**Figures 6**, **7**). Moss health state changes were observed between healthy, stressed and moribund over the 11 years of monitoring at ASPA 135 and Robinson Ridge (**Figure 8**). As described by Robinson et al. (2018) and shown here, both sites experienced a marked increase in stress and decline in health in 2008, in both the Bryophyte and Transitional communities. This was likely due to an unusual occurrence of freezing rain in that season (Robinson et al., 2018). The significant shift in moss color from green (healthy) in 2003 to red (stressed) in 2008 (**Figures 6**, **7**) was followed by recovery of health by 2013–2014, although associated with increased stressed moss in the Robinson Ridge Bryophyte community (**Figure 6B**), and increased moribund moss in the Robinson Ridge Transitional community (**Figure 7B**). Our methodology using handheld

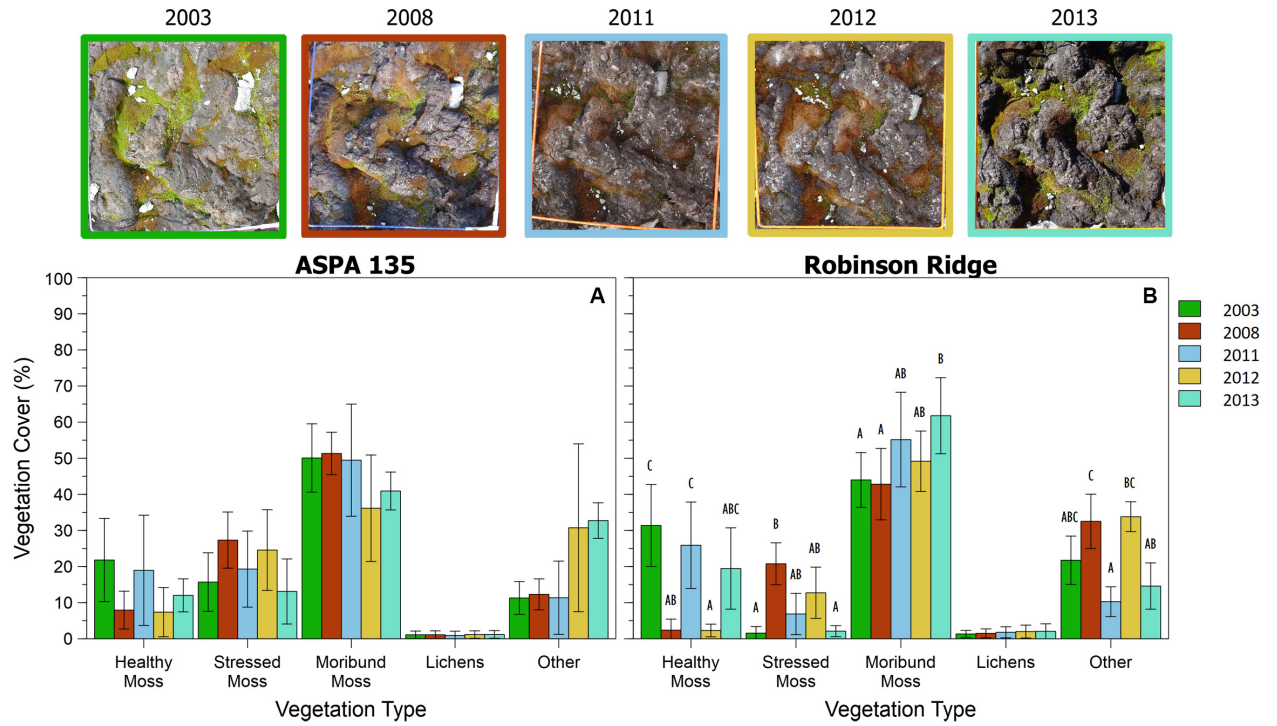


FIGURE 7 | Trends in Transitional community vegetation cover from 2003 to 2013 at two sites in the Windmill Islands, A2 **(A)** and RR **(B)**. Photographs depict a representative quadrat (site A2, quadrat T2) at each of the five time points. NB: Transitional communities were not monitored in 2014. Data are mean percent cover of vegetation ± 95% CI. Different letters denote significant differences between years per vegetation type.

digital photographs allows assessment of short and long term moss stress at a community level, prior to species-level changes in response to such stress (Robinson et al., 2018). We have seen resilience in these communities, as they have used changes in their protective pigments (becoming red in color) to respond to environmental stressors and subsequently "regreened" when conditions became more favorable. Despite this resilience, we have seen a remarkable increase in the submergence-intolerant species *Ceratodon purpureus* (Wasley et al., 2006, 2012; Robinson et al., 2018) and increased moribund moss.

Long-term monitoring coupled with image analysis improves our understanding of Antarctic vegetation health by underpinning a conceptual framework for how moss changes with stress and recovery (**Figure 8**). Antarctic moss health changes rapidly between stressed and healthy under ideal laboratory conditions, dependent on water availability, but further research is required to determine how long it takes for these changes under field conditions (Malenovský et al., 2015; Waterman et al., 2018). The color changes are caused by changes in the types of protective pigments (including anthocyanins and carotenoids) produced in the leaves of the moss (Post, 1990; Lovelock and Robinson, 2002; Waterman et al., 2018). ASPA mosses suffered stress resulting in pigment changes from green to red between 2003 and 2008, but this was reversed by subsequent loss of red pigments (red arrow) or growth of new leaves (green arrows). Both Robinson Ridge communities experienced the same stress response in 2008, however, the bryophyte community only partially recovered by 2014. In contrast, the transitional community demonstrates how mosses transition from green to red and finally moribund (red and blue arrows, **Figure 8**). These physiological, individual plant level changes occurred alongside species level changes, with an increase in the desiccation-tolerant mosses *Ceratodon purpureus* and *Bryum pseudotriquetrum* and a decline in the endemic *Schistidium antarctici* (Robinson et al., 2018). This appears to be linked to climatic changes such as regional drying and fewer days above zero degrees, in response to changes in the Southern Annular Mode (SAM) (Robinson et al., 2018). Further assessment of these changes through ongoing long-term monitoring will be vital for future management and conservation of these ecosystems (Bergstrom, 2017). There are, however, indications that moribund moss can regreen within 1 month, under the extreme flooding caused by recent Antarctic heatwaves (Robinson et al., 2020).

There are significant advantages of using semi-automated classification for long-term vegetation monitoring projects. The classification of vegetation health is objective and consistent, and each sampling season has the same rules, so all quadrats are assessed in the same way and valid comparisons can be made between photographs taken in different sampling periods over time, whether days, weeks or years apart. Digital photography of the vegetation quadrats means much less time needs to be spent in the field and all analyses can be completed in the comfort and safety of the laboratory (Bennett et al., 2000; Booth et al., 2005a). This is particularly useful in areas where fieldwork is

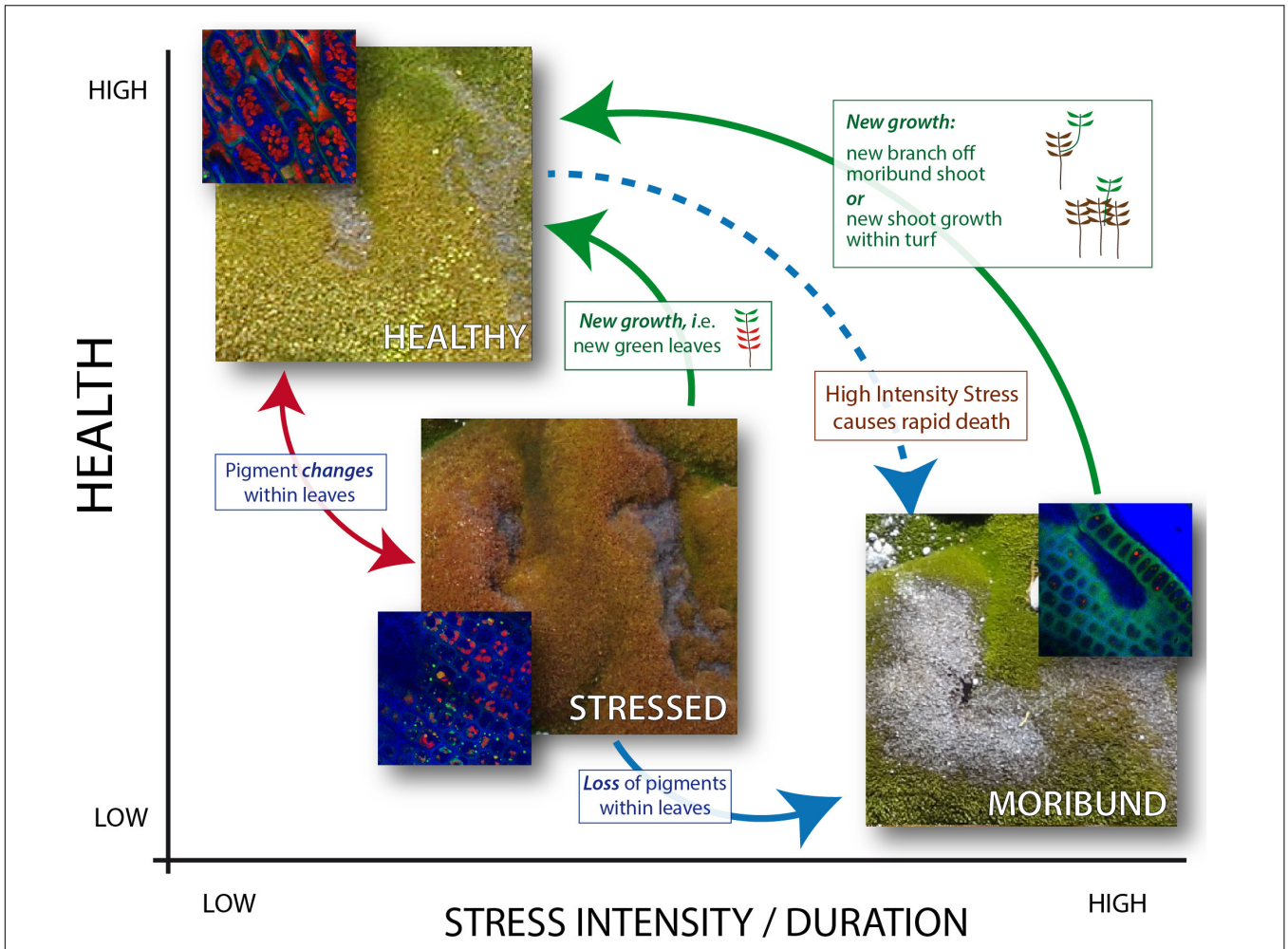


FIGURE 8 | Diagrammatic representation of East Antarctic moss health state relationships. The intensity or duration of stress (*x* axis) impacts the health state (*y* axis). Solid arrows indicate observed heath state changes (2003–2014) of monitoring moss health in the Windmill Islands, with red and blue arrows indicating pigment changes within the cells of an individual plant, while green arrows indicate growth of new healthy leaves/plants at a turf level. Note that the three images show the same area of moss in different years, and that the moribund moss is the gray moss in the lower right image. Inset images (blue) depict confocal microscope images of cells at each stage of moss health, with bright red chloroplasts obvious in healthy moss but less clear in stressed moss and absent in moribund moss. Extreme high intensity stress is thought to cause immediate state change from healthy to moribund (dashed blue arrow), however, evidence for this is only anecdotal.

difficult, dangerous or expensive. Digital photography of each quadrat can be accomplished in less than 5 min in the field, whereas field estimation of percent cover can take up to 15 min per quadrat. This reduces the time in the field for Antarctic vegetation monitoring from 15 to 5 h, which means that a single day per field site is feasible to complete all the required sampling each season. The digital photographs are easily stored, and are thus available for future retrospective studies examining new questions or using improved methods (Bennett et al., 2000; Booth et al., 2005a). For a long-term monitoring project, the rule development can take some time initially, as the rules used must be modified to suit that particular ecosystem. However, once a rule set is established with good accuracy for that ecosystem, this method is easily repeatable and consistent, and subsequent pre-processing and analysis of photographs is quite fast. Pre- and post-processing of the photographs does, however, require a trained expert in the field in order to accurately assess the vegetation types in the photographs, particularly when moss is wet. If quadrats are flooded with water or covered by snow, vegetation classification is difficult, whether using traditional field-based methods or photographs. Once the methodology protocol has been established, the combined field and lab work of OBIA can be as quick as, or quicker than, field estimations of percent cover, and is quicker than manual digitization, while reducing the effects of observer bias and allowing spatial and temporal comparisons of quadrats.

Recommendations to improve the accuracy include using a multispectral or hyperspectral camera that includes a near infrared band, enabling the use of the normalized difference vegetation index (NDVI), which would enable analysis of the Lichen community, as well as significantly reduce pre-processing time by allowing automated removal of non-vegetated areas of

the image and better classification of vegetation health (Stow et al., 2004; Torres-Sánchez et al., 2015; Lawley et al., 2016). However, this would not allow comparisons with the earlier data collected in the monitoring study from 2003 to 2014. Accuracy could also be improved by standardizing the conditions for photography, for example ensuring that photos are not taken in direct sunlight, in order to reduce shadows and reflections within quadrat photographs, although time constraints with Antarctic fieldwork mean this is not always possible, and including color references in each image for color calibration (Kolyaie et al., 2019). The use of fuzzy classification methods in addition to rule-based methods may also increase the classification accuracy, although it would increase the time required for analysis as samples would have to be selected for each class in each image (Laliberte et al., 2007b). The use of multispectral very high resolution (VHR) satellite imagery has recently enabled the development of semi-automated satellite-based remote sensing methods of assessing Antarctic vegetation distribution (Jawak et al., 2015). However, those methods cannot yet distinguish the difference between vegetation types (such as moss, lichens, algae or cyanobacteria), and cannot determine vegetation distribution, both of which can be achieved using handheld digital photography. Using unmanned aerial systems (UAS's) equipped with multispectral sensors is another good option for assessing vegetation distributions and health in the field, with minimal disturbance to the vegetation and the ability to monitor entire sites rather than just single quadrats, although ground observations are still required for data validation (Malenovský et al., 2017; Turner et al., 2018). It is likely that UASs will be increasingly utilized for vegetation monitoring in future, particularly in remote regions.

Here, we have shown that the use of object-based image analysis to classify digital photographs of quadrats provides an objective, repeatable, robust and fast method of assessing vegetation health over time. This method is particularly suited to remote, difficult to access locations with short stature vegetation communities, where field work and disturbance must be minimized, such as in polar, alpine and desert ecosystems. By analyzing long-term vegetation monitoring datasets with image analysis, we have informed a conceptual model for healthy, stressed and moribund moss in response to environmental stress and recovery. In East Antarctica it reveals that whilst moss health changes can be quite dynamic, a long-term decline in health is also apparent.

DATA AVAILABILITY STATEMENT

All data associated with this article is archived with the Australian Antarctic Data Centre (https://data.aad.gov.au/aadc/, doi: 10.26179/5eaf718e45748, doi: 10.26179/5eaf945e9b1ef, and doi: 10.4225/15/59fbbb909cf71).

AUTHOR CONTRIBUTIONS

JW and SR conceived the study. SR, JW, ER-C, and AL performed the field work. ER-C conducted the pilot study. DK designed the methods and analyzed the data. MA, AL, LC, and SR assisted with the method development. MA assisted with the data analysis. DK wrote the manuscript. All authors contributed to editing the manuscript.

FUNDING

Funding was provided by the Australian Research Council DP110101714 and DP180100113, Antarctic Science Grants 2780, 1313, 3042, 3129, 3130, and 4046. We acknowledge financial support from the University of Wollongong's Global Challenges Program as part of the Sustaining Coastal and Marine Zones challenge. DK was supported by an Australian Postgraduate Award/Research Training Program Scholarship.

ACKNOWLEDGMENTS

The authors wish to thank Dana Bergstrom, Jessica Bramley-Alves, Johanna D. Turnbull, Rebecca E. Miller, Zbyněk Malenovský, Anna Nydahl, and other ANARE expeditioners for assistance in the field. Taylor Benny for lab assistance. Heidi Brown for technical support; and Andrew Netherwood for production of **Figure 8**.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00766/full#supplementary-material

REFERENCES

Abramov, I., Gordon, J., Feldman, O., and Chavarga, A. (2012). Sex and vision II: color appearance of monochromatic lights. *Biol. Sex Differ.* 3, 1–16. doi: 10.1186/2042-6410-3-21

Aksoy, S., Akçay, H. G., and Wassenaar, T. (2010). Automatic mapping of linear woody vegetation features in agricultural landscapes using very high resolution imagery. *IEEE Trans. Geosci. Remote Sens.* 48, 511–522. doi: 10.1109/TGRS.2009.2027702

Barrett, J. E., Virginia, R. A., Wall, D. H., Doran, P. T., Fountain, A. G., Welch, K. A., et al. (2008). Persistent effects of a discrete warming event on a polar desert ecosystem. *Glob. Chang. Biol.* 14, 2249–2261. doi: 10.1111/j.1365-2486.2008.01641.x

Bennett, L. T., Judd, T. S., and Adams, M. A. (2000). Close-range vertical photography for measuring cover changes in perennial grasslands. *J. Range Manag.* 53, 634–641. doi: 10.2458/azu_jrm_v53i6_bennett

Berberoglu, S., Akin, A., Atkinson, P. M., and Curran, P. J. (2010). Utilizing image texture to detect land-cover change in Mediterranean coastal wetlands. *Int. J. Remote Sens.* 31, 2793–2815. doi: 10.1080/01431160903111077

Bergstrom, D. M. (2017). Ecosystem shift after a hot event. *Nat. Ecol. Evol.* 1, 1226–1227. doi: 10.1038/s41559-017-0262-z

Blaschke, T. (2004). "Object-based contextual image classification built on image segmentation," in *Proceedings of the IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*, (Greenbelt, MD: IEEE), 113–119. doi: 10.1109/WARSD.2003.1295182

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65, 2–16. doi: 10.1016/j.isprsjprs.2009.06.004

Blaschke, T., and Strobl, J. (2001). What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *Geo Inform. Syst.* 14, 12–17. doi: 10.1364/AO.52.007629

Bollard-Breen, B., Brooks, J. D., Jones, M. R. L., Robertson, J., Betschart, S., Kung, O., et al. (2015). Application of an unmanned aerial vehicle in spatial mapping of terrestrial biology and human disturbance in the McMurdo Dry Valleys, East Antarctica. *Polar Biol.* 38, 573–578. doi: 10.1007/s00300-014-1586-7

Booth, D. T., Cox, S. E., and Johnson, D. E. (2005b). Detection-threshold calibration and other factors influencing digital measurements of ground cover. *Rangel. Ecol. Manag.* 58, 598–604. doi: 10.2111/05-060R1.1

Booth, D. T., Cox, S. E., Fifield, C., Phillips, M., and Williamson, N. (2005a). Image analysis compared with other methods for measuring ground cover. *Arid. L. Res. Manag.* 19, 91–100. doi: 10.1080/15324980590916486

Brown, R. O., and MacLeod, D. I. A. (1997). Color appearance depends on the variance of surround colors. *Curr. Biol.* 7, 844–849. doi: 10.1016/S0960-9822(06)00372-1

Bunting, P., and Lucas, R. (2006). The delineation of tree crowns in Australian mixed species forests using hyperspectral compact airborne spectrographic imager (CASI) data. *Remote Sens. Environ.* 101, 230–248. doi: 10.1016/j.rse.2005.12.015

Chen, Z., Chen, W., Leblanc, S. G., and Henry, G. H. R. (2010). Digital photograph analysis for measuring percent plant cover in the Arctic. *Arctic* 63, 315–326. doi: 10.14430/arctic1495

Conchedda, G., Durieux, L., and Mayaux, P. (2007). "Object-based monitoring of land cover changes in mangrove ecosystems of Senegal," in *Proceedings of the 2007 International Workshop on the Analysis of Multi-Temporal Remote Sensing Images*, (Leuven: IEEE), 1–6. doi: 10.1109/MULTITEMP.2007.4293039

Congalton, R. G., and Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL: CRC Press.

Cserhalmi, D., Nagy, J., Kristóf, D., and Neidert, D. (2011). Changes in a wetland ecosystem: a vegetation reconstruction study based on historical panchromatic aerial photographs and succession patterns. *Folia Geobot.* 46, 351–371. doi: 10.1007/s12224-011-9099-4

Drăguţ, L., Tiede, D., and Levick, S. R. (2010). ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *Int. J. Geogr. Inf. Sci.* 24, 859–871. doi: 10.1080/13658810903174803

Dunn, J. L., and Robinson, S. A. (2006). Ultraviolet B screening potential is higher in two cosmopolitan moss species than in a co-occurring Antarctic endemic moss: implications of continuing ozone depletion. *Glob. Chang. Biol.* 12, 2282–2296. doi: 10.1111/j.1365-2486.2006.01283.x

Ehlers, M., Gaehler, M., and Janowsky, R. (2006). Automated techniques for environmental monitoring and change analyses for ultra high resolution remote sensing data. *Photogramm. Eng. Remote Sens.* 72, 835–844. doi: 10.14358/PERS.72.7.835

Elzinga, C., Salzer, D., and Willoughby, J. (1998). "Measuring & monitering plant populations," in *U.S. Bureau of Land Management Papers*, 17. Available online at: http://digitalcommons.unl.edu/usblmpub/17 (accessed April 20, 2020).

Fensham, R. J., and Fairfax, R. J. (2002). Aerial photography for assessing vegetation change: a review of applications and the relevance of findings for Australian vegetation history. *Aust. J. Bot.* 50, 415–429. doi: 10.1071/BT01032

Gorrod, E. J., and Keith, D. A. (2009). Observer variation in field assessments of vegetation condition: implications for biodiversity conservation. *Ecol. Manag. Restor.* 10, 31–40. doi: 10.1111/j.1442-8903.2009.00437.x

Greenwood, D. L., and Weisberg, P. J. (2009). GIS-based modeling of pinyon-juniper woodland structure in the great basin. *For. Sci.* 55, 1–12. doi: 10.17221/96/2008-jfs

Hájek, F. (2008). Process-based approach to automated classification of forest structures using medium format digital aerial photos and ancillary GIS information. *Eur. J. For. Res.* 127, 115–124. doi: 10.1007/s10342-007-0188-0

Hall, F. C. (2002). *Photo Point Monitoring Handbook: Part B-Concepts and Analysis. Portland, OR*. Available online at: http://npshistory.com/publications/interdisciplinary/im/pnw-gtr526b.pdf (accessed May 2, 2020).

Hay, G. J., Castilla, G., Wulder, M. A., and Ruiz, J. R. (2005). An automated object-based approach for the multiscale image segmentation of forest scenes. *Int. J. Appl. Earth Obs. Geoinf.* 7, 339–359. doi: 10.1016/j.jag.2005.06.005

Haywood, A., and Stone, C. (2011). Semi-automating the stand delineation process in mapping natural eucalypt forests. *Aust. For.* 74, 13–22. doi: 10.1080/00049158.2011.10676341

Howard-Williams, C., Peterson, D., Lyons, W. B., Cattaneo-Vietti, R., and Gordon, S. (2006). Measuring ecosystem response in a rapidly changing environment: the latitudinal gradient project. *Antarct. Sci.* 18, 465–471. doi: 10.1017/S0954102006000514

IPCC (2014). "Climate change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects," in *Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds C. B. Field, V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, T. E. Bilir, et al. (Cambridge: Cambridge University Press).

Jawak, S. D., Raut, D. A., and Luis, A. J. (2015). Iterative spectral index ratio exploration for object-based image analysis of Antarctic coastal oasis using high resolution satellite remote sensing data. *Aquat. Procedia* 4, 157–164. doi: 10.1016/j.aqpro.2015.02.022

Jensen, J. R. (2013). *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd Edn. Harlow: Pearson Education Limited.

Jobin, B., Labrecque, S., Grenier, M., and Falardeau, G. (2008). Object-based classification as an alternative approach to the traditional pixel-based classification to identify potential habitat of the Grasshopper Sparrow. *Environ. Manage.* 41, 20–31. doi: 10.1007/s00267-007-9031-0

Johansson, P., and Thor, G. (2008). Lichen species density and abundance over ten years in permanent plots in inland Dronning Maud Land, Antarctica. *Antarct. Sci.* 20, 115–121. doi: 10.1017/S0954102007000855

Kim, M., Warner, T. A., Madden, M., and Atkinson, D. S. (2011). Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: Scale, texture and image objects. *Int. J. Remote Sens.* 32, 2825–2850. doi: 10.1080/01431161003745608

Ko, D., Bristow, N., Greenwood, D., and Weisberg, P. (2009). Canopy cover estimation in semiarid woodlands: comparison of field-based and remote sensing methods. *For. Sci.* 55, 132–141. doi: 10.1093/forestscience/55.2.132

Kolyaie, S., Treier, U. A., Watmough, G. R., Madsen, B., Bøcher, P. K., Psomas, A., et al. (2019). Transferability and the effect of colour calibration during multi-image classification of Arctic vegetation change. *Polar Biol.* 42, 1227–1239. doi: 10.1007/s00300-019-02491-7

Laliberte, A. S., and Rango, A. (2011). Image processing and classification procedures for analysis of sub-decimeter imagery acquired with an unmanned aircraft over arid rangelands. *GIScience Remote Sens.* 48, 4–23. doi: 10.2747/1548-1603.48.1.4

Laliberte, A. S., Browningb, D. M., Herrickb, J. E., and Gronemeyera, P. (2010). Hierarchical object-based classification of ultra-high-resolution digital mapping camera (DMC) imagery for rangeland mapping and assessment. *J. Spat. Sci.* 55, 101–115. doi: 10.1080/14498596.2010.487853

Laliberte, A. S., Fredrickson, E. L., and Rango, A. (2007a). Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogramm. Eng. Remote Sensing* 73, 197–207. doi: 10.14358/PERS.73.2.197

Laliberte, A. S., Rango, A., Herrick, J. E., Fredrickson, E. L., and Burkett, L. (2007b). An object-based image analysis approach for determining fractional cover of senescent and green vegetation with digital plot photography. *J. Arid Environ.* 69, 1–14. doi: 10.1016/j.jaridenv.2006.08.016

Lathrop, R. G., Montesano, P., and Haag, S. (2006). A multi-scale segmentation approach to mapping seagrass habitats using airborne digital camera imagery. *Photogramm. Eng. Remote Sens.* 72, 665–675. doi: 10.14358/PERS.72.6.665

Lawley, V., Lewis, M., Clarke, K., and Ostendorf, B. (2016). Site-based and remote sensing methods for monitoring indicators of vegetation condition: an Australian review. *Ecol. Indic.* 60, 1273–1283. doi: 10.1016/j.ecolind.2015.03.021

Lee, J. R., Raymond, B., Bracegirdle, T. J., Chadès, I., Fuller, R. A., Shaw, J. D., et al. (2017). Climate change drives expansion of Antarctic ice-free habitat. *Nature* 547, 49–54. doi: 10.1038/nature22996

Liu, D., and Xia, F. (2010). Assessing object-based classification: advantages and limitations. *Remote Sens. Lett.* 1, 187–194. doi: 10.1080/01431161003743173

Lovelock, C. E., and Robinson, S. A. (2002). Surface reflectance properties of antarctic moss and their relationship to plant species, pigment composition and photosynthetic function. *Plant Cell Environ.* 25, 1239–1250. doi: 10.1046/j.1365-3040.2002.00916.x

Luscier, J. D., Thompson, W. L., Wilson, J. M., Gorham, B. E., and Dragut, L. D. (2006). Using digital photographs and object—based image analysis to estimate percent ground cover in vegetation plots. *Front. Ecol. Environ.* 4:408. doi: 10.1890/1540-929520064[408:UDPAOI]2.0.CO;2

Malenovský, Z., Lucieer, A., King, D. H., Turnbull, J. D., and Robinson, S. A. (2017). Unmanned aircraft system advances health mapping of fragile polar vegetation. *Methods Ecol. Evol.* 8, 1842–1857. doi: 10.1111/2041-210X.12833

Malenovský, Z., Turnbull, J. D., Lucieer, A., and Robinson, S. A. (2015). Antarctic moss stress assessment based on chlorophyll content and leaf density retrieved from imaging spectroscopy data. *New Phytol.* 208, 608–624. doi: 10.1111/nph.13524

McCarthy, D. P., and Zaniewski, K. (2001). Digital analysis of lichen cover: a technique for use in lichenometry and licnenology. *Arctic Antarct. Alp. Res.* 33, 107–113. doi: 10.1080/15230430.2001.12003411

Michaels, A., and Power, A. G. (2011). *Long-Term Ecological Research Program. A Report of the 30 Year Review Committee*. Available online at: papers3://publication/uuid/1898E571-88FC-4FC1-A3A7-48AAC2430420 (accessed April 20, 2020).

Michel, P., Mathieu, R., and Mark, A. F. (2010). Spatial analysis of oblique photo-point images for quantifying spatio-temporal changes in plant communities. *Appl. Veg. Sci.* 13, 173–182. doi: 10.1111/j.1654-109X.2009.01059.x

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., et al. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* 355:eaai9214. doi: 10.1126/science.aai9214

Post, A. (1990). Photoprotective pigment as an adaptive strategy in the antarctic moss Ceratodon purpureus. *Polar Biol.* 10, 241–245. doi: 10.1007/BF00238420

Pringle, R. M., Syfert, M., Webb, J. K., and Shine, R. (2009). Quantifying historical changes in habitat availability for endangered species: Use of pixel- and object-based remote sensing. *J. Appl. Ecol.* 46, 544–553. doi: 10.1111/j.1365-2664.2009.01637.x

Robinson, S. A., King, D. H., Bramley-Alves, J., Waterman, M. J., Ashcroft, M. B., Wasley, J., et al. (2018). Rapid change in East Antarctic terrestrial vegetation in response to regional drying. *Nat. Clim. Chang.* 8, 879–884. doi: 10.1038/s41558-018-0280-0

Robinson, S. A., Klekociuk, A. R., King, D. H., Pizarro Rojas, M., Zúñiga, G. E., and Bergstrom, D. M. (2020). The 2019/2020 summer of Antarctic heatwaves. *Glob. Chang. Biol.* doi: 10.1111/gcb.15083 [Epub ahead of print].

Rogers, G. F., Turner, R. M., and Malde, H. E. (1983). "Using matched photographs to monitor resource change," in *Proceedings, International Conference Renewable Resource Inventories for Monitoring Changes and Trend*, eds J. F. Bell and T. Atterbury (Corvallis, OR: College of Forestry), 90–92.

Ryan-Colton, E. (2007). *Long-Term Monitoring of the Impacts of Climate Change on Antarctic Terrestrial Communities: Baseline and Method Developments*. Wollongong NSW: University of Wollongong.

Stow, D. A., Hope, A., McGuire, D., Verbyla, D., Gamon, J., Huemmrich, F., et al. (2004). Remote sensing of vegetation and land-cover change in Arctic Tundra Ecosystems. *Remote Sens. Environ.* 89, 281–308. doi: 10.1016/j.rse.2003.10.018

Torres-Sánchez, J., López-Granados, F., and Peña, J. M. (2015). An automatic object-based method for optimal thresholding in UAV images: application for vegetation detection in herbaceous crops. *Comput. Electron. Agric.* 114, 43–52. doi: 10.1016/j.compag.2015.03.019

Trussell, H. J., and Vrhel, M. J. (2008). *Fundamentals of Digital Imaging*. Cambridge: Cambridge University Press.

Turner, D., Lucieer, A., and Watson, C. (2012). An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sens.* 4, 1392–1410. doi: 10.3390/rs4051392

Turner, D., Lucieer, A., Malenovský, Z., King, D. H., and Robinson, S. A. (2014). Spatial co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over Antarctic Moss Beds. *Remote Sens.* 6, 4003–4024. doi: 10.3390/rs6054003

Turner, D., Lucieer, A., Malenovský, Z., King, D. H., and Robinson, S. A. (2018). Assessment of Antarctic moss health from multi-sensor UAS imagery with random forest modelling. *Int. J. Appl. Earth Obs. Geoinf.* 68, 168–179. doi: 10.1016/j.jag.2018.01.004

Van Coillie, F. M. B., Gardin, S., Anseel, F., Duyck, W., Verbeke, L. P. C., and De Wulf, R. R. (2014). Variability of operator performance in remote-sensing image interpretation: the importance of human and external factors. *Int. J. Remote Sens.* 35, 754–778. doi: 10.1080/01431161.2013.873152

Verbyla, D. L. (2002). *Practical GIS Analysis*. Milton Park: Taylor & Francis.

Wasley, J., Robinson, S. A., Lovelock, C. E., and Popp, M. (2006). Some like it wet – biological characteristics underpinning tolerance of extreme water stress events in Antarctic bryophytes. *Funct. Plant Biol.* 33, 443–455. doi: 10.1071/FP05306

Wasley, J., Robinson, S. A., Turnbull, J. D., King, D. H., Wanek, W., and Popp, M. (2012). Bryophyte species composition over moisture gradients in the Windmill Islands, East Antarctica: development of a baseline for monitoring climate change impacts. *Biodiversity* 13, 257–264. doi: 10.1080/14888386.2012.712636

Waterman, M. J., Bramley-Alves, J., Miller, R. E., Keller, P. A., and Robinson, S. A. (2018). Photoprotection enhanced by red cell wall pigments in three East Antarctic mosses. *Biol. Res.* 51, 13–49. doi: 10.1186/s40659-018-0196-1

Whiteside, T. G., Boggs, G. S., and Maier, S. W. (2011). Comparing object-based and pixel-based classifications for mapping savannas. *Int. J. Appl. Earth Obs. Geoinf.* 13, 884–893. doi: 10.1016/j.jag.2011.06.008

Yu, Q., Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., et al. (2006). Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogramm. Eng. Remote Sens.* 72, 799–811.

Zhang, L., Li, L. J., Liang, L. Q., and Li, J. Y. (2007). Monitoring of vegetation coverage based on high-resolution images. *For. Stud. China* 9, 256–261. doi: 10.1007/s11632-007-0040-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 June 2020
doi: 10.3389/fpls.2020.00747

Transcriptional Landscapes of Divergent Sporophyte Development in Two Mosses, *Physcomitrium (Physcomitrella) patens* and *Funaria hygrometrica*

*Alexander Kirbis[1], Manuel Waller[1], Mariana Ricca[2], Zoe Bont[3], Anna Neubauer[1], Bernard Goffinet[4] and Péter Szövényi[1]**

[1] Department of Systematic and Evolutionary Botany, University of Zurich, Zurich and Zurich-Basel Plant Science Center, Zurich, Switzerland, [2] Department for BioMedical Research (DBMR), University of Bern, Bern, Switzerland, [3] Institute of Plant Sciences, University of Bern, Bern, Switzerland, [4] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, United States

OPEN ACCESS

Edited by:
Stefan de Folter,
Centro de Investigación y Estudios Avanzados, Instituto Politécnico Nacional de México (CINVESTAV), Mexico

Reviewed by:
Lydia Gramzow,
Friedrich Schiller University Jena, Germany
Caspar Christian Cedric Chater,
The University of Sheffield, United Kingdom

***Correspondence:**
Péter Szövényi
peter.szoevenyi@systbot.uzh.ch

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: 19 February 2020
Accepted: 11 May 2020
Published: 10 June 2020

Citation:
Kirbis A, Waller M, Ricca M, Bont Z, Neubauer A, Goffinet B and Szövényi P (2020) Transcriptional Landscapes of Divergent Sporophyte Development in Two Mosses, Physcomitrium (Physcomitrella) patens and Funaria hygrometrica. Front. Plant Sci. 11:747. doi: 10.3389/fpls.2020.00747

Understanding the molecular basis of morphological shifts is a fundamental question of evolutionary biology. New morphologies may arise through the birth/death of genes (gene gain/loss) or by reutilizing existing gene sets. Yet, the relative contribution of these two processes to radical morphological shifts is still poorly understood. Here, we use the model system of two mosses, *Funaria hygrometrica* and *Physcomitrium (Physcomitrella) patens*, to investigate the molecular mechanisms underlying contrasting sporophyte architectures. We used comparative analysis of time-series expression data for four stages of sporophyte development in both species to address this question in detail. We found that large-scale differences in sporophytic architecture are mainly governed by orthologous (i.e., shared) genes frequently experiencing temporal gene expression shifts between the two species. While the absolute number of species-specific genes expressed during sporophyte development is somewhat smaller, we observed a significant increase of their proportion in preferentially sporophyte expressed genes, suggesting a fundamental role in the sporophyte phase. However, further functional studies are necessary to determine their contribution to diverging sporophyte morphologies. Our results add to the growing set of studies suggesting that radical changes in morphology may rely on the heterochronic expression of conserved regulators.

Keywords: transcriptomics, sporophyte development, mosses, Funariaceae, RNAseq

INTRODUCTION

The genome is constantly reshaped by diverse types of mutations providing raw material for evolution to work with. Genomic changes can ultimately lead to new phenotypes possessing a novel set of morphological characters potentially enabling adaptation to new environmental conditions (Orr, 2005). Various molecular mechanisms may underlie the origin of morphological novelties including (i) the rise of novel genes (de novo evolution and introgression of genes)

(Chen et al., 2013; Schlötterer, 2015; Li et al., 2016; Zhang et al., 2019), (ii) the utilization of existing genes for new functions (Davidson, 2010; Pires and Dolan, 2012; Das Gupta and Tsiantis, 2018; Bowman et al., 2019), (iii) and the loss of genes or gene function (Albalat and Cañestro, 2016; Xu et al., 2019). Nevertheless, in reality molecular mechanisms underlying the origin of morphological novelties are more complex and can be best described by various combinations of these three basic scenarios.

Although the primary molecular mechanisms contributing to novel phenotypes are well-documented, their relative importance is poorly understood and difficult to predict (True and Carroll, 2002; Khalturin et al., 2009; Kaessmann, 2010; Wagner and Lynch, 2010). For instance, *de novo* evolution of genes is thought to have boosted the diversification of plants by enabling the evolution of new key characters both at the molecular and macromorphological levels (Furumizu et al., 2015; Soltis and Soltis, 2016; Jill Harrison, 2017; Van de Peer et al., 2017; Clark and Donoghue, 2018; Landis et al., 2018; Whitewoods et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019; Bowles et al., 2020). Similarly, new genes acquired by horizontal gene transfer or introgression/hybridization seem to have also been crucial in the evolution of key morphological features (Suarez-Gonzalez et al., 2018; Cheng et al., 2019; Wickell and Li, 2020). In contrast, the evolution of many key phenotypic characters has taken another path by co-opting existing genes or complete regulatory networks to create new morphological features (Rast-Somssich et al., 2015; Rebeiz and Tsiantis, 2017). Often, temporal shifts in the expression of conserved regulatory modules (e.g., heterochronic expression) is sufficient to give rise to new morphological innovations both in plant and animal systems (Geuten and Coenen, 2013; Buendía-Monreal and Gillmor, 2018). For instance, heterochronic expression of some key genes are major determinants of organ size and number in *Arabidopsis* (Sun et al., 2014). Finally, evidence is mounting that new phenotypes can also be acquired by loss of genes or gene functions. For instance, rapid evolution of new phenotypes seems to have proceeded by compromising gene function both in animals and plants (Olson, 1999; Nachman et al., 2003; Gujas et al., 2012; MacArthur et al., 2012; Nadeau et al., 2016; Sun et al., 2018). Therefore, the underlying molecular processes leading to the evolution of new morphological structures are diverse and it is currently unclear why evolution of particular phenotypes would follow one or the other trajectory (see i-iii above). Furthermore, whether evolutionary trajectories are canalized by various currently poorly known constraints or the prevalence of one trajectory is rather determined by random chance is highly debated (Galis et al., 2018). Revealing the molecular processes underlying phenotypic changes in a diverse set of model systems may help to discover key commonalities of the evolutionary process and assess how gene gain/loss and co-option of existing genes for new functions may contribute to morphological evolution.

Here, we use a model system of two species from a single family of mosses, *Physcomitrium (Physcomitrella) patens* (hereafter referred to as *P. patens*, see in Medina et al., 2019) and *Funaria hygrometrica*, to begin investigating the molecular mechanisms shaping the evolution of their highly distinct sporophyte morphologies. In contrast to flowering plants, the moss life cycle possesses a dominant haploid gametophyte (1n) (consisting of the filamentous protonema and leafy shoot-like gametophores) alternating with an unbranched diploid sporophyte (2n) phase. The diploid sporophyte phase is multicellular, photosynthetic although nutritionally dependent on the maternal gametophyte to which it is permanently attached (Jonathan Shaw and Goffinet, 2000). The primary function of the sporophyte is to produce spores and control their dispersal, and therefore the sporophyte is likely under severe selection (Haig, 2013).

The moss family Funariaceae comprises about 300 species, displaying a relatively uniform gametophyte morphology, but highly variable sporophyte stature varying in size from about a millimeter to 5 cm and in complexity from a sessile sporangium lacking a differentiated mode of dehiscence to a long stalked capsule bearing highly specialized structures for controlled spore dispersal (Medina et al., 2018, 2019). Phylogenomic analyses revealed that the subfamily Funarioideae comprises the monophyletic *Funaria* and its sister lineage, the *Entosthodon-Physcomitrium* (EP) complex (**Figure 1**) with an estimated divergence time of 60 million years (95% CI: 35–70 million years ago; Medina et al., 2018). *Funaria* is characterized by an architecturally rather complex sporophyte, with an elongated seta, asymmetric capsule, dehiscing via a revoluble double annulus revealing a double peristome regulating spore dispersal (Fife, 1982; Liu et al., 2012; Medina et al., 2018). By contrast, the sporophyte of the E-P complex is more variable, spanning a gradient of architectural complexity extending to the simple sporophyte of *P. patens* whose sporophyte is composed of a short seta and a small spherical capsule lacking differentiated structures associated with spore dispersal (**Figures 1**, **2**). Given the resolution of *P. patens* within a grade of *Entosthodon* and *Physcomitrium* species, all with more complex sporophytes, it is assumed that the sporophyte morphology of *P. patens* arose through reduction (Liu et al., 2012; Medina et al., 2018). Ontogenetic transformations of the sporophyte are correlated or at least followed by changes in the development of the calyptra, the protective maternal gametophytic tissue covering the apex of the developing sporophyte (Budke and Goffinet, 2016). Reduction of the diploid generation occurred multiple times in the complex, giving rise to several species with a *Physcomitrium* phenotype (Medina et al., 2019). Whether such transformations were triggered by similar genetic processes is not known, and in fact the molecular mechanisms underlying sporophyte development in mosses, in particular that of contrasting sporophyte morphologies in the Funariaceae, are poorly understood (Goffinet and Buck, 2013).

We used comparative gene expression and genome analysis to begin investigating the molecular changes that potentially contribute to the evolution of the contrasting sporophyte morphology in Funariaceae. To do so, we generated time-series gene expression data for four comparable stages of sporophyte development in two funarioid mosses, *F. hygrometrica*

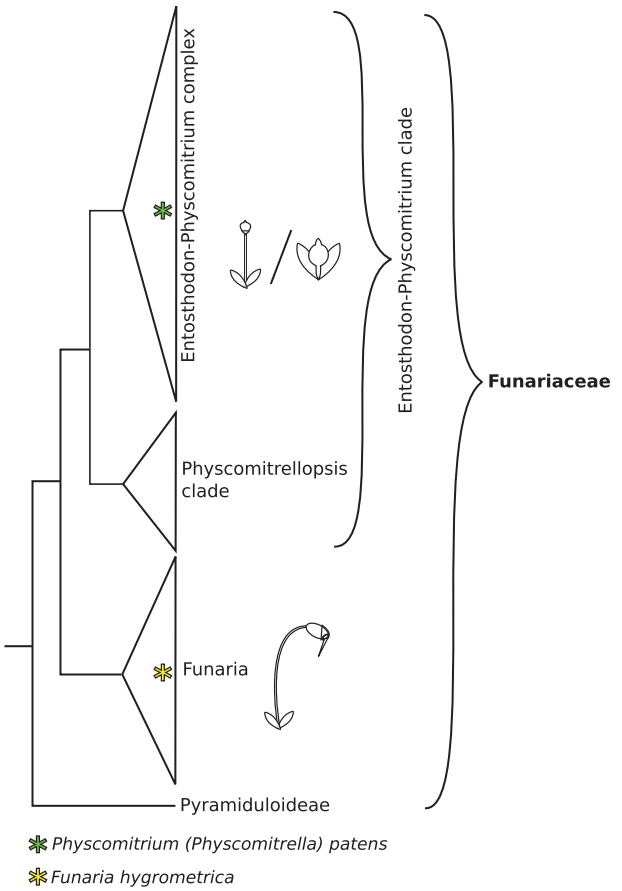


FIGURE 1 | Phylogeny backbone of the moss family Funariaceae, redrawn from Medina et al. (2018, 2019). Species of the *Funaria* clade show a complex sporophyte phenotype, characterized by an elongated seta and structures aiding controlled spore dispersal (operculum, revoluble annulus, peristome teeth). The *Entosthodon-Physcomitrium* (EP) clade is mainly comprised of species with an intermediate sporophyte phenotype, lacking one or several structures for improved control of spore dispersal and displaying an overall reduced size. Within the EP clade severe reduction of sporophyte complexity occurred at least four times independently. These highly reduced sporophytes are short and lack specialized structures aiding spore dispersal, mainly forming simple, indehiscent capsules.

(complex sporophyte) and *P. patens* (simple sporophyte), representing the two extremes of sporophyte complexity in the family. We contrasted gene expression in the two species to test whether shifts in sporophyte complexity were associated with changes in expression of shared genes or with changes in gene content, e.g., gene loss/gain. Our analyses suggest that heterochronic expression of conserved sets of developmental genes govern the development of sporophytes with contrasting complexities, although the contribution of gene gain/loss is also considerable. Our study adds to the growing set of observations that reorganization of highly conserved regulatory networks is a critical mechanism underlying major morphological shifts in evolution (Rosin and Kramer, 2009; Zhao et al., 2016; Das Gupta and Tsiantis, 2018; Bowman et al., 2019).

MATERIALS AND METHODS

Extraction of RNA From Sporophyte Tissue of *F. hygrometrica* and *P. patens*

The *F. hygrometrica* isolate used in this study was established from a single spore of a sporophyte collected around the city of Sankt Gallen, Switzerland. Gametophores were grown and vegetatively propagated in petri dishes on KNOP medium (Reski and Abel, 1985). To obtain sporophytes, gametophores were coarsely ground, and fragments spread on sterilized soil. Cultures were regularly sprayed with water to facilitate fertilization and sporophyte development was checked weekly. We collected sporophytes of *F. hygrometrica* corresponding to the four developmental stages described in Budke et al. (2012) and depicted in **Figure 2A**. We also sampled *P. patens* sporophytes with comparable developmental stages as shown in **Figure 2B**. For RNA-sequencing, we sampled sporophytes using forceps and scissors. We also used spores from freshly collected sporophytes to obtain plant material for various developmental stages of the gametophyte phase. We spread spores of surface sterilized sporophytes onto Knop medium overlaid with cellophane and collected germinating spores after 3 days (germinating spores 3 days) and 2 weeks (protonema 2 weeks). Young gametophores (young buds) emerging after 3 weeks were also collected. We immediately submerged collected sporophytes and gametophyte tissues into RNAlater® and stored them at 4°C until RNA extraction. RNA was extracted using the Spectrum™ Plant Total RNA Kit from Sigma-Aldrich. The calyptra was removed from all sporophytic tissues prior RNA extraction. Poly-A RNA of three biological replicates were sequenced with paired-end or single-end sequencing on a HiSeq 2500 or 4000 machine (see **Supplementary Table S1**).

Draft Genome Sequence of *F. hygrometrica*

We extracted DNA using a modified CTAB protocol (Porebski et al., 1997) from axenically grown gametophytes of the very same *F. hygrometrica* isolate we used for the RNA-seq experiment. We first prepared 454 libraries, which were sequenced with the titanium technology on a 454 FLX Roche machine at the Functional Genomic Center Zurich (FGCZ). We prepared short-insert DNA libraries and 3 and 5 kb jumping libraries using the Illumina DNA-seq and mate-pair library preparation kits and sequenced them at the FGCZ on HiSeq 2000 and HiSeq 2500 machines (short-insert libraries with a depth of 120x, mate-pair libraries with 26x). We quality filtered raw reads with Trimmomatic (Bolger et al., 2014) using default values for paired-end and single-end sequencing data. We assembled the genome with SOAPdenovo2 (Luo et al., 2012) using short-insert Illumina and 454 reads in the contig assembly step and 3 and 5 kb jumping libraries in the scaffolding step. We filled gaps in the final assembly using all 454 and Illumina data with Gap-closer (Luo et al., 2012). Our final assembly consisted of 13,000 scaffolds longer than 100 bp with an N50 value of ca. 100 kb and a total length of 340 Mbp. Genomic scaffolds can be found in **Supplementary Data S1**. Quality of the genome

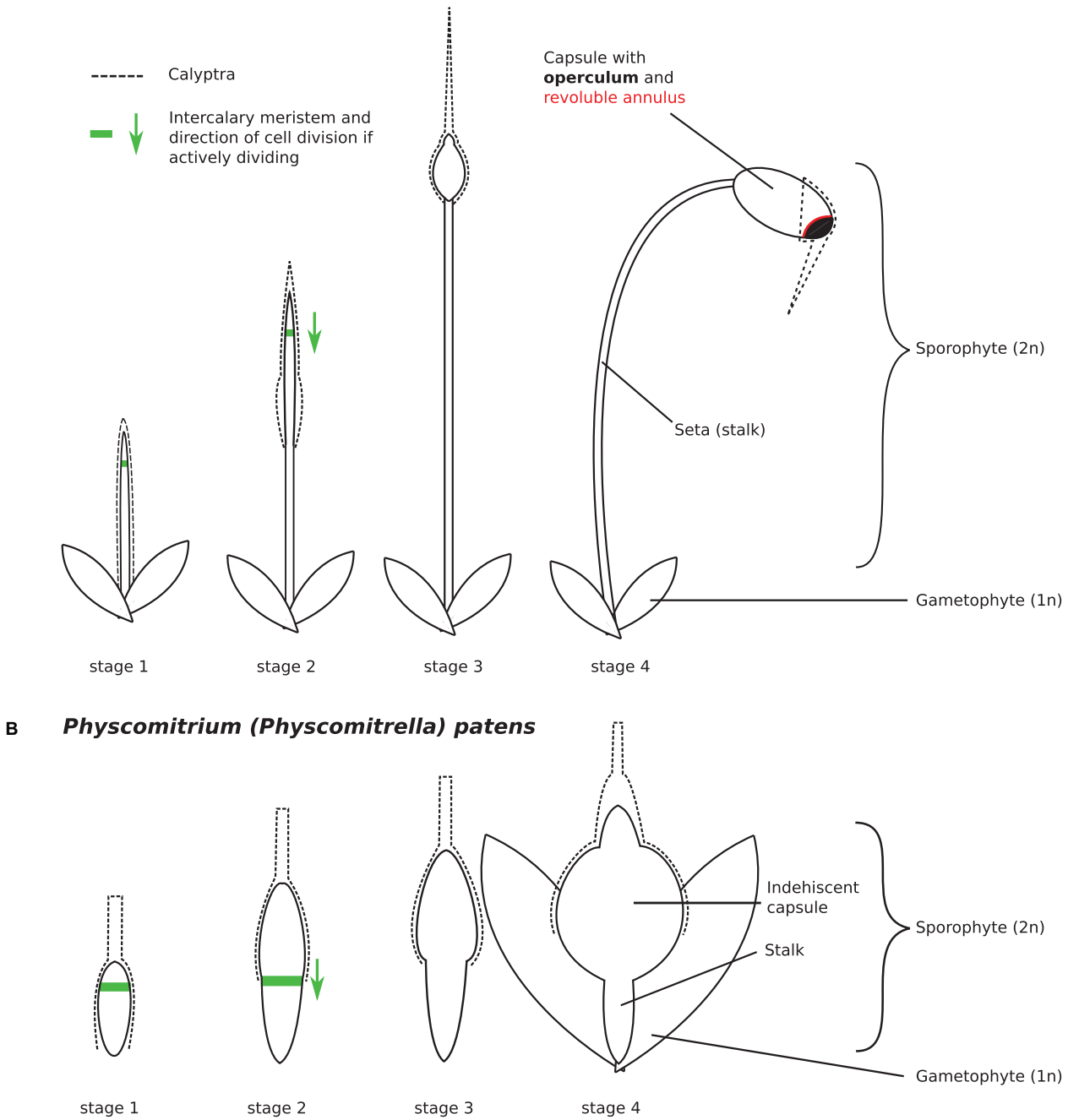


FIGURE 2 | Developmental stages of **(A)** *Funaria hygrometrica* and **(B)** *Physcomitrium patens* sporophytes investigated in this study. The stages were characterized as follows: stage 1 – sporophyte emerges from gametophytic tissue, stage 2 – elongation of the seta occurs, stage 3 – sporophyte reaches its final height, capsule begins to swell, stage 4 – capsule has swollen to its final size and all specialized structures are established but meiosis has not occurred yet. Presence of the green arrow indicates that the intercalary meristem (green line) is likely active and shows the direction of cell division. The *F. hygrometrica* sporophyte reaches a final size of 5 cm, the *P. patens* sporophyte grows to a size of 1 mm. The calyptra (dashed line) is developed from the archegonium and as such, is not part of the sporophyte.

assembly was assessed by searching translated peptide sequences of the reconstructed gene set (see in section "Read Mapping and Quantification of Expression") against hidden Markov models of Benchmarking Universal Single-Copy Orthologs (BUSCO) (Seppey et al., 2019). For this purpose, we used the Viridiplantae and Embryophyta datasets retrieved from OrthoDB v10.1 (Kriventseva et al., 2019). To put these results into context, publicly available gene coding sequences of *Marchantia polymorpha* (Bowman et al., 2017), *P. patens* (Lang et al., 2018), *Pleurozium schreberi* (Pederson et al., 2019), and *Sphagnum*

fallax were retrieved from Phytozome v12 (Goodstein et al., 2012) and searched against the same BUSCO sets. Results of all completeness assessments are shown in **Table 1**. Our results show that the reconstructed gene set captures the complete sequence of 80.5% BUSCOs from the Embryophyta dataset and of 91.8% from the Viridiplantae dataset. These BUSCO values are close to the figures obtained for high-quality genome assemblies generated using long-read technologies (*M. polymorpha*, *P. patens,* and *S. fallax*) and clearly exceed the values of the *P. schreberi* genome, which was likewise assembled using only short-reads. Overall, our BUSCO analysis implies that the *F. hygrometrica* draft genome is of good quality.

Read Mapping and Quantification of Expression

Mapping of the reads obtained by Illumina sequencing was performed following the HISAT2/StringTie pipeline (Pertea et al., 2016). First, adapter sequences were removed from the libraries and the reads were quality filtered using Trimmomatic v36 (Bolger et al., 2014) (ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25). The reads were then mapped to the *P. patens* reference genome, version 3.1 (Lang et al., 2018), and the *F. hygrometrica* draft genome (**Supplementary Data S1**), respectively, using HISAT2 (Kim et al., 2015). Detailed statistics of the read mapping are available in **Supplementary Table S1**. To calculate transcript abundance for the *P. patens* data set, we used the version 3.1 gene annotation (Lang et al., 2018) in combination with quantification in StringTie. The transcriptome of *F. hygrometrica* was reconstructed using StringTie, including *de novo* assembly of transcripts and a subsequent quantification of transcript abundance (Pertea et al., 2015). To get a complete set of transcripts expressed during sporophyte development for each species, the assembled transcripts of all samples were merged using the –merge function of StringTie (**Supplementary Data S1**). The transcriptome reconstruction of *F. hygrometrica* yielded 25,904 unique loci. We discarded all gene models with a summarized read count over all samples lower than 10 to remove gene models with very few, potentially misaligned, reads. This filtering step reduced the gene set to 25,460 (*F. hygrometrica*) and 22,690 (*P. patens*) predicted gene models which show detectable expression in the analyzed samples.

Preferential Expression in the Sporophyte Stage

Genes preferentially expressed in the sporophyte stage were identified using the R package DESeq2 version 1.16.1 (LoveI, Huber and Anders, 2014). The sporophyte expression data comprised RNAseq raw count data from three replicates of four developmental stages each (for sporophyte developmental stage 1 from *P. patens* only two replicates were used). Expression data from the gametophyte phase included raw read counts from six (*F. hygrometrica*) and eight (*P. patens*) samples from different stages of gametophyte development. After defining sporophytic and gametophytic samples, a differential expression analysis was conducted using the unmodified DESeq2 algorithm. To obtain genes highly expressed in the sporophyte stage in contrast to gametophytic stages, significant results ($p_{adj} \leq 0.05$) were filtered for gene models with a $\log_2$-fold change ≥ 2 (**Supplementary Data S2**, **S3**). We used conventional chi-square statistics to test for enrichment of genes in the set showing preferential sporophyte expression.

Identification of Homologs and Orthologs

Homologous genes (including orthologs and paralogs) were identified by the BLASTp algorithm (Altschul et al., 1990) and the software Orthofinder2 (Emms and Kelly, 2018). We used this approach to identify *F. hygrometrica* and *P. patens* gene models that are species-specific (i.e., that have no detectable homolog in the alternate species' proteome). Besides Orthofinder, we used BLASTp to detect distant homologs that might be missed by the

TABLE 1 | Assessment of completeness of the *F. hygrometrica* draft genome using Benchmarking Universal Single-Copy Orthologs (BUSCO).

	Complete			Fragmented	Missing
	Total	Single-copy	Duplicated		
Embryophyta (*n* = 1614)					
F. hygrometrica	1300 (80.5%)	1101 (68.2%)	199 (12.3%)	43 (2.7%)	271 (16.8%)
M. polymorpha	1412 (87.5%)	1385 (85.8%)	27 (1.7%)	20 (1.2%)	182 (11.3%)
P. patens	1423 (88.1%)	1174 (72.7%)	249 (15.4%)	28 (1.7%)	163 (10.2%)
P. schreberi	736 (45.6%)	540 (33.5%)	196 (12.1%)	278 (17.2%)	600 (37.2%)
S. fallax	1447 (89.7%)	1257 (77.9%)	190 (11.8%)	22 (1.4%)	145 (8.9%)
Viridiplantae (*n* = 425)					
F. hygrometrica	390 (91.8%)	336 (79.1%)	54 (12.7%)	7 (1.6%)	28 (6.6%)
M. polymorpha	411 (96.7%)	409 (96.2%)	2 (0.5%)	3 (0.7%)	11 (2.6%)
P. patens	416 (97.8%)	361 (84.9%)	55 (12.9%)	1 (0.2%)	8 (2.0%)
P. schreberi	227 (53.4%)	160 (37.6%)	67 (15.8%)	107 (25.2%)	91 (21.4%)
S. fallax	417 (98.1%)	375 (88.2%)	42 (9.9%)	2 (0.5%)	6 (1.4%)

For comparison, BUSCO statistics for the Marchantia polymorpha, Physcomitrium patens, Pleurozium schreberi, and Sphagnum fallax genomes are shown. Data was retrieved from Phytozome v12 (Goodstein et al., 2012) and analyzed using the Viridiplantae and Embryophyta BUSCO datasets obtained from OrthoDB v10.1.

more stringent approach of Orthofinder. The BLASTp search included reference peptide sequences for each gene model in the *P. patens* and *F. hygrometrica* data sets. Hits were filtered for alignments with excess similarity by applying an *E*-value threshold of $\leq 10^{-6}$. To remove alignments that cover only small conserved domains within the query we applied a second filtering step for alignments with $\geq$80% query coverage and $\geq$35% sequence similarity between the query and the corresponding hit. Both filtering methods are reported to reliably detect homologous sequences even between distantly related species (Rost, 1999; Pearson, 2013). We also ran Orthofinder2 v2.3.3 (Emms and Kelly, 2018) to identify orthogroups (comprising both orthologs and paralogs) between *P. patens* and *F. hygrometrica* using default parameters (**Supplementary Data S4**). We included the following proteomes to generate orthogroups: *Citrus clementina, M. polymorpha, Theobroma cacao, Vitis vinifera, Prunus persica, Cucumis sativus, Amborella trichopoda, Physcomitrium patens, F. hygrometrica, Selaginella moellendorffii, Zea mays, Oryza sativa v7_JGI, Brassica oleracea, Arabidopsis thaliana, P. trichocarpa, Medicago truncatula,* and *Daucus carota*. All proteomes, except the one for *F. hygrometrica*, were retrieved from Phytozome v12 (Goodstein et al., 2012). We also let Orthofinder2 v2.3.3 automatically calculate gene trees for each orthogroup and identify one-to-one orthologs of *P. patens* and *F. hygrometrica* using phylogenetic information. We used one-to-one orthologs between *P. patens* and *F. hygrometrica* to compare gene expression in the two species.

PCA

To gain insights into the overall change of gene expression associated with sporophyte development in *F. hygrometrica* and *P. patens,* we used principal component (PCA) analyses (Hotelling, 1933) on normalized and standardized RNA-seq expression data (**Figure 3A**). Because we assumed that divergent sporophyte morphologies may be partially associated with gene gain/loss we generated PCAs separately using each species' full gene set (including both homologs and species-specific genes), the gene set with homologs in the alternate species (including orthologs and paralogs) and the species-specific gene set (non-homologs, for details see section "Identification of Homologs and Orthologs"). We argue here that the amplitude of a gene's expression variation (expression dynamics) throughout sporophyte developmental stages likely correlates with its functional relevance. Following this reasoning, we expected more pronounced differentiation across sporophyte developmental stages in gene expression using the species-specific gene set if gene gain/loss is more important in contributing to the divergent sporophyte morphologies than genes that are part of the shared gene set (homologous genes). Alternatively, expression variation across sporophyte developmental stages may be more pronounced using the shared gene set if sporophyte development is rather driven by gene expression changes in homologous genes and not the expression dynamics of species-specific genes.

To further investigate which sporophyte developmental stages showed similar or rather divergent gene expression patterns in the two species we calculated an expression divergence matrix between developmental stages, using one-to-one orthologs that showed expression in both species in the corresponding stage. The expression divergence matrix was built by calculating pairwise Spearman's rank correlation coefficients between expression profiles of developmental stages of *F. hygrometrica* and *P. patens*. Results were plotted using the R package ggplot2 v. 3.2.1 (Wickham, 2009).

Functional Annotation

We retrieved functional annotation of *P. patens* gene models from Phytozome v11 (Goodstein et al., 2012). We translated and annotated transcript sequences of *F. hygrometrica* by running tBLASTn searches (*E*-value threshold 10^{-6}) against the plant proteomes available in PLAZA v2.0 (Proost et al., 2009). GO annotation for each transcript was obtained by transferring the GO annotation of the respective best hit protein to the *F. hygrometrica* transcript.

Key positions in gene regulatory networks are often occupied by transcription factors, which trigger expression of an array of downstream target genes resulting in a regulatory cascade setting off developmental transitions (Davidson and Erwin, 2006; Erwin and Davidson, 2009). Identification of gene models that encode transcription factors was done based on conserved DNA binding motifs using the PlantTFDB 4.0 resource (Jin et al., 2017). The algorithm reported 1,061 (*P. patens*) and 849 (*F. hygrometrica*) transcription factor encoding genes, which are expressed in at least one sample.

Gene Ontology Term Enrichment

To functionally characterize genes preferentially expressed in the sporophyte phase of *F. hygrometrica* and *P. patens*, we used the R package TopGO, version 2.38.1 (Kim, 2019). For the enrichment analyses we only considered GO-terms of the class "Biological Process." After defining all gene models with detectable expression in the RNA-seq data set of sporophyte and gametophyte tissue as the gene universe, and identifying all genes which are preferentially expressed in the sporophyte phase of either species as the subset of interesting genes, enriched GO-terms were computed using the Parent-Child Algorithm (Grossmann et al., 2007) and their significance tested with Fisher's exact test, both implemented in the TopGO package. Lists of enriched GO-terms (**Supplementary Data S5–S8**) were reduced using the REVIGO online resource (Supek et al., 2011) to remove redundant terms. Results of the enrichment analysis (**Figure 4**) were visualized using the R package GOsummaries version 2.22.0 (Kolde and Vilo, 2015).

Clustering by Gene Expression Profiles

The main goal of this study is to identify the genes and gene networks that may have contributed to the divergent development of the sporophyte phase in *P. patens* and *F. hygrometrica*. Since development is a dynamic, time dependent process we sought to use a method that would allow us to statistically distinguish one-to-one orthologous gene sets showing differential expression dynamics throughout sporophyte development between *P. patens* and *F. hygrometrica*. To compare expression dynamics of one-to-one orthologous genes in the two species during sporophyte development, we used

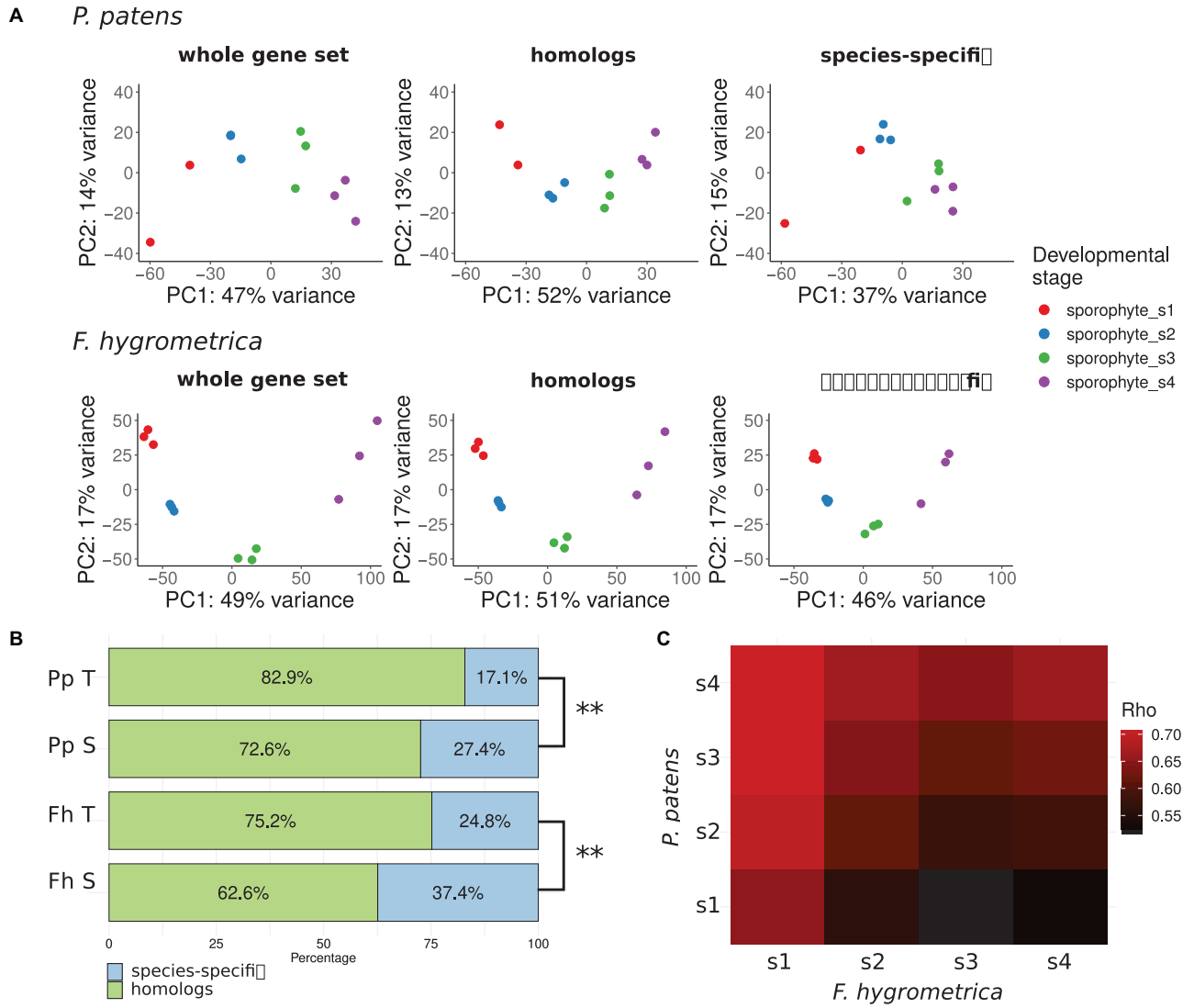


FIGURE 3 | Gene expression variation throughout the sporophyte development of *Funaria hygrometrica* and *Physcomitrium patens*. **(A)** Principal component analysis (PCA) of gene expression variation during sporophyte development in *P. patens* and *F. hygrometrica*. PCA was carried out by using gene expression data for all expressed genes ("whole gene set"), for homologous genes ("homologs"), and for species-specific genes ("species-specific"). **(B)** Proportion of homologous and species-specific genes in the transcriptome data sets of *F. hygrometrica* (Fh) and *P. patens* (Pp). For each species the total set of transcripts detected in the gametophyte and sporophyte phases (Pp T, Fh T) and the set of transcripts that is specifically expressed in the sporophyte phase (Pp S, Fh S; $\log_2$ fold change ≥ 2, $q \leq 0.05$) is shown. ** marks significant dependence of distribution between homologs and species-specific genes on the considered subset according to Chi-squared test ($p \leq 0.01$). **(C)** Correlation coefficients (Spearman's Rho) of gene expression between developmental stages 1 to 4 of *P. patens* and *F. hygrometrica* sporophytes. Only one-to-one orthologs for which expression was observed in both samples were used to calculate the corresponding rank correlation. Spearman's Rho was calculated using normalized and log-transformed raw count values.

a method enabling classification of genes by their expression pattern and mapping gene expression profiles of one species against clusters of the other species. To achieve this, we carried out fuzzy c-means clustering with the R package mfuzz (Kumar and Futschik, 2007). We first clustered the *F. hygrometrica* gene expression data and then statistically assigned genes of *P. patens* to these clusters according to their expression dynamics. Because this assignment relies on the alignment of putatively functionally conserved genes, the analysis is restricted to one-to-one orthologs as detected by the Orthofinder algorithm (see section "Materials and Methods"). To obtain the optimal cluster number, we computed the minimum centroid distance for a range between four and 40 clusters using the implemented Dmin() function. Centroid distances decreased rapidly up to a cluster number of 12. For higher cluster numbers the distance curve flattened out and changes in centroid distances were less severe. Therefore, we decided to use 12 clusters for the fuzzy c-means clustering as a compromise between describing the expression dynamics reasonably well without obtaining too

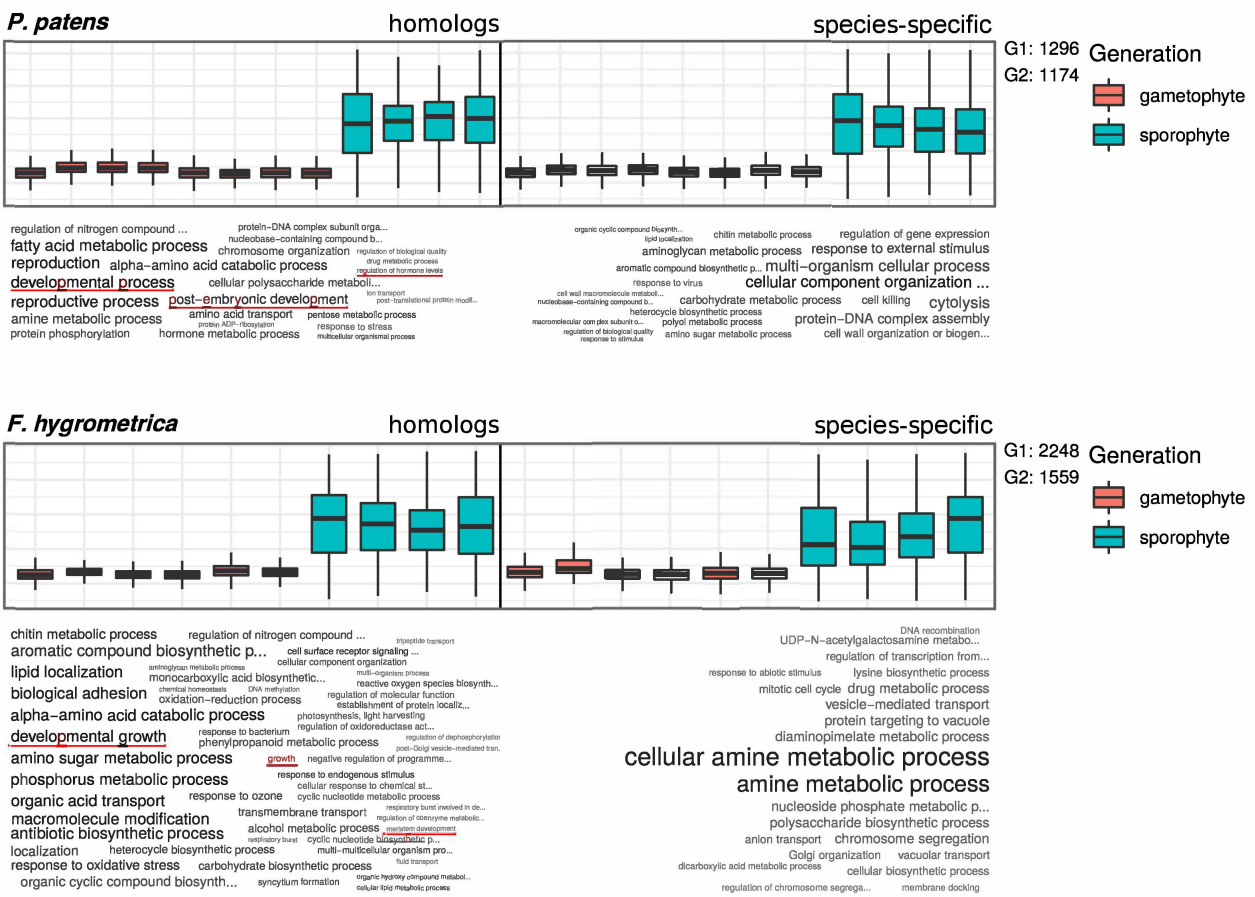


FIGURE 4 | GO term enrichment analysis of preferentially sporophyte expressed genes in *Physcomitrium patens* and *Funaria hygrometrica*. Upper panels show standardized expression values of the respective subset of genes. Homologous genes are shown on the left side and species-specific genes on the right side of the plots. Lower panels show a visualization of enriched GO terms in the respective subsets, word size is correlated with the significance value obtained from the enrichment analysis. G1 and G2 denote the number of genes in the subsets of homologous (G1) and species-specific (G2) genes, used in the enrichment analysis. GO terms potentially connected to differential sporophyte development in the two species are underlined (see red lines).

many redundant clusters. The optimal fuzzifier was computed according to the method described by Schwämmle and Jensen (2010), which is implemented in the mfuzz package (mestimate() function). Assignments of *F. hygrometrica* and *P. patens* one-to-one orthologs to the clusters can be found in **Supplementary Data S9**, **S10**.

RESULTS

To identify genes potentially underlying the difference in sporophyte development between *F. hygrometrica* and *P. patens*, we measured gene expression across four distinct stages of sporophyte development and several stages of gametophyte development in both species using RNA-sequencing (**Figure 2**). We assumed that most genes expressed or preferentially expressed in the sporophyte phase are important for proper sporophyte development. These data form the basis of our analyses described below.

Proportion of Preferentially Sporophyte Expressed Genes Is Slightly Higher in the Species With Greater Sporophyte Complexity

First, we performed a quantitative and qualitative assessment of the gene set expressed during sporophyte development in both species. After applying a relaxed expression threshold (see section "Materials and Methods"), 22,690 *P. patens* gene models were considered to be expressed in at least one of the sporophyte developmental stages, which is about 64% of the 35,307 currently annotated gene models (Lang et al., 2018). Similarly, of the total 25,904 gene models assembled using RNA-seq data of *F. hygrometrica* (see section "Materials and Methods"), 25,460 (98%) showed expression during sporophyte development. Of the gene set expressed during sporophyte and gametophyte development, 2,807 (12.4%) (*P. patens*) and 3,807 (15%) (*F. hygrometrica*) gene models were strongly upregulated in the sporophyte versus the gametophyte phase ($\log_2$ fold change

≥2, false discovery rate ≤0.05). This suggests that the proportion of preferentially sporophyte expressed genes is slightly higher in the species with a more complex sporophyte morphology, but such genes represent only a relatively small proportion of the total gene set in both species.

Sporophytic Transcriptomes of *P. patens* and *F. hygrometrica* Are Dominated by Homologs but the Contribution of Species-Specific Genes Is Not Negligible

We then asked whether transcripts expressed during sporophyte development mainly represent genes homologous (including orthologs and paralogs, see identification of homologs in section "Identification of Homologs and Orthologs") between the two species or, instead, species-specific genes. The Orthofinder analysis revealed that 19,667 and 19,267 genes of *P. patens* and *F. hygrometrica* fall into orthogroups containing sequences from both species. That is 55% (19,667 out of 35,307) and 74% (19,267 out of 25,904) of the *P. patens* and *F. hygrometrica* full gene set had homologs (orthologs and paralogs) in the alternate species' proteome, respectively. Out of these homologs, 12,741 gene models were identified as one-to-one orthologs. When expressed as a proportion of genes transcribed throughout sporophyte development in *P. patens* and *F. hygrometrica*, 85.3% (19,364 of 22,690 gene models) and 75.6% (19,245 of 25,460 gene models), respectively, had homologs in the alternate species' genome. These numbers did not change considerably when using the more relaxed BLASTp approach (see section "Identification of Homologs and Orthologs") to include more distant homologs. To ensure that genes without homologs in either one of the species are species-specific and not missed by the annotation process, we searched protein sequences of putatively non-homologous genes against the corresponding genome sequence, using the tBLASTn algorithm (Camacho et al., 2009) with a threshold of ≥80% query coverage and ≥35% sequence similarity between the query and the corresponding hit. We identified an additional 2,003 gene models of *P. patens* to have putative homologs in the *F. hygrometrica* genome, which was expected due to the fact that the gene models for *F. hygrometrica* were assembled using RNAseq data and genes, which are not expressed in the sampled stages will not be detected by this method. Similarly, 980 additional gene models in the *F. hygrometrica* gene set have homologs in the *P. patens* genome. Combined with the expression data, 81% and 78% of the sporophytic transcriptome of *P. patens* and *F. hygrometrica*, respectively, thus have homologs in the alternate species' genome. In summary, the two species express a large proportion of homologous (about 80%) genes during the development of their sporophyte. Nevertheless, about 20% of their genes expressed during sporophyte development appear species-specific, that is with no clear homologs in the other species' genome.

F. hygrometrica Shows More Expression Variation During Sporophyte Development Than *P. patens*

Sporophyte development stages showed distinct expression patterns in *F. hygrometrica* in the shared (orthologs and paralogs: homologs) and species-specific gene sets (**Figure 3A**). By contrast, differentiation in gene expression among the four sporophyte developmental stages in *P. patens* was strongly dependent on the gene set used. They were well-differentiated when we used all genes or only the set of homologs, whereas only the early and the late two developmental stages were well-distinguishable when comparing expression of species-specific genes (**Figure 3A**).

Species-Specific Genes Are Overrepresented Among Preferentially Sporophyte-Expressed Genes

Next, we tested how species-specificity or shared nature of genes is correlated with their putative function in sporophyte development. We assumed that preferential expression of genes in the sporophyte stage (compared to the gametophyte) can be used as a proxy for their functional role in sporophyte development. During gametophyte and sporophyte development 24,031 and 25,818 gene models were expressed in *P. patens* and *F. hygrometrica* of which 82.9% and 75.2% fell into orthogroups containing genes from both species, respectively. When using genes preferentially expressed in the sporophyte phase ($\log_2$ fold change ≥2, false discovery rate ≤0.05), the proportion of genes coming from orthogroups shared by the two species decreased: only 72.6% (*P. patens*, 1,792 gene models) and 62.6% (*F. hygrometrica*, 2,385 gene models) of preferentially sporophyte expressed genes were homologous between the two species (**Figure 3B**). This difference in the proportion of homologous genes between preferentially sporophyte expressed and all expressed genes was highly significant in both species according to a chi-squared test (*P. patens*: $\chi^2_{df=1} = 206.16$, $p < 2.2 \times 10^{-16}$; *F. hygrometrica*: $\chi^2_{df=1} = 384.73$, $p < 2.2 \times 10^{-16}$). Altogether, this implies that species-specific genes are significantly overrepresented while homologous genes are underrepresented in the gene set preferentially expressed in the sporophyte.

Comparing gene expression at specific developmental stages between the two species (**Figure 3C**) using one-to-one orthologs reveals that all sporophytic stages of *P. patens* were most similar to the earliest sporophyte developmental stage of *F. hygrometrica* (stage number 1, see **Figure 3C**). Furthermore, overall expression similarity increased between *P. patens* and any of the *F. hygrometrica* developmental stages along a developmental chronology with the expression in the first stage of *P. patens* being most dissimilar. This suggests that the major differences in gene expression of orthologs can be found during early sporophyte development, a stage where precursors of various tissues that will form the mature sporophyte are likely established (Schwartz, 1997).

Functional Analysis of Genes Suggests a Contribution of Homologs in Establishing Contrasting Sporophyte Phenotypes

Comparing GO-terms of species-specific and homologous genes preferentially expressed in the sporophyte stage of both species (see section "Materials and Methods" for details) revealed that 103 (homologs) and 32 (species-specific) GO terms in

F. hygrometrica and 55 (homologs) and 50 (species-specific) GO terms in *P. patens* were significantly enriched in the corresponding subset (Fisher's exact test, $p \leq 0.05$, **Figure 4**). The majority of enriched GO-terms is related to various metabolic processes in both species. Nevertheless, some significantly enriched GO terms are potentially linked to the establishment of sporophyte morphology, which among others includes terms such as "regulation of growth", "developmental transitions", and "structure morphogenesis" (**Figure 4**). Importantly, only the homologous gene set was enriched for GO terms potentially linked to the morphology and development of the sporophyte phase in both species investigated (**Figure 4**). By contrast, the species-specific subset of preferentially sporophyte expressed genes were primarily enriched for general metabolic processes. In particular, homologous genes preferentially expressed in the sporophyte of *P. patens* were enriched for terms "developmental process," "post-embryonic development," and "regulation of hormone levels." Similarly, homologous genes preferentially expressed in the sporophyte in *F. hygrometrica* were enriched for GO-terms "developmental growth," "growth," and "meristem development" that are possibly linked to the establishment of the complex sporophyte morphology. By contrast, species-specific genes preferentially expressed in the sporophyte phase in both species were primarily enriched for GO terms related to general metabolic processes and not to differential development of sporophytic structures. Furthermore, while about 50% of the preferentially sporophyte expressed homologs in both species were associated with at least one GO-term, only 12% (*F. hygrometrica*) and 22% (*P. patens*) of the species-specific genes could be functionally characterized by one or more GO-terms, and consequently the contrasting pattern of homologous and species-specific genes may result from their sparse functional annotation.

Several Orthologous Developmental Regulators Show Heterochronic Expression in *P. patens* and *F. hygrometrica*

We used fuzzy c-mean clustering to identify groups of one-to-one orthologous genes experiencing differential expression dynamics throughout sporophyte development between *P. patens* and *F. hygrometrica*. We first clustered the *F. hygrometrica* gene expression data and then statistically assigned genes of *P. patens* to these clusters according to their expression dynamics. Clustering genes according to their expression profile over sporophyte development in *F. hygrometrica* resulted in 12 clusters with distinct expression patterns (referred to as minor clusters). We assigned those clusters to early, mid, and late sporophyte development (referred to as major clusters), based on the expression peak of each cluster (**Figure 5**). Following filtering (see section "Materials and Methods"), the data set contained 3,976 one-to-one orthologs assigned to one of the minor clusters. Of these, 1,807 (45.4%) were assigned to the same major cluster in both species, and 2,169 (54.6%) to different major clusters (**Supplementary Data S9**).

Looking at only transcription factor encoding genes did not influence our conclusion. Half of the 230 transcription factor encoding one-to-one orthologs present in the analyzed data set (i.e., 114, 51.8%) shifted between major clusters in *P. patens* and *F. hygrometrica* whereas the other half (i.e., 106, 48.2%) stayed in the same major cluster. These proportions were not significantly different from those when considering all expressed genes. Among the differentially expressed transcription factor encoding genes, many are members of gene families frequently associated with developmental control and growth in plants, including AP2/ERF, ARF, GRAS, MADS-box, TALE and WOX (see **Supplementary Data S10** for a complete table).

DISCUSSION

Morphological novelties may be associated with gene birth/death (*de novo* evolution of genes or loss of genes) or can alternatively be achieved by existing genes acquiring a new function (Davidson, 2010; Pires and Dolan, 2012; Chen et al., 2013; Schlötterer, 2015; Li et al., 2016; Das Gupta and Tsiantis, 2018; Bowman et al., 2019; Zhang et al., 2019). The relative contribution of these two major processes to the evolution of morphological novelties is still poorly known. Here, we compared the transcriptomes of two moss species, *P. patens* and *F. hygrometrica*, to investigate the relative contribution of gene gain/loss and shifts in expression dynamics of orthologous genes to the contrasting morphologies of their sporophyte phase. We found that the divergent sporophyte morphologies are primarily achieved by the heterochronic expression of a conserved set of genes, and while species-specific genes are likely to be also important, their contribution remains to be clarified. Our study contributes to the growing set of observations suggesting that shifting temporal dynamics of conserved genes represents a frequent mechanism through which radically different morphologies can be achieved (Sakamoto et al., 2009; Pires and Dolan, 2012; Lemmon et al., 2016; Olson and Nedelcu, 2016; Pham et al., 2017; Fisher et al., 2019). Moreover, it provides critical information on the molecular processes contributing to the diversification of moss sporophytes, a central topic of bryophyte taxonomy and evolutionary research. In the following paragraphs we discuss the implications of our findings for the molecular mechanisms underlying divergent sporophyte morphologies in *P. patens* and *F. hygrometrica* and for the evolution of morphological novelties in general.

Proportion of Preferentially Sporophyte-Expressed Genes Is Small

Developmental processes regulating the morphological properties of organisms are known to be composed of an interconnected set of genes organized in so-called regulatory networks (Davidson and Erwin, 2006; Davidson, 2010; Das Gupta and Tsiantis, 2018). Regulatory networks are modular, which makes them highly evolvable through the replacement of conserved modules and the rewiring of regulatory connections (Rosin and Kramer, 2009; Glassford et al., 2015; Halfon, 2017; Verd et al., 2019). Although many genes are expressed during

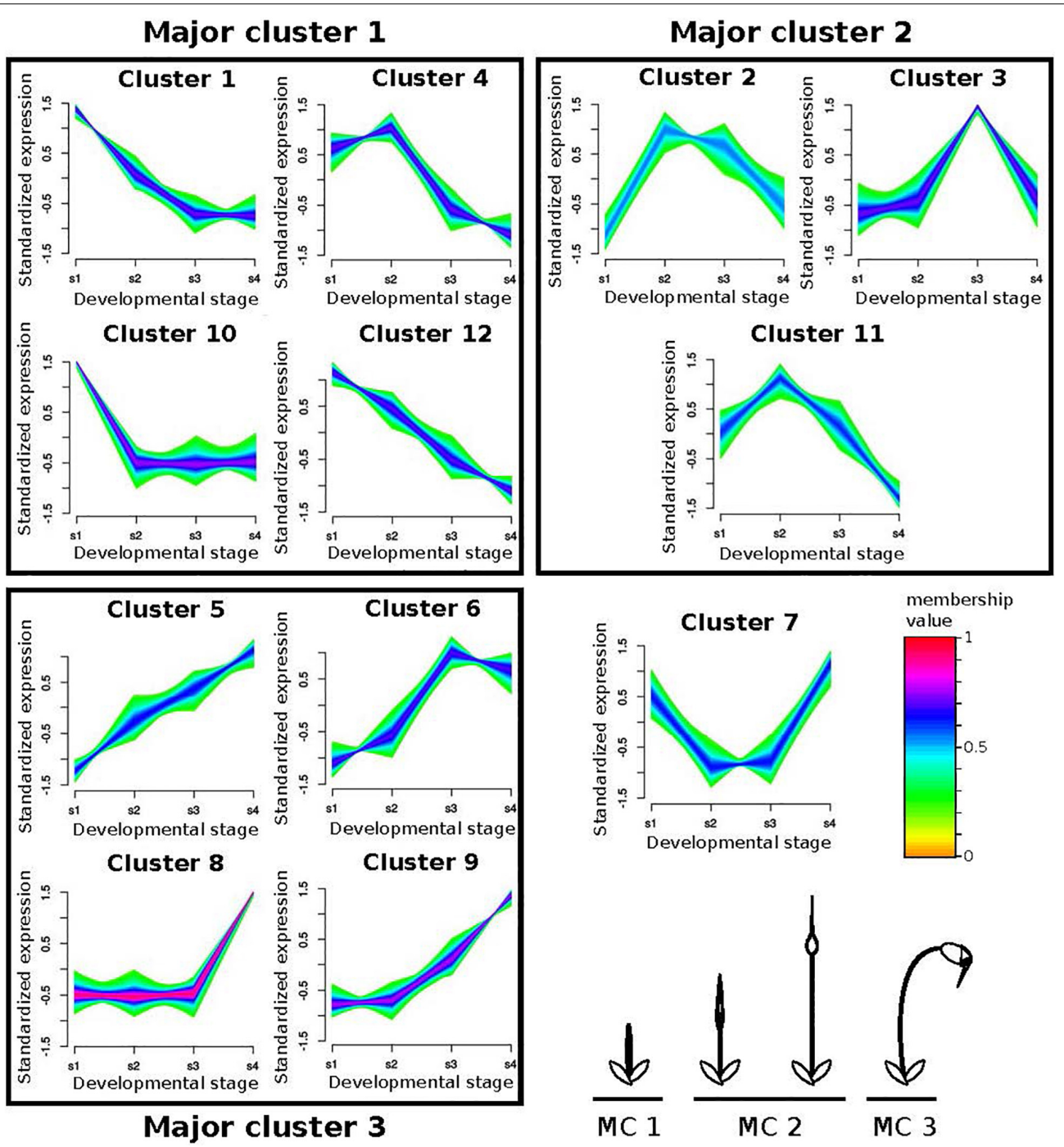


FIGURE 5 | Fuzzy clustering of gene expression data from 4 stages of sporophyte development of *Funaria hygrometrica* and *Physcomitrium patens*. We identified 12 clusters with distinct expression profiles. For each cluster, relative expression changes of the assigned genes are shown together with their color-coded membership values, indicating how strong their expression pattern correlates with the cluster core. The clusters are further condensed to Major clusters (MC), depending on whether expression peaks occurred during early (MC 1), intermediate (MC 2), or late (MC 3) sporophyte development (bottom right). Cluster 7 could not be assigned to one of the major clusters. s1 – s4: sporophyte developmental stages 1 to 4.

sporophyte development in both species, only about 12–15% indeed exhibit preferential sporophytic expression and are thus likely components of the genetic network underlying the development of the sporophyte versus the gametophyte. Consequently, shifts from the gametophyte to the sporophyte program would only require changes in expression of a relatively small set of major genes. This is in line with multiple studies reporting how shifts in temporal or local expression of a

small set of genes can lead to radical change in morphology (Frankel et al., 2011; Ichihashi et al., 2014; Sicard et al., 2014; Di Ruocco et al., 2018; Whitewoods et al., 2020). Whether this assertion may also apply to the molecular mechanisms underlying contrasting sporophytic architectures is discussed in the paragraphs below.

Evidence for Control of Divergent Sporophyte Development by Conserved Orthologs Is Mounting, but Contribution of Species-Specific Genes Remains Unclear

Novel morphological structures can be acquired by the evolution of new genes or the differential usage of already existing gene sets, but the significance of either mechanism in shaping innovation or shifts in morphology remains ambiguous, and likely spans a broad range across lineages (Davidson, 2010; Pires and Dolan, 2012; Chen et al., 2013; Schlötterer, 2015; Li et al., 2016; Das Gupta and Tsiantis, 2018; Bowman et al., 2019; Zhang et al., 2019). Although the body plan of the moss sporophyte is fundamentally simple (Goffinet and Buck, 2013), the complexity of its mature architecture varies considerably (Vitt, 1981), and homoplastic reduction seems to be common across mosses (Schwartz, 1994; Shaw et al., 2000).

Within the Funariaceae the architecture of the sporophyte varies considerably, with *P. patens* and *F. hygrometrica* representing the two extremes of the spectrum (Liu et al., 2012; Medina et al., 2018, 2019). Their large-scale morphological difference lies in the length of the seta, the (a)symmetry of the spore bearing capsule and the presence/absence of appendages controlling spore dispersal (Fife, 1982). Sporophytes of *F. hygrometrica* consist of a 2–4.5 cm long stalk (seta) bearing an asymmetric capsule with subapical differentiated cells (annulus) enabling the release of a lid (operculum), and the then exposed mouth is lined by small hygroscopic teeth (peristome). By contrast, the sporophyte of *P. patens* is composed of an extremely short stalk, ending in a spherical capsule whose epidermal wall disintegrates when spores are mature (**Figure 2**).

Our study provides the first hints on the molecular mechanisms potentially underlying contrasting sporophyte architectures in *P. patens* and *F. hygrometrica*. We found that about half of the one-to-one orthologs between *P. patens* and *F. hygrometrica* show divergent expression dynamics during sporophyte development of the two species as demonstrated by our clustering analysis. Furthermore, of the transcription factor encoding orthologs that show heterochronic expression, many are members of gene families frequently associated with developmental control and growth in plants, including AP2/ERF, ARF, GRAS, MADS-box, TALE and WOX (see **Supplementary Data S10** for a complete table). Finally, our gene ontology analysis and previous functional studies on *P. patens* also provided evidence that some of these genes are likely to be key players in contributing to the divergent sporophyte phenotypes. Altogether, our data supports the notion that drastic macromorphological differences in sporophyte morphology (seta length and opening mechanisms of the capsule) are likely achieved by the heterochronic expression of orthologous transcription factors.

Besides the overwhelming support for the contribution of orthologous transcription factors to the two contrasting sporophyte architectures, our analysis provides strong evidence that species-specific genes are also fundamental. We found that a large proportion of all (roughly 20%) and preferentially sporophyte expressed genes (between 30 and 40%) had no orthologs/homologs in the alternate species' genome and thus were truly species-specific. The proportion of species-specific genes between *P. patens* and *F. hygrometrica* appears to be high but not exceptional when considering their estimated divergence time of 60 million years (Medina et al., 2018). Comparable values can also be observed between closely related species pairs in angiosperms. For instance, *Arabidopsis thaliana* and *Arabidopsis lyrata* diverged 10 million years ago and roughly 16–18% of their gene sets are species-specific, respectively (Hu et al., 2011). Similar proportion of species-specific genes were also observed between the sister species *Zostera marina* and *Zostera muelleri*, which diverged from one another roughly 14 million years ago (Lee et al., 2016; Olsen et al., 2016). We further found that genes upregulated in the sporophyte phase were significantly enriched for species-specific genes suggesting their preferential recruitment for sporophytic functions. Unfortunately, species-specific genes are poorly annotated and thus their specific contribution in establishing divergent sporophyte architectures remains unclear and awaits clarification. In the next two paragraphs, we will focus on some major morphological features and their suggested molecular basis supported by our analyses.

Molecular Basis of Differential Seta Length in *P. patens* and *F. hygrometrica*

The seta develops through the activity of an intercalary meristem located immediately below the presumptive tissue of the capsule (Garner and Paolillo, 1973; French and Paolillo, 1976; Sano et al., 2005; **Figure 2**). The intercalary meristem is basipetal, such that cells are added below it and elongating, contributing to the growth of the seta. Thus, the length of the seta is likely correlated with the duration of activity of the intercalary meristem during sporophyte development. Given that the sporophyte of *P. patens* completes its development faster than *F. hygrometrica* (French and Paolillo, 1975; Sano et al., 2005; Coudert et al., 2019) as its meristem soon ceases to add new cells, it is therefore likely that expression of the underlying regulatory networks is shut down sooner or shortened compared to *F. hygrometrica*. That is, heterochronic expression of the meristem activity regulatory network can contribute to the seta length difference.

The class I KNOX gene *PpMKN2* promotes sporophyte axis extension by activating cytokinin biosynthesis pathways in *P. patens* (French and Paolillo, 1975; Coudert et al., 2019). In this species, *PpMKN2* is mainly expressed in the first sporophyte stage with expression quickly decreasing in more advanced stages of sporophyte development (**Figure 5**, cluster

10). By contrast, Fh_22445, the *F. hygrometrica* ortholog of *PpMKN2*, is characterized by a slower decline of expression reaching its minimum in the third and fourth phases of sporophyte development (**Figure 5**, cluster 1). The apparent correlation between difference in gene expression and degrees of seta length in the two species may suggest that the prolonged expression of the class I KNOX ortholog in *F. hygrometrica* accounts for the elongated seta phenotype.

A WUSCHEL-related homeobox13-like (WOX13-like) gene in *P. patens* (Pp3c15_20000) and its ortholog Fh_3609 also display differential gene expression dynamics in the two species. Fh_3609 is equally highly expressed during the first two developmental stages, but only weakly in other stages (**Figure 5**, cluster 4), whereas Pp3c15_20000 reaches an expression maximum only during the second stage of sporophyte development in *P. patens* (**Figure 5**, cluster 11). WOX13-like genes are important factors in initiating cell growth during zygote and stem cell formation by upregulating genes controlling cell wall loosening, including expansins (Sakakibara et al., 2014). This coincides with our finding that growth related genes are enriched in the set of preferentially sporophyte expressed genes in *F. hygrometrica*, among them three expansin homologs (Fh_21099, Fh_21985, Fh_4588). Furthermore, all of these expansins were assigned to the same major cluster as the *WOX13-like* gene Fh_3609 and two of them to the very same minor cluster. The three expansin homologs also have co-orthologous sequences in *P. patens* (Pp3c8_15200, Pp3c20_5780, Pp3c21_22300), but only Pp3c8_15200 is co-expressed with the *WOX13-like* gene Pp3c15_20000, whereas the other two show opposite expression patterns reaching their maximum expression in developmental stage 4 (**Figure 5**, cluster 5). Taken together, these results suggest that additional genes aiding in cell expansion and growth were recruited to the regulatory module initiated by expression of the *WOX13-like* gene Fh_3609 in *F. hygrometrica*, presumably to facilitate the rapid growth of the seta during early sporophyte development. Alternatively, regulatory connections between the *WOX13-like* gene of *P. patens* and some expansin homologs could have been lost in *P. patens*, contributing to the reduced growth of the seta.

Regulators of Abscission and Dehiscence Pathways Could Fulfill a Similar Function in Peristome and Annulus Development in *F. hygrometrica*

In addition to their highly contrasted seta length, the sporophytes of *F. hygrometrica* and *P. patens* differ in the presence/absence of specialized structures aiding in spore release and dispersal (**Figure 2**). The sporangium of *F. hygrometrica* opens via the loss of a lid (operculum), enabled by the differentiation of an annulus, a subapical ring of cells forming a predetermined breaking point between capsule body and operculum (Garner and Paolillo, 1973). The release of the operculum exposes the capsule mouth, which is lined by two rows of hygroscopic appendages, the peristome teeth, whose movement over the mouth at least partially controls the dispersal of spores. The development of both structures, the annulus and peristome teeth, requires the establishment of tissue layers, spatially and temporally tightly regulated cell death and cell wall break down during development of the sporophyte (Budke et al., 2007; Goffinet et al., 2009; Goffinet and Buck, 2013), a process reminiscent of abscission and dehiscence found in a wide variety of plant species.

Control of abscission and dehiscence on a genetic level are well studied in vascular plants (Lenser and Theißen, 2013; Kim et al., 2019), but regulatory pathways remain unexplored in bryophytes. Central regulators of abscission zone (AZ) formation in *Arabidopsis thaliana* are the redundantly acting transcription factors *BLADE-ON-PETIOLE1* (*BOP1*) and *BOP2* (McKim et al., 2008; Wu et al., 2012; Khan et al., 2014), for which co-orthologs are present both in *F. hygrometrica* (Fh_25624, Fh_20164, Fh_1793) and *P. patens* (Pp3c17_8330, Pp3c1_35410, Pp3c14_11190). In *P. patens* the BOP co-orthologs show expression throughout sporophyte development with peaks in stages 1, 3 and 4, respectively. By contrast, expression of the *F. hygrometrica* BOP co-orthologs is more focused to intermediate developmental stages with Fh_25624 and Fh_1793 displaying peak expression in stage 2 and Fh_20164 being most strongly expressed in stage 3. The differentiation of capsule tissues begins after the seta has reached its maximum length (Garner and Paolillo, 1973), which happens around developmental stage 3. The shift of *F. hygrometrica* BOP ortholog expression to intermediate developmental stages when the capsule is not yet differentiated (stage 2) and differentiation begins (stage 3) could point to a potential role of these genes in establishing the boundary tissues that will later allow detachment of the operculum and formation of peristome teeth. However, BOP transcription factors are also involved in various other developmental and physiological processes in plants (Khan et al., 2014) and no apparent phenotypes in the sporophyte were reported in mutant analyses of PpBOP genes (Hata et al., 2019). Therefore, functional studies of BOP co-orthologs in *F. hygrometrica* are necessary to determine if these genes play a role in capsule tissue patterning that has been lost in funarioid and potentially also other species with a reduced sporophyte phenotype.

We also identified a pair of MIKC^C-type MADS-box gene orthologs in *F. hygrometrica* (Fh_2640) and *P. patens* (Pp3c14_14900), which differ in their expression profile throughout sporophyte development. Fh_2640 shows an expression peak during late stages of sporophyte development (stage 3 and 4) whereas Pp3c14_14900 is primarily expressed during the elongation phase in stages 2 and 3. This shift in expression likely represents a change in the regulatory context of the orthologs, but whether this change is also accompanied by a functional diversification is currently unclear. MIKC^C-type MADS-box genes are both expressed in the gametophyte and sporophyte phases in *P. patens* and are redundant negative regulators of cell division and growth in the gametophyte as well as sperm formation (Koshimizu et al., 2018). Their function in sporophyte development of mosses is poorly understood and debated. Koshimizu et al. (2018)

could not identify a phenotypic effect in knockout lines on sporophyte development in *P. patens* although all six MIKC^C MADS-box genes were strongly expressed in the sporophyte phase. By contrast, other studies reported well-recognizable sporophyte phenotypes when down regulating some MIKC^C genes in *P. patens* (Tanabe et al., 2003; Singer et al., 2007). We speculate that the MIKC^C-type MADS-box genes may also be important regulators of sporophyte development, potentially as negative regulators of cell division, growth and tissue patterning. The temporal expression shift of the *F. hygrometrica* ortholog to later stages of development may play a significant yet currently unknown role in the evolution of the more elaborate sporophyte morphology. This assumption is supported by the observations that down regulation of MIKC^C-type genes in *P. patens* led to abnormal swelling of the capsule (Singer et al., 2007) and that fluorescent reporter lines in *P. patens* showed that MIKC^C-type MADS-box genes have specific and complementary expression patterns throughout the development of the sporophyte. Altogether, it is possible that temporal expression shifts of MIKC^C-type MADS-box genes have contributed to the morphological diversification of sporophytes in Funarioid mosses, which parallels the observation that MIKC^C-type genes are key regulators in various aspects of sporophyte development including the diversification of flowers in angiosperms (Gramzow and Theissen, 2010). Nevertheless, sporophytic function of the MIKC^C-type genes in *P. patens* and *F. hygrometrica* must be determined in future studies on sporophyte development.

Implications for the Evolution of Sporophytic Characters in Mosses

The sporophyte is responsible for the production of spores, which will initiate the development of a new free-living gametophytic generation. Transformations of the sporophyte during the diversification of mosses affect a broad spectrum of traits, such as spore numbers, asymmetry of the capsule, mode of dehiscence, length of the seta, presence of stomata, responses of the capsule wall to dehydration, or architecture of the peristome controlling spore release (Crum, 2001). All these modifications may have a direct impact on fitness and are thus likely under natural selection (Vitt, 1981; Rose et al., 2016). Although some mosses have atypically amplified sporophytic traits [e.g., entomophilous Splachnaceae (Marino et al., 2009), or *Buxbaumia* (Crum, 2001)], and the macroevolutionary trend is one of increased complexity, repeated reductions through loss of complexity is rampant across the diversification of mosses, leading to capsules being immersed among vegetative leaves, lacking a complete peristome, stomata, or an operculum. Reduced sporophytic architecture characterizes many lineages distributed along a decreasing humidity gradient, and is common among epiphytic mosses, and particularly among xerophytic or short-lived annual mosses (Vitt, 1981; Rose et al., 2016). Finally, given the phylogenetic distribution of taxa with reduced morphology among congeners with more complex architectures, such as the three species traditionally referred to as *Physcomitrella* (highly reduced sporophytes) scattered among species of *Physcomitrium* (with more elaborate sporophytes; Medina et al., 2019), reduction may result from repeated breakdown or shut-down of complex traits, and thus convergence may be readily achieved. Reduced sporophyte complexity of *P. patens* may be easily achieved by loss or reduced expression of particular *F. hygrometrica* genes involved in sporophyte development. This is in line with previous observations that temporal and/or spatial shifts in gene expression have significantly contributed to morphological diversification both in animal and plant systems (Frankel et al., 2011; Ichihashi et al., 2014; Rast-Somssich et al., 2015; Das Gupta and Tsiantis, 2018).

Our comparative study of transcriptomic profiles of developing sporophytes in *Physcomitrium patens* and *Funaria hygrometrica* provides the very first insights into the potential genetic tools shaping sporophyte morphologies. Our observations reveal that species-specific genes are preferentially expressed during sporophyte development in both species. While the actual function of these genes is poorly known, they may be involved in meiotic processes, physiological processes such as those related to endohydric, or metabolic pathways such as for components of the cuticle (Koch et al., 2009; Budke and Goffinet, 2016). Contrasts in genic expression levels suggest also that differential length of the seta and the gain/loss of a specific opening mechanism can be achieved relatively easily by heterochronic expression of major developmental regulators, and hence that sporophyte reduction can occur by temporal and/or spatial reprogramming of expression of conserved regulatory modules. Whether convergent evolution of reduced sporophyte morphologies is driven by similar molecular mechanisms across unrelated moss lineages must await similar comparative studies from independent lineages. Ultimately, once the critical genetic traits underlying reduction are identified, the hypothesis on their irreversibility/reversibility can be tested. Together these advances would be essential to further our understanding of trends in the diversification of the moss sporophyte, and their integration in systematic concepts reflecting the relationships among mosses with highly contrasted morphologies.

DATA AVAILABILITY STATEMENT

RNA-seq data generated for this study have been submitted to the European Nucleotide Archive (ENA) under study accession number PRJEB36328. The draft genome sequence of *F. hygrometrica,* assembled transcripts and their protein translations are provided on figshare (https://doi.org/10.6084/m9.figshare.11663892.v1).

AUTHOR CONTRIBUTIONS

AK, PS, and BG conceptualized and designed the study. MW and ZB grew *F. hygrometrica* sporophytes and gametophytes, respectively, extracted RNA and prepared libraries. MR and AN collected *P. patens* sporophytes, extracted RNA, and prepared

libraries. AK and PS analyzed the data and wrote the first version of the manuscript. All authors corrected and revised the final version of the manuscript.

FUNDING

This work was supported by the URPP Evolution in Action of the University of Zurich (PS), the Forschungskredit of the University of Zurich (MR, MW, and AK), grants of the Swiss National Science Foundation (PS 160004 and 131726), the Georges and Antoine Claraz Foundation (AK, AN, MW, and PS), Plant Fellows Postdoctoral Fellowship (MR), the Foundation of German Business (sdw) (AN), and NSF grant DEB-1753811 (BG).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00747/full#supplementary-material

DATA SHEET S1 | This data can be found at the link: https://doi.org/10.6084/m9.figshare.11663892.v1.

REFERENCES

Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. *Nat. Rev. Genet.* 17, 379–391.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bowles, A. M. C., Bechtold, U., and Paps, J. (2020). The origin of land plants is rooted in two bursts of genomic novelty. *Curr. Biol.* 30, 530–536.e6. doi: 10.1016/j.cub.2019.11.090

Bowman, J. L., Briginshaw, L. N., and Florent, S. N. (2019). Evolution and co-option of developmental regulatory networks in early land plants. *Curr. Top. Dev. Biol.* 131, 35–53. doi: 10.1016/bs.ctdb.2018.10.001

Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017). Insights into land plant evolution garnered from the marchantia polymorpha genome. *Cell* 171, 287–304.e6. doi: 10.1016/j.cell.2017.09.030

Budke, J. M., and Goffinet, B. (2016). Comparative cuticle development reveals taller sporophytes are covered by thicker calyptra cuticles in mosses. *Front. Plant Sci.* 7:832. doi: 10.3389/fpls.2016.00832

Budke, J. M., Goffinet, B., and Jones, C. S. (2012). The cuticle on the gametophyte calyptra matures before the sporophyte cuticle in the moss *Funaria hygrometrica* (Funariaceae). *Am. J. Bot.* 99, 14–22. doi: 10.3732/ajb.1100311

Budke, J. M., Jones, C. S., and Goffinet, B. (2007). Development of the enigmatic peristome of *Timmia megapolitana* (Timmiaceae; Bryophyta). *Am. J. Bot.* 94, 460–467. doi: 10.3732/ajb.94.3.460

Buendía-Monreal, M., and Gillmor, C. S. (2018). The times they are a-changin': heterochrony in plant development and evolution. *Front. Plant Sci.* 9:1349. doi: 10.3389/fpls.2018.01349

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

Chen, S., Krinsky, B. H., and Long, M. (2013). New genes as drivers of phenotypic evolution. *Nat. Rev. Genet.* 14, 645–660. doi: 10.1038/nrg3521

Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., et al. (2019). Genomes of subaerial zygnematophyceae provide insights into land plant evolution. *Cell* 179, 1057–1067.e6. doi: 10.1016/j.cell.2019.10.019

Clark, J. W., and Donoghue, P. C. J. (2018). Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* 23, 933–945. doi: 10.1016/j.tplants.2018.07.006

Coudert, Y., Novák, O., and Harrison, C. J. (2019). A KNOX-cytokinin regulatory module predates the origin of indeterminate vascular plants. *Curr. Biol.* 29, 2743–2750.e3. doi: 10.1016/j.cub.2019.06.083

Crum, H. A. (2001). *Structural Diversity of Bryophytes*. Ann Arbor, MI: University of Michigan Herbarium.

Das Gupta, M., and Tsiantis, M. (2018). Gene networks and the evolution of plant morphology. *Curr. Opin. Plant Biol.* 45, 82–87. doi: 10.1016/j.pbi.2018.05.011

Davidson, E. H. (2010). Emerging properties of animal gene regulatory networks. *Nature* 468, 911–920. doi: 10.1038/nature09645

Davidson, E. H., and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* 311, 796–800. doi: 10.1126/science.1113832

Di Ruocco, G., Bertolotti, G., Pacifici, E., Polverari, L., Tsiantis, M., Sabatini, S., et al. (2018). Differential spatial distribution of miR165/6 determines variability in plant root anatomy. *Development* 145:dev153858. doi: 10.1242/dev.153858

Emms, D. M., and Kelly, S. (2018). OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *bioRxiv*[Preprint]. doi: 10.1101/466201

Erwin, D. H., and Davidson, E. H. (2009). The evolution of hierarchical gene regulatory networks. *Nat. Rev. Genet.* 10, 141–148. doi: 10.1038/nrg2499

Fife, A. J. (1982). *A Generic Revision Of The Funariaceae (Bryophyta: Musci)*. Ann Arbor, MI: University Of Michigan.

Fisher, C. R., Wegrzyn, J. L., and Jockusch, E. L. (2019). Co-option of wing-patterning genes underlies the evolution of the treehopper helmet. *Nat. Ecol. Evol.* 4, 250–260. doi: 10.1038/s41559-019-1054-4

Frankel, N., Erezyilmaz, D. F., McGregor, A. P., Wang, S., Payre, F., and Stern, D. L. (2011). Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* 474, 598–603. doi: 10.1038/nature10200

French, J. C., and Paolillo, D. J. (1975). Effect of exogenously supplied growth regulators on intercalary meristematic activity and capsule expansion in *Funaria*. *Bryologist* 78:431. doi: 10.2307/3242165

French, J. C., and Paolillo, D. J. (1976). Effect of the calyptra on intercalary meristematic activity in the sporophyte of *Funaria* (Musci). *Am. J. Bot.* 63, 492–498. doi: 10.1002/j.1537-2197.1976.tb11838.x

Galis, F., Metz, J., and van Alphen, J. M. J. (2018). Development and evolutionary constraints in animals. *Annu. Rev. Ecol. Evol. Syst.* 49, 499–522. doi: 10.1146/annurev-ecolsys-110617-062339

Furumizu, C., Alvarez, J. P., Sakakibara, K., and Bowman, J. L. (2015). Antagonistic roles for KNOX1 and KNOX2 genes in patterning the land plant body plan following an ancient gene duplication. *PLoS Genet.* 11:e1004980. doi: 10.1371/journal.pgen.1004980

Garner, D., and Paolillo, D. J. (1973). A time-course of sporophyte development in *Funaria hygrometrica* hedw. *Bryologist* 76, 356–360.

Geuten, K., and Coenen, H. (2013). Heterochronic genes in plant evolution and development. *Front. Plant Sci.* 4:381. doi: 10.3389/fpls.2013.00381

Glassford, W. J., Johnson, W. C., Dall, N. R., Smith, S. J., Liu, Y., Boll, W., et al. (2015). Co-option of an ancestral hox-regulated network underlies a recently evolved morphological novelty. *Dev. Cell* 34, 520–531. doi: 10.1016/j.devcel.2015.08.005

Goffinet, B., and Buck, W. (2013). "The evolution of body form in bryophytes," in *Annual Plant Reviews*, Vol. 45, eds B. A. Ambrose and M. Purugganan (Chichester: John Wiley & Sons, Ltd), 51–90. doi: 10.1002/9781118305881.ch2

Goffinet, B., Buck, W. R., and Shaw, A. J. (2009). Morphology, anatomy, and classification of the Bryophyta. *Bryophyte Biol.* 2, 55–138. doi: 10.1017/cbo9780511754807.003

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

Gramzow, L., and Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biol.* 11:214. doi: 10.1186/gb-2010-11-6-214

Grossmann, S., Bauer, S., Robinson, P. N., and Vingron, M. (2007). Improved detection of overrepresentation of gene-ontology annotations with parent child analysis. *Bioinformatics* 23, 3024–3031. doi: 10.1093/bioinformatics/btm440

Gujas, B., Alonso-Blanco, C., and Hardtke, C. S. (2012). Natural *Arabidopsis* brx loss-of-function alleles confer root adaptation to acidic soil. *Curr. Biol.* 22, 1962–1968. doi: 10.1016/j.cub.2012.08.026

Haig, D. (2013). Filial mistletoes: the functional morphology of moss sporophytes. *Ann. Bot.* 111, 337–345. doi: 10.1093/aob/mcs295

Halfon, M. S. (2017). Perspectives on gene regulatory network evolution. *Trends Genet.* 33, 436–447. doi: 10.1016/j.tig.2017.04.005

Hata, Y., Naramoto, S., and Kyozuka, J. (2019). BLADE-ON-PETIOLE genes are not involved in the transition from protonema to gametophore in the moss *Physcomitrella patens*. *J. Plant Res.* 132, 617–627. doi: 10.1007/s10265-019-01132-8

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417–441. doi: 10.1037/h0071325

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., et al. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43, 476–481. doi: 10.1038/ng.807

Ichihashi, Y., Aguilar-Martínez, J. A., Farhi, M., Chitwood, D. H., Kumar, R., Millon, L. V., et al. (2014). Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2616–E2621. doi: 10.1073/pnas.1402835111

Jill Harrison, C. (2017). Development and genetics in the evolution of land plant body plans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20150490. doi: 10.1098/rstb.2015.0490

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982

Jonathan Shaw, A., and Goffinet, B. (2000). *Bryophyte Biology*. Cambridge: Cambridge University Press.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20, 1313–1326. doi: 10.1101/gr.101386.109

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., and Bosch, T. C. G. (2009). More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* 25, 404–413. doi: 10.1016/j.tig.2009.07.006

Khan, M., Xu, H., and Hepworth, S. R. (2014). BLADE-ON-PETIOLE genes: setting boundaries in development and defense. *Plant Sci.* 215-216, 157–171. doi: 10.1016/j.plantsci.2013.10.019

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kim, J., Chun, J.-P., and Tucker, M. L. (2019). Transcriptional regulation of abscission zones. *Plants* 8:E154. doi: 10.3390/plants8060154

Kim, J. H. (2019). *Genome Data Analysis*. Berlin: Springer.

Koch, K., Frahm, J.-P., and Pollawatn, R. (2009). The cuticle of the *Buxbaumia viridis* sporophyte. *Flora Morphol. Distrib. Funct. Ecol. Plants* 204, 34–39. doi: 10.1016/j.flora.2007.11.007

Kolde, R., and Vilo, J. (2015). GOsummaries: an R package for visual functional annotation of experimental data. *F1000Res.* 4:574. doi: 10.12688/f1000research.6925.1

Koshimizu, S., Kofuji, R., Sasaki-Sekimoto, Y., Kikkawa, M., Shimojima, M., Ohta, H., et al. (2018). Physcomitrella MADS-box genes regulate water supply and sperm movement for fertilization. *Nat. Plants* 4, 36–45. doi: 10.1038/s41477-017-0082-9

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., et al. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* 47, D807–D811. doi: 10.1093/nar/gky1053

Kumar, L., and Futschik, M. E. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* 2, 5–7. doi: 10.6026/97320630002005

Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C., et al. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* 105, 348–363. doi: 10.1002/ajb2.1060

Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801

Lee, H., Golicz, A. A., Bayer, P. E., Jiao, Y., Tang, H., Paterson, A. H., et al. (2016). The genome of a southern hemisphere seagrass species (*Zostera muelleri*). *Plant Physiol.* 172, 272–283. doi: 10.1104/pp.16.00868

Lemmon, Z. H., Park, S. J., Jiang, K., Van Eck, J., Schatz, M. C., and Lippman, Z. B. (2016). The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Res.* 26, 1676–1686. doi: 10.1101/gr.207837.116

Lenser, T., and Theißen, G. (2013). Conservation of fruit dehiscence pathways between *Lepidium campestre* and *Arabidopsis thaliana* sheds light on the regulation of INDEHISCENT. *Plant J.* 76, 545–556. doi: 10.1111/tpj.12321

Li, Z.-W., Chen, X., Wu, Q., Hagmann, J., Han, T.-S., Zou, Y.-P., et al. (2016). On the origin of De Novo genes in *Arabidopsis thaliana* populations. *Genome Biol. Evol.* 8, 2190–2202. doi: 10.1093/gbe/evw164

Liu, Y., Budke, J. M., and Goffinet, B. (2012). Phylogenetic inference rejects sporophyte based classification of the Funariaceae (Bryophyta): rapid radiation suggests rampant homoplasy in sporophyte evolution. *Mol. Phylogenet. Evol.* 62, 130–145. doi: 10.1016/j.ympev.2011.09.010

Love, M. I, Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. doi: 10.1186/2047-217X-1-18

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823–828. doi: 10.1126/science.1215040

Marino, P., Raguso, R., and Goffinet, B. (2009). The ecology and evolution of fly dispersed dung mosses (Family Splachnaceae): manipulating insect behaviour through odour and visual cues. *Symbiosis* 47, 61–76. doi: 10.1007/bf03182289

McKim, S. M., Stenvik, G.-E., Butenko, M. A., Kristiansen, W., Cho, S. K., Hepworth, S. R., et al. (2008). The BLADE-ON-PETIOLE genes are essential for abscission zone formation in *Arabidopsis*. *Development* 135, 1537–1546. doi: 10.1242/dev.012807

Medina, R., Johnson, M., Liu, Y., Wilding, N., Hedderson, T. A., Wickett, N., et al. (2018). Evolutionary dynamism in bryophytes: phylogenomic inferences confirm rapid radiation in the moss family Funariaceae. *Mol. Phylogenet. Evol.* 120, 240–247. doi: 10.1016/j.ympev.2017.12.002

Medina, R., Johnson, M. G., Liu, Y., Wickett, N. J., Shaw, A. J., and Goffinet, B. (2019). Phylogenomic delineation of *Physcomitrium* (Bryophyta: Funariaceae) based on targeted sequencing of nuclear exons and their flanking regions rejects the retention of *Physcomitrella*, *Physcomitridium* and *Aphanorrhegma*. *J. Syst. Evol.* 57, 404–417. doi: 10.1111/jse.12516

Nachman, M. W., Hoekstra, H. E., and D'Agostino, S. L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5268–5273. doi: 10.1073/pnas.0431157100

Nadeau, N. J., Pardo-Diaz, C., Whibley, A., Supple, M. A., Saenko, S. V., Wallbank, R. W. R., et al. (2016). The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature* 534, 106–110. doi: 10.1038/nature17961

Olsen, J. L., Rouzé, P., Verhelst, B., Lin, Y.-C., Bayer, T., Collen, J., et al. (2016). The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* 530, 331–335. doi: 10.1038/nature16548

Olson, B. J., and Nedelcu, A. M. (2016). Co-option during the evolution of multicellular and developmental complexity in the volvocine green algae. *Curr. Opin. Genet. Dev.* 39, 107–115. doi: 10.1016/j.gde.2016.06.003

Olson, M. V. (1999). When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* 64, 18–23. doi: 10.1086/302219

One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* 6, 119–127. doi: 10.1038/nrg1523

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinform.* Chapter 3, Unit3.1. doi: 10.1002/0471250953.bi0301s42

Pederson, E. R. A., Warshan, D., and Rasmussen, U. (2019). Genome sequencing of *Pleurozium schreberi*: the assembled and annotated draft genome of a pleurocarpous feather moss. *G3* 9, 2791–2797. doi: 10.1534/g3.119.400279

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Pham, T., Day, S. M., Glassford, W. J., Williams, T. M., and Rebeiz, M. (2017). The evolutionary origination of a novel expression pattern through an extreme heterochronic shift. *Evol. Dev.* 19, 43–55. doi: 10.1111/ede.12215

Pires, N. D., and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 508–518. doi: 10.1098/rstb.2011.0252

Porebski, S., Bailey, L. G., and Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15, 8–15. doi: 10.1007/bf02772108

Proost, S., Van Bel, M., Sterck, L., Billiau, K., Van Parys, T., Van de Peer, Y., et al. (2009). PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* 21, 3718–3731. doi: 10.1105/tpc.109.071506

Rast-Somssich, M. I., Broholm, S., Jenkins, H., Canales, C., Vlad, D., Kwantes, M., et al. (2015). Alternate wiring of a KNOXI genetic network underlies differences in leaf development of *A. thaliana* and *C. hirsuta*. *Genes Dev.* 29, 2391–2404. doi: 10.1101/gad.269050.115

Rebeiz, M., and Tsiantis, M. (2017). Enhancer evolution and the origins of morphological novelty. *Curr. Opin. Genet. Dev.* 45, 115–123. doi: 10.1016/j.gde.2017.04.006

Reski, R., and Abel, W. O. (1985). Induction of budding on chloronemata and caulonemata of the moss, *Physcomitrella patens*, using isopentenyladenine. *Planta* 165, 354–358. doi: 10.1007/BF00392232

Rose, J. P., Kriebel, R., and Sytsma, K. J. (2016). Shape analysis of moss (Bryophyta) sporophytes: Insights into land plant evolution. *Am. J. Bot.* 103, 652–662. doi: 10.3732/ajb.1500394

Rosin, F. M., and Kramer, E. M. (2009). Old dogs, new tricks: regulatory evolution in conserved genetic modules leads to novel morphologies in plants. *Dev. Biol.* 332, 25–35. doi: 10.1016/j.ydbio.2009.05.542

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng. Design Select.* 12, 85–94. doi: 10.1093/protein/12.2.85

Sakakibara, K., Reisewitz, P., Aoyama, T., Friedrich, T., Ando, S., Sato, Y., et al. (2014). WOX13-like genes are required for reprogramming of leaf and protoplast cells into stem cells in the moss *Physcomitrella patens*. *Development* 141, 1660–1670. doi: 10.1242/dev.097444

Sakamoto, K., Onimaru, K., Munakata, K., Suda, N., Tamura, M., Ochi, H., et al. (2009). Heterochronic shift in Hox-mediated activation of sonic hedgehog leads to morphological changes during fin development. *PLoS One* 4:e5121. doi: 10.1371/journal.pone.0005121

Sano, R., Juárez, C. M., Hass, B., Sakakibara, K., Ito, M., Banks, J. A., et al. (2005). KNOX homeobox genes potentially have similar function in both diploid unicellular and multicellular meristems, but not in haploid meristems. *Evol. Dev.* 7, 69–78. doi: 10.1111/j.1525-142x.2005.05008.x

Schlötterer, C. (2015). Genes from scratch–the evolutionary fate of de novo genes. *Trends Genet.* 31, 215–219. doi: 10.1016/j.tig.2015.02.007

Schwämmle, V., and Jensen, O. N. (2010). A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinformatics* 26, 2841–2848. doi: 10.1093/bioinformatics/btq534

Schwartz, O. M. (1994). The development of the peristome-forming layers in the Funariaceae. *Int. J. Plant Sci.* 155, 640–657. doi: 10.1086/297204

Schwartz, O. M. (1997). Apical cell segmentation and its relationship to the peristome-forming layers in the Funariaceae. *Int. J. Plant Sci.* 158, 236–248. doi: 10.1086/297435

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). "BUSCO: assessing genome assembly and annotation completeness," in *Gene Prediction: Methods and Protocols*, ed. M. Kollmar (New York, NY: Springer), 227–245. doi: 10.1007/978-1-4939-9173-0_14

Shaw, A. J., Anderson, L. E., and Mishler, B. D. (2000). Paedomorphic sporophyte development in *Bruchia flexuosa* (Bruchiaceae). *Bryologist* 103, 147–155. doi: 10.1639/0007-2745(2000)103[0147:psdibf]2.0.co;2

Sicard, A., Thamm, A., Marona, C., Lee, Y. W., Wahl, V., Stinchcombe, J. R., et al. (2014). Repeated evolutionary changes of leaf morphology caused by mutations to a homeobox gene. *Curr. Biol.* 24, 1880–1886. doi: 10.1016/j.cub.2014.06.061

Singer, S. D., Krogan, N. T., and Ashton, N. W. (2007). Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. *Plant Cell Rep.* 26, 1155–1169. doi: 10.1007/s00299-007-0312-0

Soltis, P. S., and Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* 30, 159–165. doi: 10.1016/j.pbi.2016.03.015

Suarez-Gonzalez, A., Lexer, C., and Cronk, Q. C. B. (2018). Adaptive introgression: a plant perspective. *Biol. Lett.* 14:20170688. doi: 10.1098/rsbl.2017.0688

Sun, B., Looi, L.-S., Guo, S., He, Z., Gan, E.-S., Huang, J., et al. (2014). Timing mechanism dependent on cell division is invoked by Polycomb eviction in plant stem cells. *Science* 343:1248559. doi: 10.1126/science.1248559

Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C., et al. (2018). Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat. Commun.* 9:2683. doi: 10.1038/s41467-018-04721-8

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800. doi: 10.1371/journal.pone.0021800

Tanabe, Y., Uchida, M., Hasebe, M., and Ito, M. (2003). Characterization of the *Selaginella remotifolia* MADS-box gene. *J. Plant Res.* 116, 71–75.

True, J. R., and Carroll, S. B. (2002). Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* 18, 53–80. doi: 10.1146/annurev.cellbio.18.020402.140619

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26

Verd, B., Monk, N. A., and Jaeger, J. (2019). Modularity, criticality, and evolvability of a developmental gene regulatory network. *eLife* 8:42832. doi: 10.7554/eLife.42832

Vitt, D. H. (1981). Adaptive modes of the moss sporophyte. *Bryologist* 84, 166–186.

Wagner, G. P., and Lynch, V. J. (2010). Evolutionary novelties. *Curr. Biol.* 20, R48–R52. doi: 10.1016/j.cub.2009.11.010

Whitewoods, C. D., Cammarata, J., Nemec Venza, Z., Sang, S., Crook, A. D., Aoyama, T., et al. (2018). CLAVATA was a genetic novelty for the morphological innovation of 3D growth in land plants. *Curr. Biol.* 28, 2365–2376.e5. doi: 10.1016/j.cub.2018.05.068

Whitewoods, C. D., Gonçalves, B., Cheng, J., Cui, M., Kennaway, R., Lee, K., et al. (2020). Evolution of carnivorous traps from planar leaves through simple shifts in gene expression. *Science* 367, 91–96. doi: 10.1126/science.aay5433

Wickell, D. A., and Li, F.-W. (2020). On the evolutionary significance of horizontal gene transfers in plants. *New Phytol.* 225, 113–117. doi: 10.1111/nph.16022

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer Science & Business Media.

Wu, X.-M., Yu, Y., Han, L.-B., Li, C.-L., Wang, H.-Y., Zhong, N.-Q., et al. (2012). The tobacco BLADE-ON-PETIOLE2 gene mediates differentiation of the corolla abscission zone by controlling longitudinal cell expansion. *Plant Physiol.* 159, 835–850. doi: 10.1104/pp.112.193482

Xu, Y.-C., Niu, X.-M., Li, X.-X., He, W., Chen, J.-F., Zou, Y.-P., et al. (2019). Adaptation and phenotypic diversification in *Arabidopsis* through loss-of-function mutations in protein-coding genes. *Plant Cell* 31, 1012–1025. doi: 10.1105/tpc.18.00791

Zhang, L., Ren, Y., Yang, T., Li, G., Chen, J., Gschwend, A. R., et al. (2019). Rapid evolution of protein diversity by de novo origination in *Oryza*. *Nat. Ecol. Evol.* 3, 679–690. doi: 10.1038/s41559-019-0822-5

Zhao, Y., Zhang, T., Broholm, S. K., Tähtiharju, S., Mouhu, K., Albert, V. A., et al. (2016). Evolutionary co-option of floral meristem identity genes for patterning of the flower-like asteraceae inflorescence. *Plant Physiol.* 172, 284–296. doi: 10.1104/pp.16.00779

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 19 June 2020
doi: 10.3389/fpls.2020.00858

Are All Paraphyllia the Same?

Ulyana N. Spirina[1,2*], *Tatiana V. Voronkova*[2] *and Michael S. Ignatov*[2,3]

[1] Faculty of Biology, Tver State University, Tver, Russia, [2] Tsitsin Main Botanical Garden, Russian Academy of Sciences, Moscow, Russia, [3] Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia

Moss paraphyllia, the trichome-like or foliose structures on moss stem surfaces, are usually treated as epidermal outgrowths. However, in some taxa of the moss families Leskeaceae, Neckeraceae, and Amblystegiaceae their distribution along the stem is consistently correlated with parts of the stem surface near branch primordia. In other moss families, Climaciaceae, Hylocomiaceae, and Pseudoleskeaceae the specific paraphyllia-generating epidermal layer produces paraphyllia evenly all along the stem. These results suggest that there are at least two different types of regulation of paraphyllia development; however, both of them may be involved in the morphogenesis of paraphyllia in some families, for example in the Thuidiaceae. Exogenous abscisic acid treatment consistently increases the number of paraphyllia of the *Leskea*-type, and it also induces the development of proximal branch leaves that normally do not develop a lamina above the stem surface. This fact supports conclusions regarding the homology of the *Leskea*-type of paraphyllia with leaves.

Keywords: paraphyllia, *Leskea*, abscisic acid, morphogenesis, pseudoparaphyllia, branch primordia, fluridone

OPEN ACCESS

Edited by:
Bernard Goffinet,
University of Connecticut,
United States

Reviewed by:
Chun-Lei Xiang,
Kunming Institute of Botany,
Chinese Academy of Sciences, China
Paulo Camara,
University of Brasilia, Brazil

***Correspondence:**
Ulyana N. Spirina
ulayspirina@mail.ru

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *21 February 2020*
Accepted: *27 May 2020*
Published: *19 June 2020*

Citation:
Spirina UN, Voronkova TV and Ignatov MS (2020) Are All Paraphyllia the Same? Front. Plant Sci. 11:858. doi: 10.3389/fpls.2020.00858

INTRODUCTION

Mosses are small plants, and their morphology remains incompletely understood despite a long history of studies. Among the most controversially interpreted features are foliose and filamentose appendages on the stem surface in pleurocarpous mosses, commonly called paraphyllia and pseudoparaphyllia. An interest in their structure has been recently revived after molecular phylogenetic studies showed that many morphological characters used in "classical taxonomy" are enormously homoplasious and can no longer be used as key diagnostic characters. The classification of pleurocarpous mosses at family level is especially flawed, as sporophyte characters have been shown to be extremely plastic during the transition to epiphytism (Huttunen et al., 2004; Huttunen et al., 2012; Hedenäs, 2012). Conversely, previously neglected features of branch primordia, and specifically the arrangement of foliose structures around branch initials, appear stable enough to serve as diagnostic characters for family-level classification (Ignatov, 1999). The terminology of these structures and interpretation of their homology, however, remain incongruent, as different manuals sometimes refer to the same cases in conflicting ways.

Paraphyllia are filamentose or narrowly lanceolate, rarely ovate, multicellular, branched or unbranched structures, densely or sparsely covering the stems or solitary (**Figure 1**). They were described and illustrated originally by Hedwig (1801), who called them "stupae" (tow), while the authors of "Bryologia Europaea" introduced the name "paraphyllia" (Bruch et al., 1836–1855), a term widely used up to now. Later it was observed that some foliose structures on the stem occur only in proximity to branch primordia, and Warnstorf (1904–1906) introduced another term, "pseudoparaphyllia," exemplified by the foliose structures around the branch primordia in *Rhynchostegium* (Brachytheciaceae) and *Rhytidiadelphus* (Hylocomiaceae). Ironically, the first world-wide comprehensive overview of pseudoparaphyllia by Ireland (1971) interpreted these two genera as lacking pseudoparaphyllia. Several attempts were made to define the distinction between paraphyllia, pseudoparaphyllia, and proximal branch leaves, and the best-

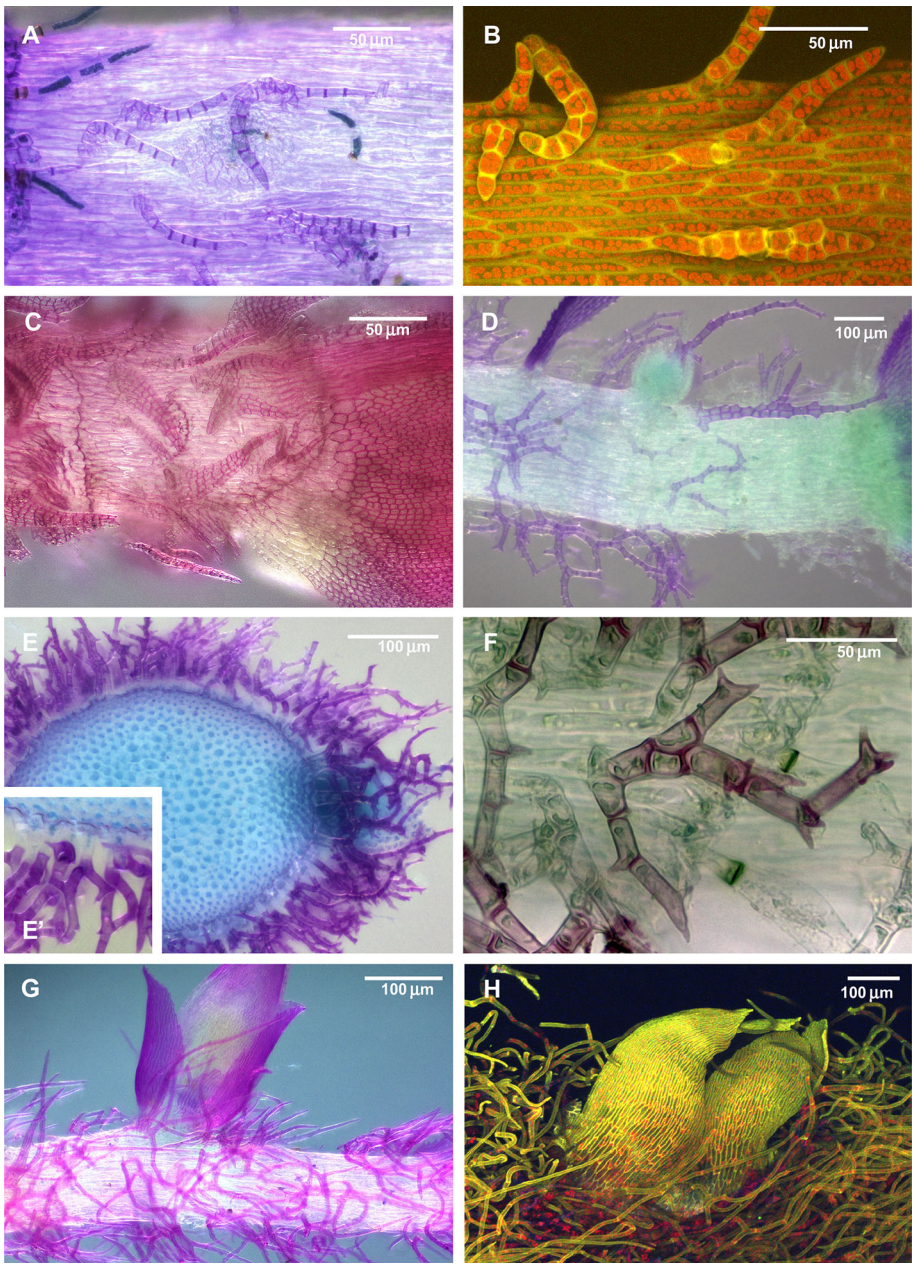


FIGURE 1 | Diversity of paraphyllia in mosses: scattered over stems **(A–C)** or appearing on the majority of cells **(E–H)** or numerous on the stem and more scattered on branches **(D)**. **(A, B)** *Leskea polycarpa*; **(C)** *Cratoneuron filicinum*; **(D)** *Thuidium tamariscinum*, secondary stem with primordium of a tertiary branch; **(E, F)** *Thuidium recognitum*, stem cross-section (E and E', close up of E) and surface view **(F)**; **(G)** *Hylocomium splendens* (Hedw.) Schimp., part of the stem with branch primordium; **(H)** *Climacium dendroides* (Hedw.) F. Weber & D. Mohr, part of the stem with branch primordium. **(A, C–G)**: LM; **(B, H)**: LSCM.

known overviews of structures around branch primordia are those by Akiyama (1990a; 1990b) and Akiyama and Nishimura (1993). These authors suggested that pseudoparaphyllia should be differentiated from proximal branch leaves, for which they proposed the terms "scale-like leaves" or "scaly leaves". Paraphyllia were treated by these authors as adventive structures (appendages) of the stem epidermis, in contrast to foliose or filamentose structures around branch primordia referred to as pseudoparaphyllia.

Paraphyllia have attracted less attention, and the most detailed overview remains a publication by Bonnot (1967), who discussed possible homologies of paraphyllia, which were treated as appendages of the stem epidermis, additional leaf-like structures, reduced leaves, trichomes, and modified protonematal structures, but always without relation to the branch primordia. Ignatov and Hedenäs (2007) challenged the latter statement showing that in *Alsia, Cratoneuron*, *Leptodon*, and *Leskea,* "paraphyllia" are more abundant near branch primordia or situated, often in pairs, at the locations of undeveloped branch primordia. These authors concluded that in the studied taxa "paraphyllia" may be assumed to be pseudoparaphyllia that have escaped for a certain distance from the branch primordia. However, this suggestion did not meet with wide acceptance.

Later it was shown that moss leaves are not always entire. Proximal branch leaves can be deeply incised, or subdivided to the base, and their parts sometimes occur spaced from each other for a certain distance, appearing to be independent structures, which have been called "compound leaves" (Spirina and Ignatov, 2008; Ignatov and Spirina, 2012; Spirina and Ignatov, 2015). **Figure 2** illustrates the variation in the structure of proximal branch leaves (the outermost leaves around the branch primordium). Most pleurocarps in their development follow the scheme shown in **Figure 2A**, with the first proximal branch leaf appearing in the "four o'clock position" (**Figures 2B–D**). A different situation occurs in some groups, e.g. *Hypnum cupressiforme* Hedw. (**Figures 2E, G, I**), where the first three branch merophytes produce leaves divided into filiform segments which look like independent structures, originating, however, from a single cell. Furthermore, the parts of such compound leaves can be somewhat displaced (**Figures 2F, H**), and ultimately stand apart from the branch initial (**Figure 2J**). The cases where the parts of compound branch leaves appear distant from the branch primordia raise questions about the distinction between paraphyllia and compound proximal branch leaves. The homology of the previously mentioned structures thus requires better definition, and this is in the main focus of the present paper. Specifically, the hypothesis that we test is whether foliose and filamentose structures on the surface of moss stems, excluding rhizoids, axillary hairs, and gemmae, can be classified in two or more groups based on their structure and distribution along the stem.

In addition to morphological studies we conducted experiments with exogenous abscisic acid (ABA). The selection of ABA is based on preliminary studies with three phytohormones, cytokinin, auxin, and ABA, on the growth of *Leskea polycarpa* Hedw. Only ABA showed an apparent effect on morphology, specifically the conspicuous increase in the number of paraphyllia after exogenous ABA application, and therefore we undertook more detailed studies.

ABA is an important compound in plants, involved in the regulation of plant growth, germination, stomatal movement, responses to abiotic stresses, as well as the stresses caused by viruses and bacteria (Zhang, 2014; Alazem and Lin, 2017; Alazem et al., 2017). Although it is widely accepted that ABA regulates stresses (Osakabe et al., 2014; Yoshida et al., 2014; Shi et al., 2015), there are still few reports that show a direct role of ABA.

Cho et al. (2009) review ABA studies in *Physcomitrella*, confirming the similarity of its physiological responses to that in seed plants. For the morphogenetic role of ABA, however, Cho et al. (2009) noted only membrane fragility after freezing, which is increased in untreated protonematal cells. There are also reports of the thickening of cell walls (Tintelnot, 2006), induction of brood cells (Goode et al., 1993), induction of tmema cells (Decker et al., 2006) and inhibition of cytokinin-induced bud formation on protonema (Christianson, 2000).

The putative importance of ABA studies to the paraphyllia/pseudoparaphyllia problem can be related to its ability to promote seed dormancy in angiosperms (Nee et al., 2016; Shu et al., 2016; Shu et al., 2017; Shuai et al., 2017) and to mediate plasmodesmata closure in the bud meristems of *Populus*, thereby inducing bud dormancy (Tylewicz et al., 2018). Kitagawa et al. (2019) demonstrated the regulation of plasmodesmata permeability in the moss *Physcomitrella patens* (Hedw.) Bruch & Schimp. and showed that the molecular trafficking through them is rapid and reversibly restricted by abscisic acid. Thus a kind of release from dormancy may be better understood using deeper ABA studies. Branch primordia in the majority of pleurocarpous moss genera have a specific variant of dormancy: branch primordia start to develop, produce small-sized foliose protective structures covering the bud, and then revert to a dormant state but are promptly released from it in the case of damage to a stem apical cell. Although ABA was never considered for dormancy in mosses, its effect on structures associated with dormancy seems to merit being tested.

MATERIALS AND METHODS

Morphological Studies

Species Sampling

Leskea polycarpa was the main focus of the present study as its paraphyllia are the easiest to count. This is a widespread epiphytic moss, common on tree trunks in the temperate zone of the Northern Hemisphere. Its habitats periodically experience drought lasting from several days to several weeks. Intensive growth of *Leskea polycarpa* in the Moscow region starts in the second half of spring. As the species was the main subject of our ABA investigation, we undertook anatomical studies by sectioning to elucidate the sub-surface structure of its branch primordia.

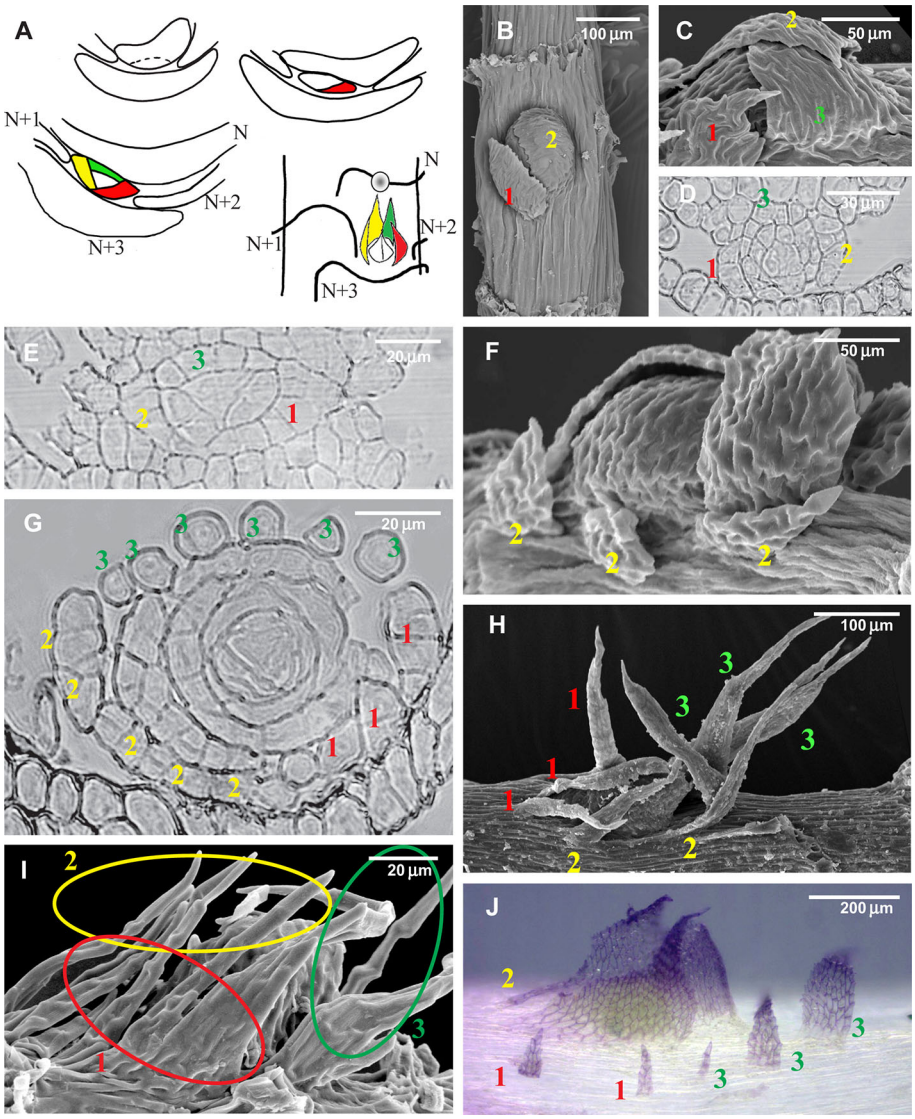


FIGURE 2 | Development of the branch primordium in mosses, illustrating variants of the foliose structures around it. **(A)** Scheme of branch primordium differentiation from leaf N and subsequent displacement towards the axil of leaf N+3 along the path between the decurrences of leaves N+1 and N+2; **(B–D)** Variants of the most common arrangement in pleurocarpous mosses of the proximal branch leaves in *Orthostichella hexasticha* (Schwägr.) W.R. Buck **(B)**, *Campylium stellatum* (Hedw.) C. Jens. **(C)** and *Amblystegium serpens* (Hedw.) Schimp. **(D)**, with the outermost leaf in "four o'clock position" [with the possible switching of left and right spirals, the "four o'clock position" may appear as the "eight o'clock position", which is still called hereafter the "four o'clock position"]; **(E, G, I)** Development of the branch primordium in *Hypnum cupressiforme*, from the initial, showing formation of merophytes from the trifacial apical cell **(E)**, then the development of leaves strongly incised from the beginning **(G)**, and finally a mature "dormant bud" surrounded by three compound leaves, each of which were traditionally called pseudoparaphyllia **(I)**; **(F, H)** Examples of compound broader proximal branch leaves in *Thamnobryum alopecurum* (Hedw.) Nieuwl. ex Gangulee **(F)** and *Alleniella complanata* (Hedw.) S. Olsson, Enroth & D. Quandt **(H)**, recognized as such due to the apparent phyllotaxis of the branch leaves; **(J)** Branch primordium of *Cratoneuron filicinum*, with first branch leaves compound and remote from the bud. **(B, C, F, H, I)**: SEM; **(J)**: LM, basic fuchsine, and picroindigocarmine.

Two other mosses studied for paraphyllia distribution along the stem and also for the effect of ABA were chosen from genera of different families: *Leptodon* (Neckeraceae) and *Cratoneuron* (Amblystegiaceae).

Genera from the following families known to have paraphyllia were re-studied with light microscopy to observe the distribution pattern of the paraphyllia: Amblystegiaceae (*Cratoneuron*, *Palustriella*), Climaciaceae (*Climacium, Pleuroziopsis*), Leskeaceae (*Leskea*), Neckeraceae (*Leptodon*, *Neckera* p.p.), Pseudoleskeaceae (*Lescuraea* s.l., and probably related to it *Rhytidiopsis*), Theliaceae (*Thelia*), Thuidiaceae (*Abietinella*, *Actinothuidium*, *Boulaya*, *Bryochenea*, *Bryonoguchia*, *Haplocladium*, *Hylocomiopsis*, *Helodium*, *Pelekium*, *Rauiella*, *Thuidium*).

We include illustrations of some examples of other taxa from previous studies, and some additional images prepared for better illustration of the structure of the paraphyllia and proximal branch leaves.

A list of taxa and studied specimens is available in **Supplementary Materials 1**.

Distribution of Paraphyllia Along the Stem

Observations of the distribution of the paraphyllia along the stem were made in *Leskea polycarpa, Leptodon smithii* D. Mohr, and *Cratoneuron filicinum* (Sull.) Spruce. The area of the stem was subdivided into three zones (**Figures 3A, B**), based on the observation that paraphyllia of *Leskea*-type occur on the stem mainly along the route of the branch primordium from "mother-leaf" N to a destination in the axil of the leaf N+3. The area between the decurrencies of leaves N+1 and N+2 has paraphyllia more often than the rest.

All leaves were removed from the shoots with thin tweezers, stems were stained with fuchsine, and photographed under a light microscope Olympus-CX41 with digital camera Infinity 2-2. Paraphyllia were counted from the images, consulting the original glass slide in uncertain cases. Stems on a glass slide were covered by another slide, allowing for observations of both sides of the stem. Plants from herbaria were studied in the same way.

Anatomical Studies

Anatomical studies were conducted with *Leskea polycarpa* to characterize the structure of the branch primordia. To prepare anatomical sections, the upper parts of shoots (with leaves removed) were fixed in a 2.5% glutaraldehyde solution for 7 days, post-fixed in a 1% OsO_4 water solution for 3 hours, washed in water. After washing, the material was dehydrated in an alcohol series (20%, 40%, 60%, 80% and 96% alcohol), alcohol–acetone mixture (1:1), and acetone for 1 hour in each solution, soaked in an acetone–resin mixture series (3:1, 1:1, 1:3) for 12, 24 and 3 hours respectively, and embedded in epon-araldite resin as recommended by the manufacturer. The resin was polymerized at 60°C for 24 hours. Serial longitudinal, transverse, and oblique sections were cut 2 μm thick with glass knives, placed on glass slides without mounting medium, stained with 0.01% berberine or its combination with DAPI and scanned under a laser scanning confocal microscope Olympus FV-1000 based on Olympus BX61, using or combining 405 and 473 nm lasers. Z-stacks of several scans were usually obtained and are presented here.

Two series of longitudinal sections and three series of transverse ones were made, and some interesting cases observed under the microscope were studied.

Supplementary Studies

The structure of branch primordia was studied for pleurocarpous species of several families, mainly for taxonomic purposes. In the present paper we illustrate the least known cases involving the complete reduction of proximal branch leaves, and their division to the base, causing them to appear to be compound.

These studies involved light microscopy (LM), scanning electron microscopy (SEM), and laser scanning confocal microscopy (LSCM).

LM analysis: paraphyllia were described with regard to their shape and structure, position, and arrangement on the stem, contrast being enhanced by fuchsine and picroindigocarmine staining.

For **SEM** analysis the shoots were fixed in a 4% glutaraldehyde solution for 7 days, incubated in a 1% OsO_4 water solution for 6 hours, washed in water, dehydrated through an ascending alcohol–acetone series, dried at the critical point, and coated with gold. Prepared shoots were observed and photographed under a LEO430 scanning electron microscope (Carl Zeiss, Germany). For some shoots SEM observations of gold-coated samples were performed with the SEM Jeol 6380, without additional preparation.

Living material was observed and photographed with **LSCM** with similar staining by berberine or berberine/DAPI, without fixation. Combinations of lasers 405, 473, and 576 were used for maximally detailed pictures. Scan series (mostly at 1024×1024 pixels, 20–70 scans) are presented here as Z-stacked images, or, in some cases, photo-galleries of parts of series.

ABA Experiments

Sampling

Observations were conducted on *Leskea polycarpa, Cratoneuron filicinum. Leptodon smithii,* and *Thuidium tamariscinum* (Hedw.) Schimp., representatives of four different families. The selection depended on the position and number of paraphyllia which are readily countable. *Thuidium tamariscinum* has very dense and numerous paraphyllia on stem and primary branches, but on secondary branches their number is comparable with that in *Leskea*; thus only secondary branches were studied for this species. Species with very dense and interwoven paraphyllia, e.g., *Hylocomium* and *Climacium* were not studied, because it is almost impossible to count paraphyllia on their stems and branches, and paraphyllia are situated on all cells.

Cultivation

Leskea polycarpa was placed in Petri dishes (diameter 9 cm) as ten tufts of ca. 2×2 cm, without separating plants into individuals. They were put on one layer of filter paper moistened with 5 ml of distilled water or an aqueous solution of ABA (Sigma, Germany) at a concentration of 0.1 μmol, 0.04 μmol, 0.01 μmol, 0.002 μmol, hereafter referred to according to the amount applied in μmol as ABA-0.5, ABA-0,2, ABA-0.05, and ABA-0.01. Petri dishes were placed in a Sanyo Environmental Test Chamber MLR-352H: temperature + 7°C/+ 12°C (night/day), light period 10 hours, PPFD - 14 $\mu molm^{-2}s^{-1}$ for 21 days.

Field Experiment

In addition to the experiments in the test chamber, we explored the effect of ABA on *Leskea in vivo* to exclude the possibility that the observed effect on paraphyllia depends on cultivation in the permanently humid condition of the Petri dish. ABA was applied directly to tufts of *Leskea polycarpa* growing in the park of the Tsitsin Main Botanical Garden in Moscow, on four trunks: two of *Acer tataricum* L., and one each of *A. platanoides* L. and *Malus sylvestris* Mill., at 1.5 m above the ground. Two tufts 5 x 5 cm of *Leskea* were moistened on each trunk: one with ABA-0.5 solution and one with distilled water. A soft painting brush was used for wetting moss turfs.

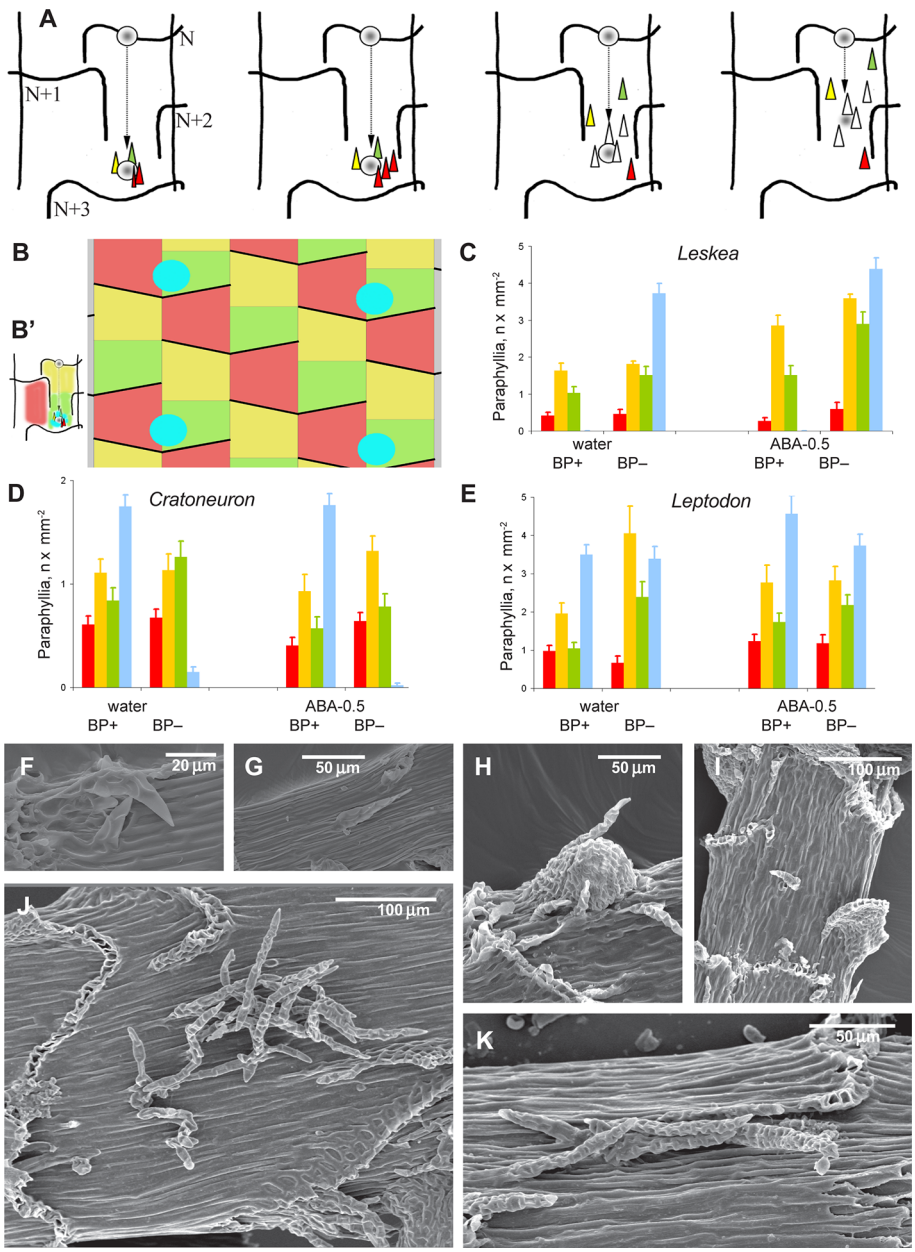


FIGURE 3 | Distribution of paraphyllia of the *Leskea*-type along the stem. **(A)** The scheme, based on observations of the branch primordium and paraphyllia in *Leskea*-type positions (cf. examples in **(F–K)**; the scheme illustrates the most common position of these paraphyllia along the path of branch primordium displacement from the "mother-leaf ", N, to the axil on the N+3 leaf (cf. also **Figure 2A**); **(B)** Scheme of whole stem surface circumference for a moss with 2/5 phyllotaxis, showing zones used for paraphyllia counts (B' helps to link these zones to the scheme in **A**) colors denote: blue is the area immediately close to the branch primordium; green is the lower half of the "internode", up to the upper level of the leaf N+2, yellow is the area above the green zone up to the leaf N, and red is the part of the stem below leaf N on the side other than where the branch primordium originated; note that leaf decurrences in the scheme are not shown for simplicity; **(C–E)**. The number of paraphyllia in experiments with three species, *Leskea polycarpa* **(C)**, *Cratoneuron filicinum* **(D)**, *Leptodon smithii* **(E)**, for the three zones of the stem as defined in the scheme in "B" and the area near branch primordia (blue); the count was performed separately for "internodes" with branch primordia (BP+) and without them (BP–); some of the plants were studied after ABA experiments, and their paraphyllia were counted separately, showing an especially large increase of paraphyllia in the yellow zone. See more data and statistics in **Supplementary Materials 2**. **(F–K)** examples of the position of foliose structures between decurrences of leaves N+1 and N+2 (cf. **Figures 2A** and **3B**), in *Hygroamblystegium varium* **(F)**: branch primordium, **(G)**: paraphyllium, *Leskea polycarpa* **(H)**: branch primordium, **(I)**: paraphyllium, *Neckera californica* **(J)** and *Leptodon smithii* **(K)**. **(F–K)**: SEM.

All turfs were situated on the same eastern side of the trunk in similar conditions. The experiment was conducted in the period following 25 April 2017, i.e. when most epiphytic mosses had resumed growth in Moscow (in the rainy and warmer days after winter).

Counting and Statistics

After 21 days of cultivation, 50 shoots from each Petri dish were selected. All leaves were removed from the stems with thin tweezers. The stems were stained with fuchsine, placed on a glass slide, and covered by another glass slide, allowing observations from both sides of the stem. The stems were examined under the light microscope (Olympus-CX41) and photographed with a digital camera (Infinity 2-2). Stem diameter was measured, and the number of paraphyllia was counted from the images. The number of paraphyllia was calculated for one square millimeter of the stem surface. Cultivated plants, plants from the field, and herbarium materials were studied in the same way.

The significance of the ABA effect was evaluated by the T-test in MS Excel (**Supplementary Materials 2**).

Fluridone Test

To make sure that it is the presence of ABA that causes changes in the morphogenesis of paraphyllia, we used a test with fluridone, an ABA biogenesis inhibitor (Popova, 1995; Klíčová et al., 2002; Velini et al., 2010; Shu et al., 2017) that may reduce the content of the free form of ABA by up to 40% (Stetsenko et al., 2015). Fluridone (Sigma, Germany) at a concentration of 15 μmol was applied in 1 ml quantities to *Leskea* cultivated with ABA and without it. The first test, which consisted of adding fluridone only once at the beginning of the cultivation period, showed inconsistent results. The absence of effect could be explained by the decomposition of fluridone as direct light destroys its molecules over a single day (Müller and Applebyki, 2011). Therefore a second test was performed with the application of fluridone three times a week (altogether nine times for 21 days). The significance of the impact of fluridone was evaluated by an ANOVA test in PAST (Hammer et al., 2001).

Brachytheciaceae Test

Brachytheciaceae is known as a family in which the branch primordia lack leaves formed from the first and second branch merophytes, which therefore appear underdeveloped, forming no foliose structures above the stem surface (Ignatov, 1999; Spirina and Ignatov, 2005). We have used ABA to check if such "hidden" leaves can develop a lamina above the stem surface. The experiment was conducted with the same protocol as in *Leskea* in the test chamber. Plants of *Brachythecium rutabulum* (Hedw.) Schimp. which were used for this study, were collected from the park in the Tsitsin Botanical Garden in Moscow.

RESULTS

Morphological Studies

Distribution of Paraphyllia Along the Stem

The evaluation of the distribution of paraphyllia along the stem was performed quantitatively using zones associated with the morphogenesis of branch primordia (**Figures 3A, B**). The number and distribution of paraphyllia in *Leptodon smithii*, *Cratoneuron filicinum*, and *Leskea polycarpa,* species from three different families, show approximately the same pattern. Paraphyllia are most abundant near the branch primordium and along the route that the primordium passes in its ontogenetic history on the path from mother leaf N (cf. also **Figures 2A** and **3A**) to the axil of the leaf N+3. The lowest number of paraphyllia occurs in the "red area" (**Figures 3C–E**), i.e. the area below a half of the leaf that cut off the cell which is later developed into the internode with the branch primordium (**Figures 2A** and **3A**).

Visual evaluation of a wide selection of taxa with paraphyllia confirms that the pattern seen in *Leskea, Cratoneuron,* and *Leptodon* is also characteristic for *Neckera menziesii* Drumm., *N. californica* Hook. & Arn., *Palustriella commutata* (Hedw.) Ochyra, *P. decipiens* (De Not.) Ochyra, *Leskea gracilescens* Hedw., *L. obtusa* Renauld & Cardot, and *Hygroamblystegium varium* (Hedw.) Mönk. (**Figures 3F–K**): either numerous or scattered paraphyllia occur in the "green" and "yellow" zones as they are delimited in **Figure 3B**.

Within the proximity to branch primordia (the "green" zone in **Figure 3D**) the arrangement of paraphyllia around the branch primordia often has a more or less clear phyllotaxis. Therefore the branch merophyte numbers to which paraphyllia belong can be interpreted (**Figures 2J** and **4A**) by comparison with classical schemes of stem morphogenesis (Berthier, 1971), partly reproduced in **Figures 2A** and **3A**. The further from branch primordia, the less regular they are, grading to an irregular arrangement, or even occasionally forming short longitudinal rows.

Anatomical Studies

Paraphyllia or compound proximal branch leaves in *Leskea polycarpa* are more numerous around branch primordia (**Figure 4A**); however, in some cases paraphyllia retain a definite phyllotaxis around a point which has no indications of a branch initial on the surface (**Figure 4B**). Transverse cross-sections made at this particular location allowed us to observe only one slightly enlarged subepidermal cell (**Figure 4D**), quite unlike the most common appearance of a branch primordium shown in **Figure 4C**. The concentration of paraphyllia around the branch primordia can also be seen in transverse sections (rectangles in **Figure 4C**). The transverse section of *Leskea* shoot close to the apex (**Figures 4E, F**) shows that the cells forming the "internode" and the branch initials (cf. **Figure 2A**) have a very broad base, which probably causes the apical and subapical branch cells to be rather indefinitely delimited from the surrounding cells. Foliose structures around the branch initial cells are only moderately tightly appressed to the apical cell and later the whole space between young stem leaves is filled by foliose structures around the branch initial cell (**Figures 4F–F**).

Paraphyllia of Different Types

Unlike the above-mentioned genera where paraphyllia are concentrated around the branch primordia, there are moss genera in which the paraphyllia are dense, situated on almost all the cells of the stem surface, or if scattered then still evenly

distributed along the stem, and usually more or less arranged in longitudinal rows. This pattern is characteristic for *Climacium* (Climaciaceae), *Hylocomiastrum*, *Hylocomium*, *Loeskeobryum*, and *Rhytidiopsis* (Hylocomiaceae) (**Figure 5**). *Pleuroziopsis,* another genus of Climaciaceae, usually has rows of unusually inflated cells, with sparse paraphyllia.

Species of the Pseudoleskeaceae also have paraphyllia in rows, but they do not cover the stem completely; sometimes they are only scattered, and there are species in the family that lack paraphyllia completely.

Species of the Thuidiaceae have stems often densely covered by paraphyllia, which are occasionally arranged in conspicuous rows (e.g. in *Bryonoguchia*), although sometimes they are very homogeneous (e.g. in *Bryochenaea, Pelekium*). At the same time, their branches are less abundantly paraphyllose, and the concentration of paraphyllia near the branch bases is obvious

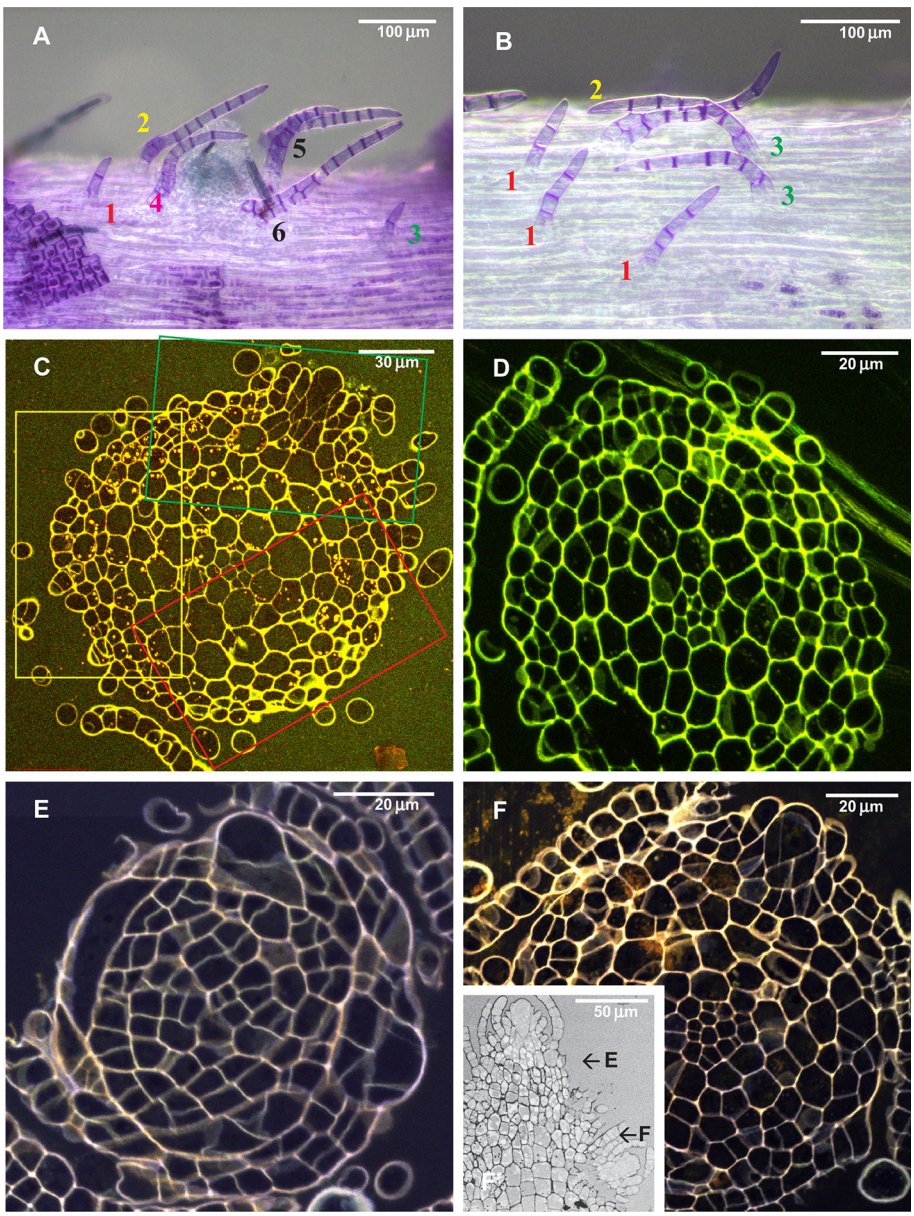


FIGURE 4 | Arrangement of paraphyllia around developed and undeveloped branch primordia of *Leskea polycarpa*, as seen on the stem **(A, B)** and in stem transverse sections **(C, D)**, and the early stages of branch primordium development **(E, F)**. **(A)** Compound proximal branch leaves ("escaped pseudoparaphyllia") around branch primordium (for numbering cf. **Figure 2**); **(B)** Paraphyllia arranged with apparent phyllotaxis around a vestigial primordium; **(C, D)** Transverse section across the "apices" of the primordium and vestigial primordium shown in "A" and "B"; **(E, F)** Transverse section of *Leskea* shoot at 46 and 78 µm from the apex, showing that the branch initials have a very broad base, which probably causes the branch itself to be rather indefinitely delimited from the surrounding cells, differentiated from the same cell at an early stage of development shown in **Figure 2A**. **(A, B)**: LM, basic fuchsine, and picroindigocarmine.

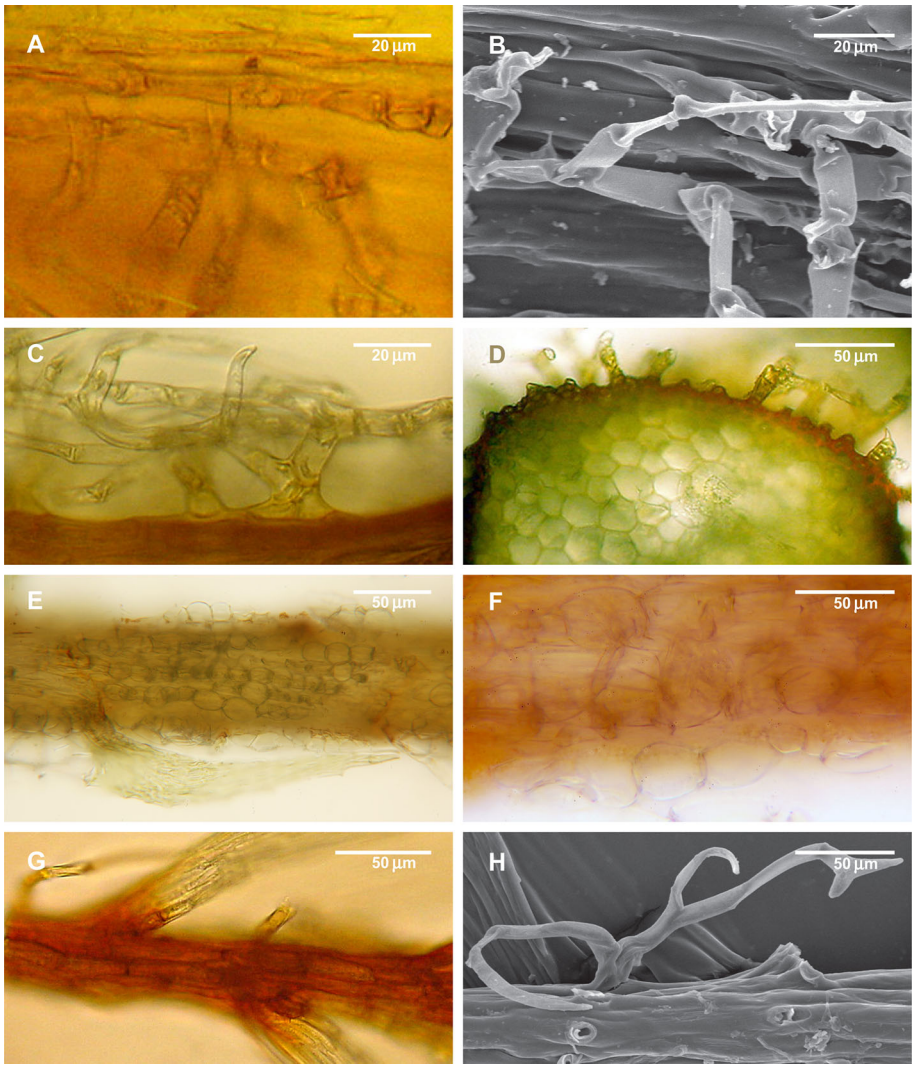


FIGURE 5 | Structure of paraphyllia of the *Climacium*-type. **(A–D)** *Climacium dendroides*, showing paraphyllia in moderately inflated longitudinal ridges; **(E–F)** *Pleuroziopsis ruthenica* (Weinm.) Kindb. ex E. Britton, longitudinal ridges of inflated cells, very rarely producing also paraphyllia; **(G, H)** *Hylocomium splendens,* branch apical part with a row of more or less inflated cells; stem surface with most of the paraphyllia removed. **(A, C–G)**: LM; **(B, H)**: SEM.

(**Figure 1D**), and their arrangement is similar to that in *Leskea*. We interpret these paraphyllia as being largely of the *Climacium*-type, but in thinner branches their development is regulated in the same way as in *Leskea*.

ABA Experiments

ABA Experiments *In Vitro*

The cultivation of *Leskea polycarpa, Cratoneuron filicinum, Leptodon smithii*, and *Thuidium tamariscinum* with ABA solutions of different concentrations *in vitro* led to an increase in the number of paraphyllia on the tips of growing shoots during the treatment period (**Figure 6**). This increase was significant for all four studied species.

In the first three species paraphyllia were counted on the stem, while in the tripinnate branched plants of *Thuidium tamariscinum* secondary branches were used, as on the stem and primary branches paraphyllia are too numerous to count.

ABA Experiments *In Vivo*

Since *Leskea polycarpa* occurs on many trees within the park area of the Tsitsin Main Botanical Garden in Moscow, we also conducted a "field experiment", moistening tufts of *Leskea* on the trunk of maple with ABA solution at the time of its growing season in spring, using the same series of concentrations as *in vitro* experiments in the test chamber. The results obtained show the same response as in cultivation experiments (**Supplementary Materials 3**).

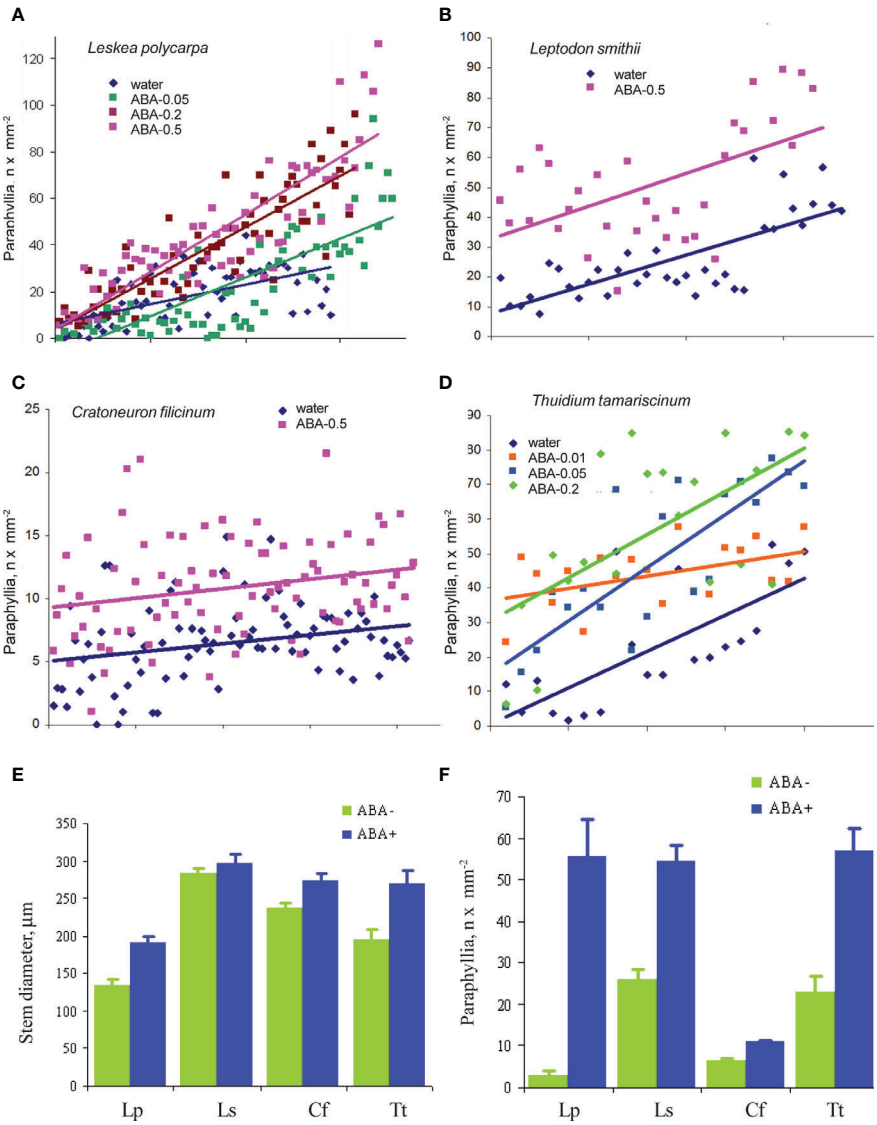


FIGURE 6 | The effect of exogenous ABA on the formation of paraphyllia in four moss species after four weeks of growth in Petri dishes with aqueous solutions of ABA. Axis Y: number of paraphyllia per mm^2. Counts of paraphyllia number per square unit are arranged along the X axis by the following method: studied stems were numbered and then sorted within each series (i.e. ABA-0.5, ABA-0.2, etc) by stem diameter, from narrowest to widest, thus the position on X axis is the position number of a given stem in its series. Such distributions allow us to compare clouds of dots in different series. **(A)** *Leskea polycarpa*; **(B)** *Leptodon smithii*; **(C)** *Cratoneuron filicinum*; **(D)** *Thuidium tamariscinum*. Color corresponds to ABA concentration; **(E)** Stem diameter from experiments shown in "A–D"; **(F)** Number of paraphyllia from experiments shown in **(A–D)**. See also **Supplementary Materials 3**.

Fluridone Test on *Leskea*

The application of fluridone (FLU) inhibited the increase in the numbers of paraphyllia by exogenous ABA (**Figure 7**). The reduction in paraphyllia numbers in the variant [ABA+ FLU+] compared with [ABA+ FLU–] was 63 and 51% in two series of experiments with plants from two populations (**Figure 7**), and was statistically significant supported by an ANOVA test (**Supplementary Materials 4**), $p < 0.001$. The effect on the stem diameter was less apparent, although significant at $p < 0.05$ (**Supplementary Materials 4**).

Brachythecium Test

Since previous experiments found that ABA induces the development of leaf-like structures on the stem in places where they are commonly absent, we undertook another test with mosses that are plastic in branch primordium development. **Figures 8A–C** show a parallel situation in the family Lembophyllaceae s.l., where there is great variation in branch primordium structure, allowing stages of reduction to be illustrated. By contrast, in Brachytheciaceae the branch primordium structure is conserved in terms of proximal

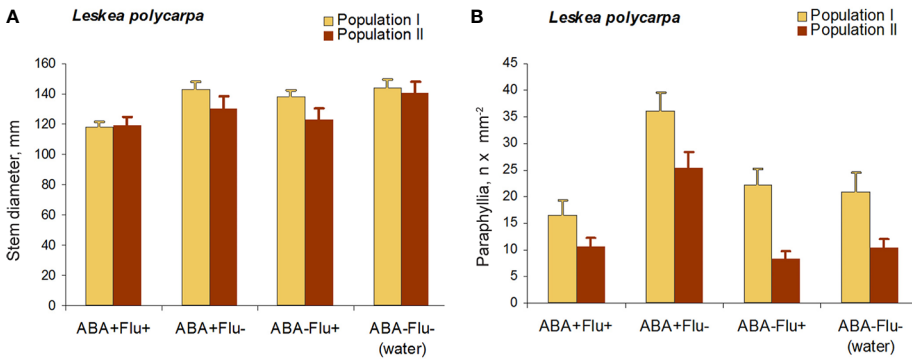


FIGURE 7 | Fluridone inhibition of the ABA effect on the growth of paraphyllia in *Leskea polycarpa*. Experiments were conducted with plants from two populations, the first with somewhat larger plants and more numerous paraphyllia. **(A)** stem width, changing slightly after growth with ABA and indistinctly narrowed in the series with fluridone; **(B)** Number of paraphyllia per square µm, showing their considerable increase after ABA treatment, which is, however, inhibited by fluridone. See also **Supplementary Materials 4**.

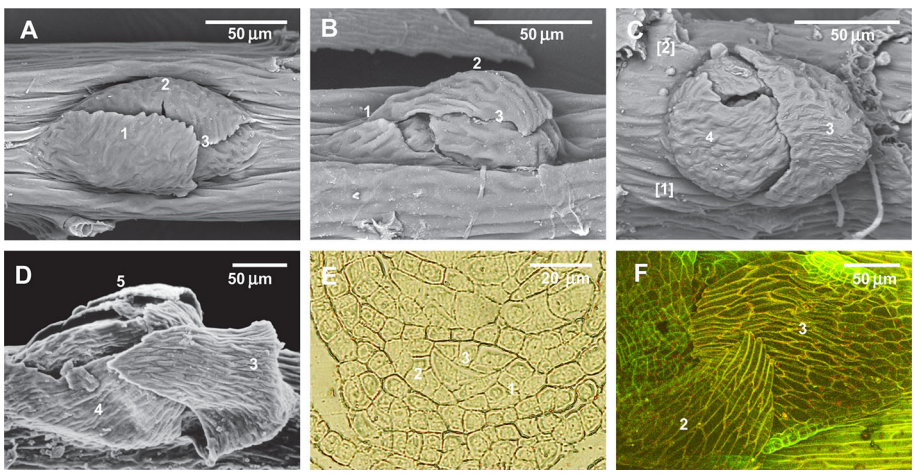


FIGURE 8 | Illustration of the reduction of the first proximal branch leaves in Lembophyllaceae s.l. (incl. Orthostichellaceae), the common absence of the first and second proximal branch leaves in Brachytheciaceae and their occasional re-appearance in response to ABA treatment, and the indifference to ABA of paraphyllia of Climaciaceae and Hylocomiaceae. **(A)** *Orthostichella hexasticha*, with a normally developed first proximal branch leaf; **(B)** *Pilotrichella cuspidata* Broth. with strongly reduced first proximal branch leaf; **(C)** *Camptochaete angustata* (Mitt.) Reichardt with totally reduced first and second proximal branch leaves; **(D)** *Brachythecium glareosum* (Bruch ex Spruce) Schimp., dormant bud, typical for the Brachytheciaceae family, with the outermost leaf in the "twelve o'clock position" due to complete reduction of the first and second proximal branch leaves; **(E)** *Brachythecium rivulare* Schimp., transverse section of branch primordium, showing the very early stage, confirming that the first merophytes are normally differentiated, but later do not form leaves; **(F)** *Brachythecium rutabulum* from cultivation with ABA, with developed second (or first)? proximal branch leaf in the "four o'clock position". **(A–D)**: SEM; **(F)**: LSCM.

branch leaf arrangement, so that the outermost leaf is always pointed downwards (**Figure 8D**). Spirina and Ignatov (2005) showed, however, that at the early developmental stage all three leaves exist (**Figure 8E**, cf. **Figure 2A**); therefore the outermost existing leaf (**Figure 8D**) is morphologically the third in sequence. Our attempt to reinstate the reduced leaves in *Brachythecium rutabulum* achieved a certain success. We found that five out of 50 studied primordia developed the second proximal branch leaf (**Figure 8F**) in the ABA-0.5 series, and one poorly developed second branch leaf was also noticed in the control series with distilled water. Statistics are difficult to apply in this case, but hundreds of observations on this character in studies for worldwide taxonomic treatments (Ignatov and Huttunen, 2002) and on specimens of Brachytheciaceae for

biodiversity studies never revealed any cases where the outermost leaf in any genus of Brachytheciaceae was seen in a lateral position. They were always found to be stable in retaining the "*Brachythecium*-type of branch primordia" as defined by Ignatov (1999); Ignatov and Huttunen (2002), and Ignatov and Hedenäs (2007).

DISCUSSION

The observations presented here demonstrate that there are at least two different patterns in the morphogenetic regulation of moss paraphyllia. The examples of *Leskea, Cratoneuron*, and *Leptodon* illustrate a pattern centered on the branch primordium, characterized by phyllotaxis in the paraphyllia arrangement and their concentration along the route passed by the bud in its morphogenetic history.

The homology of the *Leskea*-type of paraphyllia as a principally foliose structure probably has to be assumed as a compromise. On the one hand, there is an obvious regulation causing phyllotaxis even around a hardly discernible branch apical cell (**Figures 4A–D**); on the other hand, the paraphyllia appear irregularly arranged in the parts of internodes distant from the branch primordium. Such flexibility in paraphyllia development makes *Leskea* potentially a model object for studies of morphogenetic regulation in mosses. The release of some "normally undeveloped" structures from the "reduced state", like the second proximal branch leaf in *Brachythecium* (**Figure 8F**), is similar to the response of *Leskea* to ABA. Therefore this could be additional evidence for the homology of *Leskea*-type paraphyllia with leaves. Thus, the general definition of paraphyllia as structures not related to branch primordia is no longer valid. We suggest naming this pattern as "paraphyllia of the *Leskea*-type", which is characterized by phyllotaxis and a higher concentration of paraphyllia in the "green" and "yellow" zones (**Figure 3B**).

Paraphyllia of another type, the *Climacium*-type, are arranged mostly in longitudinal rows that cover the stem evenly and are not concentrated around branch primordia. Proximal branch leaves in these species are contrastingly different from paraphyllia (**Figure 1**). In some genera, e.g. *Lescuraea*, paraphyllia may be moderately dense to very sparse, but even in the latter case their even distribution along the stem is apparent and the "red zone" (**Figure 3B**) is never devoid of paraphyllia. Norris and Ignatov (2000) suggested that paraphyllia of the *Climacium*-type may be compared with micronemata ("microrhizoids") in Mniaceae; the differences of micronemata from macronemata ("normal rhizoids") in this family were described by Koponen (1968). The development of paraphyllia of the *Climacium*-type and micronemata is initiated from longitudinally oriented structures of modified and somewhat inflated epidermal cells: ridges in *Climacium* and *Pleuroziopsis* (**Figure 5**), and micronemata initials in Mniaceae. Micronemata are evenly distributed on stem surfaces; at least, there is no definite zone on the stem where they are fewer, in contrast to paraphyllia of the *Leskea*-type. However, there are important distinctions between micronemata and paraphyllia of the *Climacium*-type: micronemata lack chloroplasts and look like rhizoids, and their cell walls are oblique, while in paraphyllia of the *Climacium*-type most cell walls are transverse; micronemata originate closer to the center of micronemata initials, whereas paraphyllia in *Climacium* and *Hylocomium* are located at the cell end. Thus it is difficult to confirm their homology, if understood traditionally, while admitting that each structure may be homologous to just one organ. Alternatively, we may explain the structure of paraphyllia by the interference of different regulatory networks.

Kofuji and Hasebe (2014) found that the sporophyte generation in mosses has only one type of stem cell, whereas in gametophytes there are seven types: apical cells of chlorenchyma, caulonema, gametophore, leaf, rhizoid, antheridium, and archegonium. It seems that their number is even more than seven, if we consider the still little-known axillary hairs and paraphyses. As initials most of these organs originate from the stem surface, and it is likely that their regulatory systems interfere, thus the structures of "hybrid" identity may be developed, and *Climacium*-type of paraphyllia may be an example of this.

The present study demonstrates the possibility of using pleurocarpous mosses as a promising model object for studies that cannot be conducted with acrocarpous mosses, including *Physcomitrella*, as well as *Syntrichia* and *Ceratodon*. The acrocarpous mosses lack protective structures on the stem around branch initials that are easily observable. Conversely, the species of the order Hypnales, to which most pleurocarps belong, have diverse stem structures that are suitable for microscopic studies, including numerical counting and observation of their spatial distribution. There are still numerous disadvantages for their use as model objects, such as the lack of a reference genome, their representation by only wild types, and their rather slow growth. However, the study of the transition to dormancy and back might be an important issue on which mosses of the order Hypnales might shed light. Their success in modern biota, where they constitute no less than 40% of the diversity of all moss species, exceeding that of any other moss order, probably depends on their capability to postpone the development of young branches. Most Hypnales retain branch primordia in the state of "ready" or "steady", and immediately start growth after wetting, which is especially significant for epiphytic mosses, of which pleurocarps constitute the vast majority.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

General plan (MI), plant fixation and embedding (US), anatomical study (US), microscopy (US, TV, MI), ABA experiments (TV), manuscript preparation (MI, US, TV).

FUNDING

Russian Foundation for Basic Research 19-04-00976 and MBG Institutional research project 18-118021490111-5.

ACKNOWLEDGMENTS

We thank Tom Blockeel (Sheffield, U.K.) for improving the English of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00858/full#supplementary-material

REFERENCES

Akiyama, H., and Nishimura, N. (1993). Further studies of branch buds in mosses; "Pseudoparaphyllia". *"Scaly leaves". J. Plant Res.* 106, 101–108. doi: 10.1007/BF02344412

Akiyama, H. (1990a). A morphological study of branch development in mosses with special reference to pseudoparaphyllia. *Bot. Mag. Tokyo* 103, 269–282. doi: 10.1007/BF02488639

Akiyama, H. (1990b). Morphology and taxonomic significance of dormant branch primordia, dormant buds, and vegetative reproductive organs in the suborders Leucodontineae and Neckerineae (Musci, Isobryales). *Bryologist* 93, 395–408. doi: 10.2307/3243603

Alazem, M., and Lin, N.-S. (2017). Antiviral roles of abscisic acid in plants. *Front. Plant Sci.* 8, 1760. doi: 10.3389/fpls.2017.01760

Alazem, M., He, M.-H., Moffett, P., and Lin, N.-S. (2017). Abscisic acid induces resistance against Bamboo mosaic virus through Argonaute 2 and 3. *Plant Physiol.* 174, 339–355. doi: 10.1104/pp.16.00015

Berthier, J. (1971). Recherches sur la structure et le développement de l'apex du gamétophyte feuillé des mousses. *Rev. Bryol. Lichénol.* 38, 421–551.

Bonnot, E.-J. (1967). Sur la valeur et signification des paraphylles chez les Bryales. *Mém. Soc Bot. Fr.* 114, 236–248. doi: 10.1080/00378941.1967.10838522

Bruch, P., Shimper, W. P., and Gümbel, W. (1836–1855). *Bryologia Europaea seu Generum Muscorum Europaeorum Monographice Illustrata* Vol. 6 volumes (Stuttgart: Schweizerbart).

Cho, S. H., Schwartzenberg, K., and Quatrano, R. (2009). The role of abscisic acid in stress tolerance. *Ann. Plant Rev.* 36, 282–297. doi: 10.1002/9781444316070.ch11

Christianson, M. L. (2000). ABA prevents the second cytokinin-mediated event during the induction of shoot buds in the moss Funaria hygrometrica. *Am. J. Bot.* 87, 1540–1545. doi: 10.2307/2656880

Decker, E. L., Frank, W., Sarnighausen, E., and Reski, R. (2006). Moss systems biology en route: phytohormones in *Physcomitrella patens* development. *Plant Biol.* 8, 397–406. doi: 10.1055/s-2006-923952

Goode, J. A., Stead, A. D., and Duckett, J. G. (1993). Redifferentiation of moss protonemata: an experimental and immunofluorescence study of brood cell formation. *Can. J. Bot.* 71, 1510–1519. doi: 10.1139/b93-183

Hammer, O., Harper, D. A. T., and Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontol Electronica* 4 (1), 1–9.

Hedenäs, L. (2012). Morphological and anatomical features associated with epiphytism among the pleurocarpous mosses – one basis for further research on adaptations and their evolution. *J. Bryol.* 34, 79–100. doi: 10.1179/1743282011Y.0000000049

Hedwig, J. (1801). *Species muscorum frondosorum, descriptae et tabulis aeneis LXXVII coloratis illustratae*. Lipsiae (Leipzig): sumtu J. A. Barthii.

Huttunen, S., Ignatov, M. S., Müller, K., and Quandt, D. (2004). "Phylogeny and evolution of epiphytism in the three moss families Meteoriaceae, Brachytheciaceae, and Lembophyllaceae," in *Molecular systematics of bryophytes*. Eds. B. Goffinet, V. C. Hollowell and R. E. Magill (St. Louis: Missouri Botanical Garden Press), 328–355.

Huttunen, S., Ignatov, M. S., Olsson, S., Buchbender, V., Enroth, J., Hedenäs, L., et al. (2012). Phylogeny-based comparative methods question the adaptive nature of sporophytic specializations in mosses. *PloS One* 7 (10), e48268. doi: 10.1371/journal.pone.0048268

Ignatov, M. S., and Hedenäs, L. (2007). "Homologies of stem structures in pleurocarpous mosses, especially of pseudoparaphyllia and similar organs," in *Pleurocarpous mosses: systematic and evolution*. Eds. A. E. Newton and R. Tangery (Boca Raton, FL: CRC Press), 269–286.

Ignatov, M. S., and Huttunen, S. (2002). Brachytheciaceae (Bryophyta) – a family of sibling genera. *Arctoa* 11, 245–296. doi: 10.15298/arctoa.11.20

Ignatov, M. S., and Spirina, U. N. (2012). Morphogenesis of proximal branch leaves in mosses. *Russ. J. Dev. Biol.* 43, 148–156. doi: 10.1134/S1062360412030034

Ignatov, M. S. (1999). Bryophyte flora of the Huon Peninsula, Papua New Guinea. LXIII. On the pseudoparaphyllia in Brachytheciaceae and Meteoriaceae (Musci). *Acta Bot. Fenn.* 165, 73–83.

Ireland, R. R. (1971). Moss pseudoparaphyllia. *Bryologist* 74, 312–330. doi: 10.2307/3241639

Kitagawa, M., Tomoi, T., Fukushima, T., Sakata, Y., Sato, M., Toyooka, K., et al. (2019). Abscisic Acid Acts as a Regulator of Molecular Trafficking through Plasmodesmata in the Moss Physcomitrella patens. *Plant Cell Physiol.* 60, 738–751. doi: 10.1093/pcp/pcy249

Klíčová, Š., Šebánek, J., Hudeová, M., Vítková, H., and Vlašínová, H. (2002). The effect of fluridone and flurochloridone on the incidence of albinism in pea (Pisum sativum) and on the abscission of leaves of privet (Ligustrum vulgare). *Rostlinná výroba* 48, 255–260. doi: 10.17221/4237-PSE

Kofuji, R., and Hasebe, M. (2014). Eight types of stem cells in the life cycle of the moss Physcomitrella patens. *Curr. Opin. Plant Biol.* 17, 13–21. doi: 10.1016/j.pbi.2013.10.007

Koponen, T. (1968). Generic revision of Mniaceae Mitt. (Bryophyta). *Ann. Bot. Fenn.* 5 (2), 117–151.

Müller, F., and Applebyki, A. P. (2011). "Weed Control, 2. Individual Herbicides," in *Ullmann's Encyclopedia of Industrial Chemistry*. Wiley-VCH Verlag GmbH & Co. KGaA. doi: 10.1002/14356007.o28_o01

Nee, G., Xiang, Y., and Soppe, W. J. (2016). The release of dormancy, a wake-up call for seeds to germinate. *Curr. Opin. Plant Biol.* 35, 8–14. doi: 10.1016/j.pbi.2016.09.002

Norris, D. H., and Ignatov, M. S. (2000). Observations on stem surface anatomy in Climacium and Pleuroziopsis (Climaciaceae: Musci). *Arctoa* 9, 151–154. doi: 10.15298/arctoa.09.18

Osakabe, Y., Osakabe, K., Shinozaki, K., and Tran, L.-S. (2014). Response of plants to water stress. *Front. Plant Sci.* 5, 86. doi: 10.3389/fpls.2014.00086

Popova, L. P. (1995). Effect of fluridone on plant development and stress–induced ABA accumulation in Vicia faba L. plants. *Bulg. J. Plant Physiol.* 21, 42–50.

Shi, Y., Ding, Y., and Yang, S. (2015). Cold signal transduction and its interplay with phytohormones during cold acclimation. *Plant Cell Physiol.* 56, 7–15. doi: 10.1093/pcp/pcu115

Shu, K., Liu, X. D., Xie, Q., and He, Z. H. (2016). Two Faces of One Seed: Hormonal Regulation of Dormancy and Germination. *Mol. Plant* 9, 34–45. doi: 10.1016/j.molp.2015.08.010

Shu, K., Qi, Y., Chen, F., Meng, Y., Luo, X., Shuai, H., et al. (2017). Salt Stress Represses Soybean Seed Germination by Negatively Regulating GA Biosynthesis While Positively Mediating ABA Biosynthesis. *Front. Plant Sci.* 8, 1372. 10. doi: 10.3389/fpls.2017.01372

Shuai, H., Meng, Y., Luo, X., Chen, F., Zhou, W., Dai, Y., et al. (2017). Exogenous auxin represses soybean seed germination through decreasing the gibberellin/abscisic acid (GA/ABA) ratio. *Sci. Rep.* 7, 12620. doi: 10.1038/s41598-017-13093-w

Spirina, U. N., and Ignatov, M. S. (2005). A comparison of early stages of branch development in Brachythecium and Calliergon (Bryophyta). *Arctoa* 14, 211–222. doi: 10.15298/arctoa.14.16

Spirina, U. N., and Ignatov, M. S. (2008). Branch development and pseudoparaphyllia of *Hypnum cupressiforme* (Hypnales, Musci). *Arctoa* 17, 139–160. doi: 10.15298/arctoa.17.11

Spirina, U. N., and Ignatov, M. S. (2015). Bilobed leaves in mosses? Structure and adaptive significance of proximal branch leaves in Lembophyllaceae. *Arctoa* 24, 124–140. doi: 10.15298/arctoa.24.13

Stetsenko, L. A., Vedenicheva, N. P., Likhnevsky, R. V., and Kuznetsova, V. V. (2015). Influence of abscisic acid and fluridone on the content of phytohormones and polyamines and the level of oxidative stress in plants of *Mesembryanthemum crystallinum* L. under salinity. *Biol. Bull. Russ. Acad. Sci.* 42, 98. doi: 10.1134/S1062359015020107

Tintelnot, S. (2006). *Der einfluss von Abscisinsäure auf die pflanzliche zellwand: untersuchung extrazellulärer proteine beim laubmoos Physcomitrella patens* (Freiburg: University of Freiburg).

Tylewicz, S., Petterle, A., Marttila, S., Miskolczi, P., Azeez, A., Singh, R., et al. (2018). Photoperiodic control of seasonal growth is mediated by ABA acting on cell-cell communication. *Science* 360, 212–215. doi: 10.1126/science.aan8576

Velini, E. D., Trindade, M. L. B., Barberis, L. R. M., and Duke, S. O. (2010). Growth regulation and other secondary effects of herbicides. *Weed Sci.* 58, 351–354. doi: 10.1614/WS-D-09-00028.1

Warnstorf, C. (1904–1906). *Kryptogamenflora der Mark Brandenburg und angrenzender Gebiete* (Laubmoose, Zweiter Band. Leipzig: Verlag von Gebrüder Borntraeger).

Yoshida, T., Mogami, J., and Yamaguchi-Shinozaki, K. (2014). ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Curr. Opin. Plant Biol.* 21, 133–139. doi: 10.1016/j.pbi.2014.07.009

Zhang, D.-P. (Ed.) (2014). *Abscisic acid: metabolism, transport and signaling* (Dordrecht: Springer). doi: 10.1007/978-94-017-9424-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 June 2020
doi: 10.3389/fpls.2020.00765

Different Predictors Shape the Diversity Patterns of Epiphytic and Non-epiphytic Liverworts in Montane Forests of Uganda

Karola Maul[1], Yu-Mei Wei[2], Martin Nebel[1], Federico Luebert[1,3], Boon-Chuan Ho[4], Dietmar Quandt[1] and Michael Kessler[5]*

[1] Nees Institute for Biodiversity of Plants, University of Bonn, Bonn, Germany, [2] Guangxi Key Laboratory of Plant Conservation and Restoration Ecology in Karst Terrain, Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences, Guilin, China, [3] Departamento de Silvicultura y Conservatión de la Naturaleza, Universidad de Chile, Santiago, Chile, [4] Singapore Botanic Gardens, National Parks Board, Singapore, Singapore, [5] Institute of Systematic and Evolutionary Botany, Faculty of Mathematics and Natural Sciences, University of Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Bernard Goffinet, University of Connecticut, United States

Reviewed by:
Petr Koutecký, University of South Bohemia, Czechia
Michael A. Sundue, University of Vermont, United States
Claudine Ah-Peng, Université de la Réunion, France

***Correspondence:**
Karola Maul
Karola.maul@uni-bonn.de

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *21 January 2020*
Accepted: *14 May 2020*
Published: *24 June 2020*

Citation:
Maul K, Wei Y-M, Nebel M, Luebert F, Ho B-C, Quandt D and Kessler M (2020) Different Predictors Shape the Diversity Patterns of Epiphytic and Non-epiphytic Liverworts in Montane Forests of Uganda. Front. Plant Sci. 11:765. doi: 10.3389/fpls.2020.00765

We studied the influence of regional and local variables on the liverwort diversity within natural forest vegetation of Uganda to contribute to our understanding of the mechanisms and processes determining species richness. To this end, we compared the species richness distribution patterns of epiphytic and non-epiphytic liverworts (Marchantiophytina) in 24 plots in the forests of four Ugandan national parks. We recorded a total of 119 species and subspecies from 18 families, including 16 new species records for the country. We used generalized linear models (GLMs) and the relative variable importance of regional and local climatic and environmental variables to assess their respective impact on the species diversity. We found that the richness patterns of total and epiphytic richness were largely driven by regional climatic factors related to temperature and water-availability. In contrast, species diversity of non-epiphytic and rare species was additionally strongly determined by local-scale microhabitat factors such as height of forest canopy and slope inclination, reflecting the availability of suitable microhabitats. We conclude that macroclimatic variables perform well in predicting epiphytic liverwort richness, whereas the adequate prediction of non-epiphytic richness requires site-specific variables. Also, we propose that richness of epiphytic liverworts will be impacted more directly by climate change than richness of non-epiphytic and rare species.

Keywords: epiphytic liverworts, non-epiphytic liverworts, species richness distribution, tropical montane forest, climatic predictors, Uganda

INTRODUCTION

Biodiversity is unevenly distributed in space and time (Hawkins, 2001) and understanding the causes and drivers of its geographical distribution remains an essential goal in ecology and biogeography. Research has identified many different ecological and evolutionary drivers of biodiversity patterns, whose influence varies depending on the taxonomic group under

consideration, the geographical region, and the spatial and temporal scales (e.g., Körner, 2000; McCain, 2005; McCain and Grytnes, 2010). Among these factors, climatic conditions have often been found to be closely correlated to contemporary species richness patterns (Araujo and Rahbek, 2006; Kessler et al., 2011). For instance, along elevational gradients, the balance between high temperatures and subsequent drought stress in the lowlands and low temperatures at high elevations may lead to optimal conditions and high diversity of many plant groups at mid-elevations, including ferns (Kluge and Kessler, 2011; Kluge et al., 2017) and bryophytes (Song et al., 2015). However, diversity can also be influenced by large-scale geographical factors such as land surface area (Karger et al., 2017) and geometric constraints in spatially restricted regions (Colwell et al., 2016). Besides these factors acting at broad spatial scales, diversity patterns are also influenced by localized factors, such as the presence of a specialized habitat required by a certain group of organisms. Although the importance of both regional- and local-scale factors is generally acknowledged (Crawley and Harral, 2001; Chalcraft, 2013), most studies identifying correlates of biodiversity patterns only focus on one of these scales, due to the availability of data at a given scale within a specific study. There is thus a need to combine factors affecting biodiversity at different spatial scales to better understand environmental drivers of biodiversity and to predict responses of biodiversity to changes in these drivers, e.g., as a result of climate change.

Bryophytes are an ecologically and evolutionarily distinct plant group including liverworts, mosses and hornworts, the almost globally distributed earliest extant land plant lineages. They are characterized by a poikilohydric lifestyle with a dominant haploid (gametophytic) and a short-lived diploid (sporophytic) generation. Because of their weak ability to actively regulate their water balance (Leon-Vargas et al., 2006; Proctor et al., 2007) bryophytes are poorly represented in arid regions, but in contrast are well adapted to low temperatures, so that they even occur in Antarctica and in high mountains above the treeline (Longton, 1982; Seppelt and Green, 1998; Bruun et al., 2006). It may thus not be so surprising that a global latitudinal gradient in species richness of mosses has not been supported (Shaw et al., 2005; Geffert et al., 2013). Liverworts in turn seem to be less frost tolerant than mosses (Glime, 2017), and a general latitudinal gradient in richness distribution has recently been demonstrated (Wang et al., 2017). Macroclimatic conditions including precipitation, temperature, and radiation have been found to be important drivers of these richness patterns (Sun et al., 2013; Zellweger et al., 2015; Spitale, 2016). Liverworts species richness, in particular, seems to be more closely correlated to macroclimate than moss richness (Aranda et al., 2014; Chen et al., 2015; Henriques et al., 2016).

On the other hand, bryophytes are small plants that are able to inhabit localized habitats like bare rocks and even leaves of vascular plants, when coupled with various strategies as well as physiological and structural adaptations (Kürschner, 2004; Kraichak, 2012). Hence, there are various studies showing a strong influence of small-scale biotic and abiotic habitat properties on different bryophyte species or growth forms, for instance relative air humidity and its daily fluctuations on the distribution of epiphyllous bryophytes in Costa Rica (Sonnleitner et al., 2009), vapor pressure deficit and soil moisture on growth of two moss species in Canada (Stewart and Mallik, 2006), and variations in canopy cover on species richness and species composition of terrestrial bryophytes in Ecuador (Mandl et al., 2009). Furthermore, it appears that the richness of epiphytic and non-epiphytic bryophytes may be influenced by a different set of factors, with the epiphytes being more closely linked to general climatic conditions (Gradstein, 2008; Zotz and Bader, 2009), and non-epiphytic species more closely to soil conditions or dead wood availability (Raabe et al., 2010). However, comparative studies researching the drivers of richness distribution patterns of epiphytic versus non-epiphytic liverworts in tropical forests are scarce. Several studies do not consider the ecological differences between mosses and liverworts (e.g., Sun et al., 2013), or do not distinguish between the sampled microhabitats (e.g., Grau et al., 2007; Tusiime et al., 2007). Other studies included only epiphytes (e.g., Wolf, 1993; Song et al., 2015) or terrestrials (e.g., Mandl et al., 2009), or focused on higher latitudes (e.g., Spitale, 2016).

In the present study, we compared the influence of regional-scale macroclimatic variables and local-scale factors such as relative air humidity, inclination, canopy closure, and canopy height on the species richness of epiphytic and non-epiphytic liverworts in Uganda. We addressed the following research questions:

(1) What is the relative importance of regional climatic factors and local site-specific factors for the prediction of species diversity patterns of Ugandan liverworts?
(2) Are there different factors determining the richness patterns of epiphytic and non-epiphytic liverworts?

MATERIALS AND METHODS

Study Sites and Sampling Design

Because of the fragmented nature of forests in Uganda, we conducted our survey in four separate protected areas in western Uganda (**Figure 1**). Our first study site was the northern and middle parts of Kibale National Park (KNP), where we sampled five plots at 1270–1500 m. KNP is located on a plateau bordering the western Great Rift Valley. The low to high grade metamorphic bedrock belongs to the Paleoproterozoic Buganda-Toro System (Schlüter, 2006). The soil is either mostly low to moderately fertile consisting of red sandy loam, or fertile when overlying volcanic tuff (Howard, 1991). The vegetation in northern KNP consists of tropical moist evergreen forest up to 50 m tall with *Parinari excelsa* Sabine (Chrysobalanaceae) as one of the dominant tree species (Kasenene, 2001). *Olea welwitschii* Gilg & G.Schellenb. (Oleaceae), *Chrysophyllum* spp., *Aningeria altissima* (A.Chev.) Aubrév. & Pellegr. (Sapotaceae), *Strombosia scheffleri* Engl. (Olacaceae), *Newtonia buchananii* (Baker) G.C.C.Gilbert & Boutique (Fabaceae), *Celtis* spp. (Ulmaceae), *Diospyros abyssinica* (Hiern) F.White (Ebenaceae), and *Markhamia platycalyx* Sprague (Bignoniaceae) are characteristic for the mixed forest communities in the center of KNP.

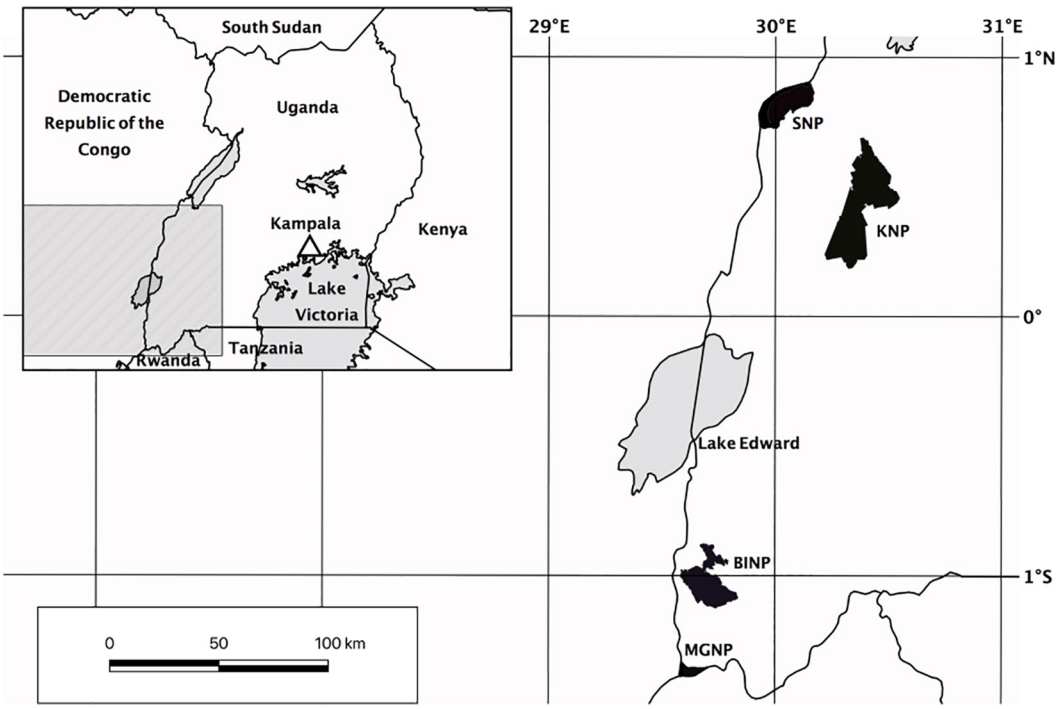


FIGURE 1 | Map of study sites in Uganda. KNP, Kibale National Park; SNP, Semliki National Park; BINP, Bwindi Impenetrable National Park; MGNP, Mgahinga Gorilla National Park.

We established three study plots at Semliki National Park (SNP), which is situated on the floor of the rift valley along Uganda's border to the Democratic Republic of the Congo. The area is flat to hilly at 670–760 m, and largely covered with moist semi-deciduous forest (Howard, 1991). Although Uganda Ironwood (*Cynometra alexandri* C.H. Wright, Fabaceae) of up to 30 m tall is the dominating tree species, which partly forms almost pure stands (Howard, 1991), more than 300 tree species have been recorded within SNP in total (Howard et al., 2000). The larger part of the ground is poorly drained low fertility gray clay alluvials and rift sediments of recent to Neogene origin (Howard, 1991; Schlüter, 2006; Ring, 2008).

Our third study site was located in Bwindi Impenetrable National Park (BINP) in the Kigezi Highlands. This area has a rough terrain with steeply sloping mountains and narrow gorges that was shaped by the upwarping of the western rift valley. The groundrock is part of the Mesoproterozoic Karagwe-Ankolean System, which is low grade to unmetamorphosed (Schlüter, 2006). The soil is largely characterized by ferralitic humic loam, of low to high acidity and extremely poor in bases (Howard, 1991). BINP has an extremely rich flora and fauna, with many regionally endemic species (Butynski and Kalina, 1993). Large parts of BINP are occupied by mixed forest with *Prunus africana* (Hook.f.) Kalkman (Rosaceae), *Newtonia buchananii* (Baker) G.C.C.Gilbert & Boutique (Fabaceae), *Symphonia globulifera* L.f. (Clusiaceae), *Chrysophyllum* spp. (Sapotaceae), *Podocarpus* spp. (Podocarpaceae), and *Strombosia scheffleri* Engl. (Olacaceae) (Howard, 1991). We established 11 plots at 1515–2450 m in medium altitude moist evergreen forest and high elevation forest including two plots within the montane bamboo zone dominated by *Arundinaria alpina* K. Schum. (Poaceae) (Howard, 1991).

Finally, Mgahinga Gorilla National Park (MGNP) is the smallest part of the Virunga Conservation Area, which is situated on the border triangle between Uganda, Rwanda and the Democratic Republic of the Congo. MGNP covers the slopes of three volcanoes that are part of the Mesoproterozoic Karagwe-Ankolean System. The soils in the Virunga Massif are of volcanic origin and generally fertile (Hitimana et al., 2006; Akayezu et al., 2019). We established two plots at high altitude forest (2500–2800 m), including one within the broad zone of *Arundinaria* montane bamboo forest at the base of Mount Sabinyo (Langdale-Brown et al., 1964). Two plots were established in the following transition zone to the ericaceous forest at 2940 m and 3050 m, characterized by *Hypericum revolutum* Vahl (Hypericaceae) and *Hagenia abyssinica* J.F. Gmel. (Rosaceae). One plot was established in the ericaceous forest at 3200 m, which is characterized by *Philippia johnstonii* Schweinf. ex Engl. and *Erica arborea* L. (Ericaceae) (Owiunji et al., 2005).

In total, we established 24 plots of 20 × 20 m^2 between 680 and 3200 m within mature natural forest, avoiding areas with vegetation disturbed by crossing trails and stream canyons. In each plot, we sampled eight subplots of 20 × 30 cm^2 each, two for each of four microhabitats (soil, rotten logs, tree trunks above 1 m, and tree branches from 2 to 4 m). The subplots were located where bryophyte abundance was high. To obtain a plot inventory as complete as possible, we also sampled outside the

subplots across the entire plot, including additional habitats (e.g., twigs, lianas, shrubs, small trees, tree bases etc.). We defined all specimens collected on tree trunks above 1 m, branches, twigs, lianas etc. as epiphytes. All specimens collected on the soil, and on rotten logs were assigned as non-epiphytes. Because tree trunk bases are a transition zone from non-epiphytic (soil, rotten logs, rocks) to epiphytic assemblages (Thomas et al., 2001; Holz et al., 2002), specimens occurring at the base of tree trunks (below 1 m), were treated as epiphytes when the majority of the other records of the species were epiphytic, or as non-epiphytic when the majority of other records was non-epiphytic. Species occurring in epiphytic and in non-epiphytic habitats were assigned to both habitats.

We used the checklist of Wigginton (2018) to assess the new records for Uganda, the literature cited therein was used as major reference to the species identification of the specimens (List S1). Classification and nomenclature followed Söderström et al. (2016). All vouchers are deposited in the herbarium of the Nees Institute for Biodiversity of Plants, University of Bonn (BONN), duplicates will be deposited in the herbarium of the Makerere University (MHU), Kampala, Uganda, after finishing the process of herbarium curation. Infraspecific taxa are treated as "species" for ease of data analyses and are simply regarded as such in the discussion.

Data Analyses

The proportion of unique and shared species among the four national parks was assessed using a Venn diagram (**Figure 2**).

We employed generalized linear models (GLMs) and two different parameter sets to assess the influence of climatic and environmental parameters on liverwort species richness. The regional dataset consisted of six climatic variables: annual mean temperature (Temp), temperature seasonality (TempS), annual precipitation (Prec), precipitation seasonality (PrecS), obtained from the CHELSA model (Karger et al., 2017), as well as solar radiation (Rad) obtained from Wilson and Gallant (2000) (**Supplementary Table S1**). Annual mean temperature was included as linear and quadratic term in the analyses ($Temp^2$), since temperature might have a unimodal influence on richness. The local dataset included six variables recorded in each individual plot during fieldwork by visual estimation by always the same observer: inclination (Inc), percentage of ground covered by plants (GCov), canopy height (HC), canopy cover (CCov), percentage of canopy branch surface covered by bryophytes (BCov) as a proxy for air humidity (Karger et al., 2012), and distance to open water, which might also influence air humidity (DW; **Supplementary Table S1**). All 12 variables were associated among each other with pairwise Pearson's correlation coefficients of $\leq$0.8 (**Supplementary Table S2**). We calculated GLMs with Poisson distribution and species richness as responding variables, and regressed them against all possible independent variable combinations from both parameter sets separately (regional: richness $\sim$ Temp + $Temp^2$ + TempS + Prec + PrecS + Rad, local: richness $\sim$ Inc + GCov + HC + CCov + BCov + DW) and for the combined dataset (regional + local: richness $\sim$ Temp + $Temp^2$ + TempS + Prec + PrecS + Rad + Inc + GCov + HC + CCov + BCov + DW). To avoid model overfitting, the possible combinations were restricted to a maximum of four variables per model in all the three approaches. The resulting models were ranked according to their respective second order Akaike information criterion (AICc) (Hurvich and Tsai, 1989). The goodness of fit of the best models (i.e., ΔAICc $\leq$ 2) was assessed using the Kullback–Leibler-divergence-based R^2 (Cameron and Windmeijer, 1997). To check the significance of the improvement of the fitted best model compared to simpler models, we reduced the independent variables stepwise according to the lowest *z*-value and compared the change in deviance running chi-square tests (**Table 1**). The relative variable importance (RVI) of each single variable was calculated by summing up the AICc weights of all the models in which the respective variable was included (**Figure 3** and **Supplementary Table S3**). To assess the effect of the additional non-standardized collections within the plots but outside subplots, we conducted

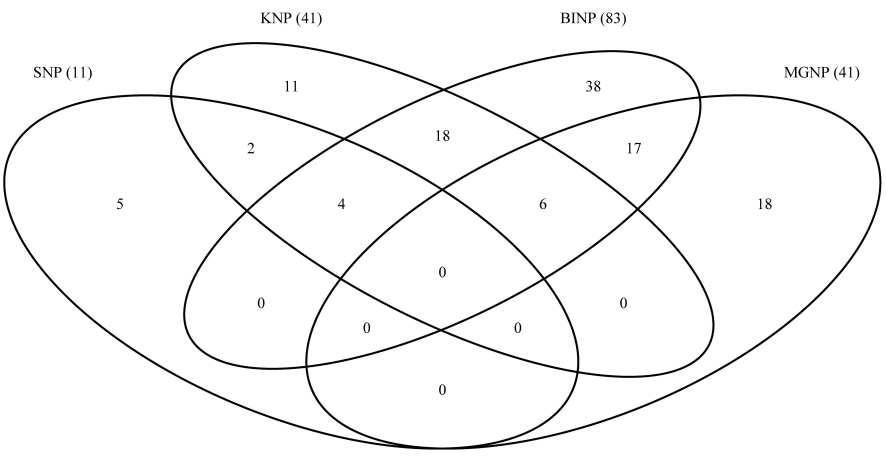


FIGURE 2 | Venn diagram showing numbers of species shared and unique to the four study sites. SNP, Semliki National Park; KNP, Kibale National Park; BINP, Bwindi Impenetrable National Park; MGNP, Mgahinga Gorilla National Park. Numbers in brackets indicate the total species numbers recorded at each site.

TABLE 1 | Summary of the best models resulting from the regional (a), local (b), and combined regional + local dataset (c) within ΔAICc = 2.

(a) Regional factors	(Intercept)	Temp	Temp^2	Prec	PrecS	TempS	Rad	df	logLik	AICc	Delta	Weight	Adj. R^2	Mean adj. R^2	Pr(>Chi)
All	3.13***	−0.46***^{4}	−0.42***^{1}		−0.62***^{2}	−0.64***^{3}		5	−71.09	155.51	0	0.999	0.50		*/./***
Epiphytes	2.87***	−0.36**^{4}	−0.46***^{1}		−0.59***^{2}	−0.61***^{3}		5	−62.99	139.31	0	0.915	0.54		./*/**
Non-epiphytes	1.86***	−0.98***^{1}		−0.76***^{2}		−0.31*^{3}		4	−68.22	146.54	0	0.168	0.25		***/*/–
(b) Local factors	**(Intercept)**	**CCov**	**HC**	**DW**	**GCov**	**Inc**	**BCov**	**df**	**logLik**	**AICc**	**Delta**	**Weight**	**Adj. R^2**		**Pr(>Chi)**
All	2.75***		0.25**^{2}	0.21**^{3}	−0.12*^{4}		0.33***^{1}	5	−80.41	174.14	0	0.251	0.23		*/*/*
	2.76***		0.29***^{2}	0.19*^{3}			0.30***^{1}	4	−82.51	175.13	0.99	0.153	0.21	0.22	*/*/–
Epiphytes	2.48***				−0.2**^{1}	−0.14.^{3}	0.29***^{2}	4	−74.60	159.31	0	0.128	0.15		**/./–
	2.48***				−0.19**^{1}		0.2**^{2}	3	−76.08	159.37	0.06	0.124	0.14		./–/–
	2.47***		0.23*^{3}	0.17.^{4}	−0.15*^{1}		0.26***^{2}	5	−73.31	159.94	0.64	0.093	0.15		**/ /.
	2.47***	−0.11^4			−0.25***^{1}	−0.16.^{3}	0.37***^{2}	5	−73.38	160.09	0.78	0.087	0.15		**/./
	2.48***		0.10^3		−0.14.^{1}		0.21**^{2}	4	−75.1	160.31	1.0	0.078	0.13		**/ /–
	2.48***	−0.1^3			−0.22**^{1}		0.25**^{2}	4	−75.17	160.44	1.13	0.073	0.13		**/ /–
	2.47***		0.10^4		−0.15*^{1}	−0.14.^{3}	0.31***^{2}	5	−73.68	160.68	1.34	0.064	0.14		**/./
	2.49***		0.17**^{2}				0.19**^{1}	3	−76.97	161.15	1.84	0.051	0.10		**/–/–
	2.48***		0.28**^{2}	0.15.^{3}			0.23**^{1}	4	−75.59	161.28	1.97	0.048	0.11	0.13	**/./–
Non-epiphytes	1.85***		0.13.^{2}				0.54***^{1}	3	−67.19	141.58	0	0.15	0.31		./–/–
	1.85***						0.49***^{1}	2	−68.63	141.83	0.26	0.13	0.31		–/–/–
	1.84***		0.27*^{2}	0.18^3			0.6***^{1}	4	−66.18	142.46	0.88	0.1	0.31		./ /–
	1.85***		−0.08^2				0.53***^{1}	3	−68.08	143.36	1.78	0.06	0.29		/–/–
	1.85***					−0.1^2	0.57***^{1}	3	−68.11	143.42	1.84	0.06	0.29	0.3	/–/–
(c) Regional + local factors	**(Intercept)**	**Temp**	**Temp^2**	**Prec**	**PrecS**	**TempS**	**Inc**	**BCov**	**df**	**logLik**	**AICc**	**Delta**	**Weight**	**Adj. R^2**	**Mean adj. R^2**	**Pr(>Chi)**
All	2.94***		−0.23***^{4}		−0.43***^{3}	−0.75***^{2}		0.61***^{1}	5	−69.01	151.36	0	0.53	0.56		*/***/***
Epiphytes	2.72***		−0.3***^{1}		−0.44***^{2}	−0.7***^{3}		0.46***^{4}	5	−62.25	137.83	0	0.41	0.57		. /*/***
	2.87***	−0.35**^{4}	−0.45***^{1}		−0.59***^{2}	−0.64***^{3}			5	−62.99	139.31	1.48	0.19	0.54	0.55	. /*/**
Non-epiphytes	1.71***			−0.43***^{3}		−1.09***^{2}	−0.19*^{4}	1.82***^{1}	5	−56.13	125.59	0	0.19	0.57		*/***/*
	1.74***			−0.41***^{3}		−0.98***^{2}		1.53***^{1}	4	−58.11	126.33	0.74	0.13	0.53	0.55	*/***/–

*Temp, annual mean temperature; Temp^2, (annual mean temperature)^2; Prec, annual precipitation; PrecS, precipitation seasonality; TempS, temperature seasonality; Rad, solar radiation; CCov, canopy cover; HC, height of canopy; DW, distance to open water; GCov, ground plant cover; Inc, inclination; BCov, bryophyte cover of branches; significance of Z-value of the respective variable is given after each coefficient estimate. df, degrees of freedom; delta, difference AICc to the best model; adj. R^2, adjusted R^2; mean adj. R^2, arithmetic mean of R^2 of the best models within delta AICc = 2, asterisks after coefficient estimates indicate significance level of variable in the respective model (Pr(>| z|). Superscripts depict the order of z-value magnitude, with ^{1} = highest absolute z-value. Pr(>Chi) = significance of improvement of best models compared to simpler models (obtained by successive removal of least significant variables based on z-values). (–) = indicate best model with less than four variables. Significance codes: ***0 < p ≤ 0.001; **0.001 < p ≤ 0.01; *0.01 < p ≤ 0.05; (.) 0.05 < p ≤ 0.1.*

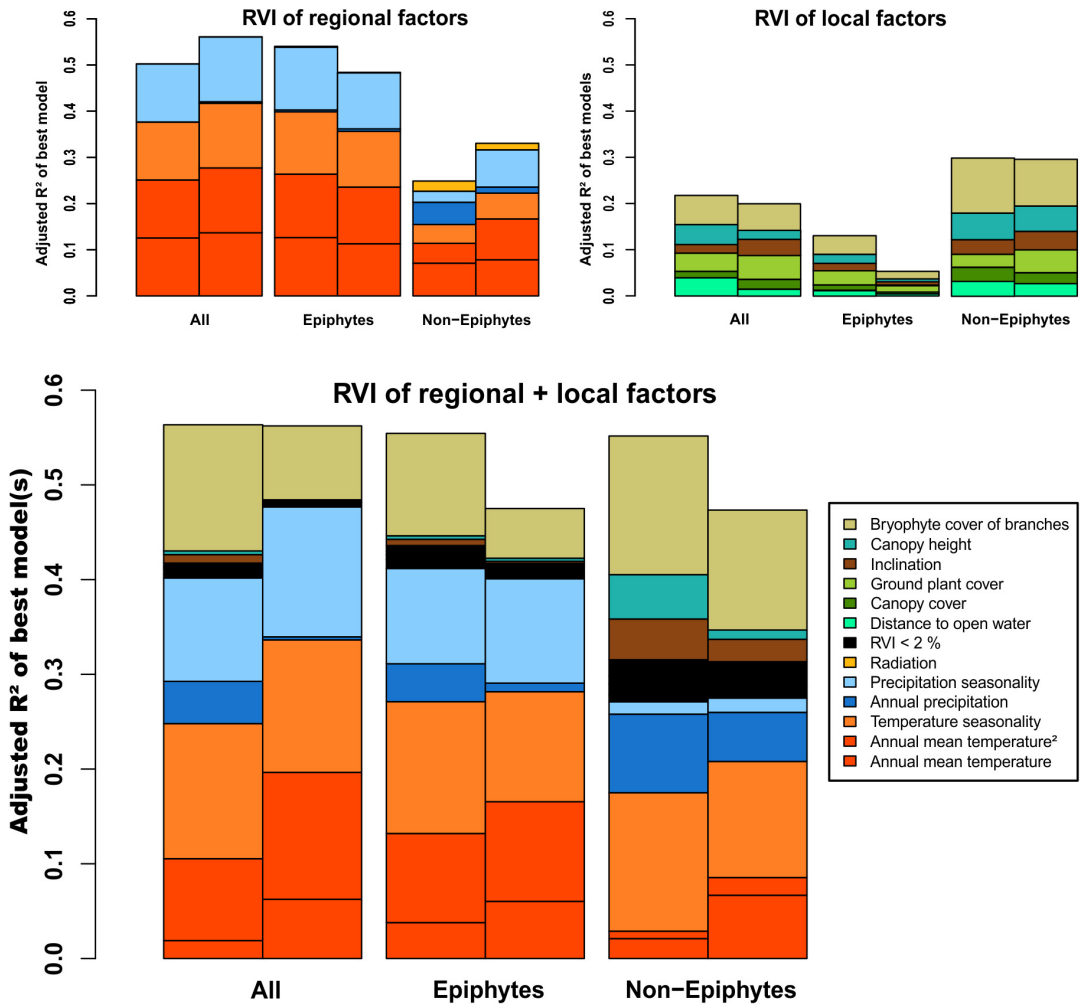


FIGURE 3 | Barplots illustrating relative variable importance (RVI) of regional, local, and regional + local factors, relative to the according goodness of fit criterion (adjusted R^2). Left bars show results from plot analyses, right bars from analyses restricted to subplot records.

the analyses including (a) all species records and (b) records from the subplots only (**Figure 3** and **Supplementary Tables S3**, **S4**). Finally, we included the four to five (non-epiphytic richness) variables with highest RVI in univariate or polynomial ($Temp^2$) GLMs to assess if the performance of the best models was a direct effect or part of a complex interplay between multiple variables (**Supplementary Figure S1**).

All computational analyses were conducted in R 3.6.0 (R Core Team, 2017) using the packages VennDiagram 1.6.20 (Chen, 2018), MuMIn 1.43.6 (Barton, 2019), and rsq 1.1 (Zhang, 2018).

RESULTS

We obtained a total of 940 species-plot records of which 653 (69.5%) were collected within the subplots. Our collection included a total of 119 species, subspecies and varieties from 19 families. No fewer than 16 species were new records for Uganda (**Table 2**). The most species rich family was Lejeuneaceae (51 species) and the most speciose genera were *Lejeunea* (22) and *Plagiochila* (18). Out of all collected species *Plagiochila kiaeri* Gottsche and *Metzgeria furcata* (L.) Corda were recorded in the majority of plots (16 and 14 plots, respectively, **Table 2**). The species occupying the widest elevational range was *Lejeunea conformis* Nees et Mont. (9 plots, 1275–3200 m). On average, we recorded 16.4 species per plot (range 3–28). A total of 48 species occurred exclusively in epiphytic and 30 species in non-epiphytic habitats; 41 species were recorded from both habitat categories. Ninety seven species and subspecies (81%) were collected from within the subplots only.

60.5% of all species recorded were found in only one national park, and no species was found in all four national parks (**Figure 2**). 8.4% of the species were shared between three national parks, and 31.1% were found in two national parks. KNP shared 28 (68%) of its recorded species with BINP, and BINP and MGNP shared 23 species, whereas SNP and MGNP did not share any species.

TABLE 2 | Liverwort inventory with collection sites and habitats.

Plot		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
GPS-coordinates:	**Latitude**	0.564722222	0.566083333	0.451680556	0.459275	0.642444444	0.829669444	0.833180556	0.837138889	−0.99268333	−0.99443611	−0.98994444	−0.98823333	−1.01033611	−1.00949722	−1.09643333	−1.08771389	−1.05808889	−1.07576667	−1.07199167	−1.38119722	−1.38446944	−1.38297222	−1.37603056	−1.37719444
	Longitude	30.3583278	30.3561389	30.3849361	30.3777111	30.3946667	30.0893778	30.0896806	30.0886083	29.614725	29.6144389	29.60675	29.6245861	29.7379222	29.7378611	29.8033	29.8080361	29.7954611	29.7641139	29.7677833	29.6026	29.5990806	29.6011222	29.6044972	29.6040833
Site		KNP	KNP	KNP	KNP	KNP	SNP	SNP	SNP	BINP	BINP	BINP	BINP	BINP	BINP	BINP	BINP	BINP	BINP	BINP	MGNP	MGNP	MGNP	MGNP	MGNP
Date of collection		15.Feb.2014	15.Feb.2014	16.Feb.2014	16.Feb.2014	17.Feb.2014	18.Feb.2014	18.Feb.2014	18.Feb.2014	23.Feb.2014	23.Feb.2014	23.Feb.2014	24.Feb.2014	25.Feb.2014	25.Feb.2014	26.Feb.2014	26.Feb.2014	27.Feb.2014	28.Feb.2014	28.Feb.2014	02.Mar.2014	02.Mar.2014	02.Mar.2014	03.Mar.2014	03.Mar.2014
Exposition		S	-	-	SW	E	-	-	-	E	E	W	NW	N	S	W	S	E	N	E	NE	N	N	N	E
Inclination [°]		5	0	0	5	8	0	0	0	5	10	15	25	45	40	15	25	20	20	15	15	45	35	5	20
Altitude [m.a.s.l.]		1500	1530	1270	1275	1450	700	680	690	1550	1514	1570	1700	1950	1980	2300	2450	2350	2100	2200	2940	3200	3050	2700	2750
Species	**Family**																								
Pseudomarsupidium decipiens (Hook.) Grolle	Adelanthaceae																					e, rl			
Plicanthus hirtellus (F.Weber) R.M.Schust.	Anastrophyllaceae													t											
Aneura pinguis (L.) Dumort.	Aneuraceae	rl									rl														
Riccardia sp. 1	Aneuraceae			rl*																					
Riccardia sp. 2	Aneuraceae									rl	t									t*					
Riccardia sp. 3	Aneuraceae							rl																	
Riccardia sp. 4	Aneuraceae									e*	rl*											rl*			rl*
Calypogeia fissa (L.) Raddi	Calypogeiaceae																								t
Cylindrocolea abyssinica (Gola) Váòa	Cephaloziellaceae																			rl					
Dumortiera hirsuta (Sw.) Nees	Dumortieraceae	t*				t*					t*														
Frullania caffraria Steph.	Frullaniaceae																		e*					e	
Frullania ericoides (Nees) Mont.	Frullaniaceae					e																			

(Continued)

TABLE 2 | Continued

Plot		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Frullania obscura (Sw.) Mont.	Frullaniaceae														e			e			e			e	
Frullania obscurifolia Mitt.	Frullaniaceae										e													e	
Frullania trinervis (Lehm.) Drège	Frullaniaceae	e	e	e	e	e	e	e		e			e												
Haplomitrium sp.	Haplomitriaceae																					t*			
Herbertus dicranus (Gottsche. Lindenb. et Nees) Trevis	Herbertaceae																					e, rl, t	e		
Acanthocoleus chrysophyllus (Lehm.) Kruijt	Lejeuneaceae	e	e																						
Caudalejeunea lehmanniana (Gottsche) A.Evans	Lejeuneaceae						e																		
Caudalejeunea lewallei Vanden Berghen	Lejeuneaceae									e	e	e	e		e										
Ceratolejeunea cornuta (Lindenb.) Steph.	Lejeuneaceae									e*		e													
Cheilolejeunea krakakammae (Lindenb.) R.M.Schust.	Lejeuneaceae													e*					e*						
Cheilolejeunea montagnei (Gottsche ex Mont.) R.M.Schust.	Lejeuneaceae													rl											
Cheilolejeunea roccatii (Gola) W.Ye. R.L.Zhu et Gradst.	Lejeuneaceae																					e, rl	e, rl		
Cheilolejeunea surrepens (Mitt.) E.W.Jones	Lejeuneaceae								e																
Cheilolejeunea xanthocarpa (Lehm. et Lindenb.) Malombe	Lejeuneaceae																			e					
Cololejeunea elegans Steph.	Lejeuneaceae										e*				e										
Cololejeunea harrisii Pócs	Lejeuneaceae										e*														
Cololejeunea lemuriana Tixier	Lejeuneaceae											e													
Diplasiolejeunea runssorensis Steph.	Lejeuneaceae														e, rl							e	e		
Drepanolejeunea physifolia (Gottsche) Pearson	Lejeuneaceae											e		e	e						e	e	e		
Lejeunea abyssinica (Gola) Cufod	Lejeuneaceae														e	e, rl									
Lejeunea alata Gottsche	Lejeuneaceae																							e	e*

(Continued)

TABLE 2 | Continued

Plot		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Lejeunea anisophylla Mont.	Lejeuneaceae	rl				e, rl	rl	e	e																
Lejeunea aphanes Spruce	Lejeuneaceae					e*				e, rl	e	e	e					e*							
Lejeunea brenanii E.W.Jones	Lejeuneaceae													e, b*											
Lejeunea cantabrigiensis E.W.Jones	Lejeuneaceae																		rl*						
Lejeunea capensis Gottsche	Lejeuneaceae																							e	
Lejeunea conformis Nees et Mont.	Lejeuneaceae	rl, t			e*					e*, rl, b*			e			rl			e, rl*, t*			rl*		e*	rl
Lejeunea eckloniana Lindenb.	Lejeuneaceae	e, rl		e		e		e				e													
Lejeunea flava (Sw.) Nees	Lejeuneaceae													e	e	e*		e, rl, b*		e, rl*				e	
Lejeunea flava ssp. *tabularis* (Spreng.) S.W.Arnell	Lejeuneaceae		e*		e					e*	e	e			e*										
Lejeunea flavovirens Ångstr.	Lejeuneaceae																e, rl*			e	e	e	e	e	
Lejeunea ibadana A.J.Harr. et E.W.Jones	Lejeuneaceae																								e
Lejeunea isophylla E.W.Jones	Lejeuneaceae																	e		rl	e, rl			e	e
Lejeunea lomana E.W.Jones	Lejeuneaceae		e							e						e	e, rl*	e, rl	e, rl		rl*			e	rl
Lejeunea papilionacea Prantl	Lejeuneaceae	e											e		e	e									
Lejeunea phyllobola Nees & Mont.	Lejeuneaceae	e	e																						
Lejeunea ramosissima Steph.	Lejeuneaceae									e*, rl*	e, rl	t		e, b*	e, rl										
Lejeunea setacea (Steph.) Steph.	Lejeuneaceae							e																	
Lejeunea sp. 1	Lejeuneaceae																			rl					
Lejeunea tuberculosa Steph.	Lejeuneaceae	e	e			e				e		e	e*		e	e			e						
Lejeunea villaumei (Steph.) Grolle	Lejeuneaceae												e*												
Lopholejeunea revoluta E.W.Jones	Lejeuneaceae	e*																							
Marchesinia excavata (Mitt.) Schiffn.	Lejeuneaceae									rl*															
Marchesinia nobilis (Gottsche) X.Q.Shi. R.L.Zhu et Gradst.	Lejeuneaceae									t															
Mastigolejeunea sp.	Lejeuneaceae		e																						
Microlejeunea africana Steph.	Lejeuneaceae		e									e		e				e						e	

(Continued)

TABLE 2 | Continued

Plot		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Microlejeunea sp. 1	Lejeuneaceae																			e					
Microlejeunea sp. 2	Lejeuneaceae										e														
Microlejeunea ulicina (Taylor) A.Evans	Lejeuneaceae																		e						
Prionolejeunea grata (Gottsche) Schiffn.	Lejeuneaceae										e	e													
Ptychanthus africanus Steph.	Lejeuneaceae	e*	e	e*	e	e, rl*				e, rl		e	e												
Schiffneriolejeunea altimontana Vanden Berghen	Lejeuneaceae													e	e					e, rl					
Schiffneriolejeunea pappeana (Nees) Gradst. var. *pappeana*	Lejeuneaceae		e							e	e*														
Schiffneriolejeunea polycarpa (Nees) Gradst.	Lejeuneaceae		e*							e*															
Spruceanthus abbreviatus (Mitt.) X.Q.Shi. R.L.Zhu et Gradst.	Lejeuneaceae			e	e		e	e	e*																
Thysananthus auriculatus (Wilson) Sukkharak et Gradst. var. *auriculatus*	Lejeuneaceae						e																		
Bazzania decrescens (Lehm. et Lindenb.) Trevis.	Lepidoziaceae													t*											
Bazzania nitida (F.Weber) Grolle	Lepidoziaceae											t		e, rl, t											
Lepidozia cupressina (Sw.) Lindenb. ssp. *cupressina* Pócs	Lepidoziaceae													rl											
Lepidozia stuhlmannii Steph.	Lepidoziaceae																				b*	rl			
Lepidozia succida Mitt.	Lepidoziaceae										t*														
Telaranea coactilis (Spruce) J.J.Engel et G.L.Merr.	Lepidoziaceae										rl*														
Telaranea nematodes (Gottsche ex Austin) M.Howe	Lepidoziaceae										t	rl	rl	t	t			t		rl*	t				
Telaranea diacantha (Mont.) J.J.Engel et G.L.Merr.	Lepidoziaceae										rl, t	rl, t													
Cryptolophocolea martiana ssp. *martiana* (Nees) L.Söderstr. Crand.-Stotl. et Stotler	Lophocoleaceae	rl	rl							rl, t	rl, t, b*	rl, t	rl	rl	rl					rl, t					
Leptoscyphus infuscatus (Mitt.) E.W.Jones ex Grolle	Lophocoleaceae																					e, rl		rl*	t

(Continued)

TABLE 2 | Continued

Plot		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Lophocolea bidentata (L.) Dumort.	Lophocoleaceae															rl, t		e, rl*	rl	rl	rl*	rl*	rl*	e, rl	rl, t*
Lophocolea concreta Mont.	Lophocoleaceae		rl															rl		rl					
Lophocolea difformis Nees	Lophocoleaceae					rl																			
Lophocolea muricata (Lehm.) Nees	Lophocoleaceae										e	rl		t	rl	t		t	rl	rl				rl*	t*
Andrewsianthus bilobus (Mitt.) Grolle	Lophoziaceae																					rl			
Metzgeria consanguinea Schiffn.	Metzgeriaceae					e													e, t*	e*	rl*	rl*		e	rl
Metzgeria crassipilis (Lindb.) A.Evans	Metzgeriaceae																	e						e*	
Metzgeria furcata (L.) Corda	Metzgeriaceae	e	e	e	e	e				e		e		e	e	t			e, rl	e				e	e
Metzgeria madagassa Steph.	Metzgeriaceae	e								e, rl			e*	e*	e, rl	e*		rl		e					
Metzgeria quadrifaria Steph.	Metzgeriaceae									e		e				e				e		e*			
Pallavicinia lyellii (Hook.) Gray	Pallaviciniaceae											t, b*													
Plagiochila colorans Steph.	Plagiochilaceae																					e, rl*			
Plagiochila ericicola Steph.	Plagiochilaceae																					rl*			
Plagiochila fusifera Taylor	Plagiochilaceae											e		e, t	e, rl										
Plagiochila heterostipa Steph.	Plagiochilaceae																				e*				
Plagiochila integerrima Steph.	Plagiochilaceae	rl, t		t*		rl*				rl	rl	e, rl*	e, rl*												
Plagiochila kiaeri Gottsche	Plagiochilaceae	e, t	e	e	e	e, rl				e, rl, t	e, rl, t		e		e, rl	e, rl, t	e	e*, rl	e	e, rl, b*				e, b*	e, t*
Plagiochila kiaeri var. *myriocarpa* (Pearson) Pócs	Plagiochilaceae		e									e	e*	e*	e										
Plagiochila lastii Mitt.	Plagiochilaceae											e, rl		e*	e				t*, b*	e, rl					
Plagiochila pectinata Lindenb.	Plagiochilaceae																		e						
Plagiochila pinniflora Steph.	Plagiochilaceae		e			e*, rl*																			
Plagiochila sp. 1	Plagiochilaceae			rl*	e*		rl*																		
Plagiochila sp. 3	Plagiochilaceae																						e *		
Plagiochila sp. 4	Plagiochilaceae																							rl *	
Plagiochila squamulosa Mitt.	Plagiochilaceae																				e*			e, rl	e*, rl
Plagiochila squamulosa var. *crispulo-caudata* (Gottsche) Vanden Berghen	Plagiochilaceae										t														
Plagiochila squamulosa var. *sinuosa* (Mitt.) Vanden Berghen	Plagiochilaceae															rl		e, rl, b*		e, rl	e, b*			e, rl	e, rl

(Continued)

TABLE 2 | Continued

Plot		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Plagiochila strictifolia Steph.	Plagiochilaceae	e, rl		e	rl	e, rl*	e	e	e	e, rl*, t			e												
Plagiochila terebrans Nees et Mont.	Plagiochilaceae		e			e				e	e, rl, t	e	e*	e, t	e				e, b*						
Porella abyssinica var. *hoehnelii* (Steph.) Pócs	Porellaceae	rl				e																			
Porella capensis (Gottsche) Mitt.	Porellaceae										e														
Porella subdentata (Mitt.) Steph. var. *subdentata*	Porellaceae	e	e, t	e, rl	e, rl	e, rl*					e														
Porella subdentata var. *camerunensis* E.W.Jones	Porellaceae									e*															
Radula ankefinensis Gottsche	Radulaceae					e						e, rl			e					rl*					
Radula boryana (F.Weber) Nees ex Mont.	Radulaceae									e	e, rl	e, b*		e*											
Radula flaccida Lindenb. et Gottsche	Radulaceae	e				e	e	e*	e			e													
Radula fulvifolia (Hook.f. et Taylor) Gottsche. Lindenb. et Nees	Radulaceae	e		e						e, rl, b*	e, t			e											
Radula quadrata Gottsche	Radulaceae													e*							e				
Radula stenocalyx Mont.	Radulaceae					e*																			
Radula voluta Taylor	Radulaceae													e					e*, t*					e	
Solenostoma borgenii (Gottsche) Steph.	Solenostomataceae																					t*			

Sites: KNP, Kibale National Park; SNP, Semliki National Park; BINP, Bwindi Impenetrable National Park; MGNP, Mgahinga Gorilla National Park. Underlined species names indicate the first record of the taxon for Uganda. Habitats: t, terrestrial; e, epiphytic; rl, rotten log; b, tree base (below 1 m), asterisks mark records from outside subplots.

The best models for the regional variables explained 50% of the variance in total species richness (**Figure 3** and **Table 1**). Roughly half of the explained variance (25.1%), was accounted for by annual mean temperature (Temp + $Temp^2$) followed by temperature seasonality (TempS; 12.5%) and precipitation seasonality (PrecS; 12.5%). Epiphytic species richness showed a very similar pattern, with 54% of the variance explained by the same factors in roughly similar proportions. Non-epiphytic species richness showed a different pattern, with 25% of variance explained, half of which was accounted for by annual mean temperature (Temp + $Temp^2$; 11.4%), followed by annual precipitation (Prec; 4.8%), and temperature seasonality (TempS; 4%) (**Figure 3**).

The models considering local environmental variables alone had lower explanatory power than the regional variables for total and epiphytic richness, whereas the reverse was true for non-epiphytic richness (**Figure 3** and **Table 1**). For total richness, the models explained 21.8% of the variance, with bryophyte cover of branches (BCov) as the most important contributor (6.3%), followed by canopy height (HC; 4.3%) and distance to open water (DW; $\sim$ 4%) as well as plant cover of ground (GCov; $\sim$ 4%). For epiphytic richness, variance explained was only 13.1%, with bryophyte cover of branches (BCov) and plant cover of ground (GCov) as main factors (4% and 3%, respectively). For non-epiphytic species richness, the models explained 29.9% of the variance, largely accounted for by bryophyte cover of branches (BCov; 11.9%) followed by canopy height (HC; 5.8%) (**Figure 3**).

Combining regional and local environmental variables explained 56.3% of the variance of the overall species richness, largely accounted for by regional factors (40.6%) and with bryophyte cover of branches (BCov; 13.3%) as the only important local factor (**Figure 3**). Epiphytic richness showed a very similar pattern with 55.4% of variance explained. Non-epiphytic richness had a similar total explained variance (55.2%), but with local factors (27.3%) equally important as the regional factors (27.8%). In particular, canopy height (HC 4.7%) and inclination (Inc 4.3%), which played essentially no role for the total and epiphytic richness, had important contributions for the non-epiphytes (**Figure 3**).

Regressing richness against $Temp^2$ only had significant independent effects on total and epiphytic species richness (adjusted $R^2 = 0.15$ and $R^2 = 0.24$, respectively, $p < 0.001$ each). TempS accounted for 10% of variance explained ($p < 0.01$) in non-epiphytic richness, BCov only explained 16% and 34% of variance in total and non-epiphytic richness, respectively ($p < 0.001$; **Supplementary Figure S1**). The other variables had no noteworthy independent effects on species richness ($R^2 > 0.2$ and $p < 0.01$).

Analyzing species richness based on the subplot records only resulted in higher explanatory power of the regional variables regarding the total species richness and the non-epiphytic species richness (56% and 33%), and almost similar explanatory power of the local variable set (**Figure 3**). Both variable sets performed better in explaining total epiphytic species diversity of the plots compared to the subplot records. The combined dataset performed roughly equally for the total species richness of the plots and the subplots, respectively. In contrast, the analyses of the subplot records decreased 7.9% of explained variance in epiphytic and non-epiphytic species richness with respect to the plot records. The relative importance of the local variables declined markedly in all three subplot-based species richness analyses compared to the plot-based diversity. Contrasting with this trend, the proportion of annual mean temperature (Temp + $Temp^2$) increased.

DISCUSSION

The main result of our study is that both regional and local environmental factors influence liverwort richness in montane forests in Uganda, and that they interact in complex ways that affect epiphytic and non-epiphytic bryophytes differently. Similar interactions of regional and local factors have also been found for other plant groups such as ferns (Weigand et al., 2019).

Specifically, we found that total and epiphytic liverwort diversity was largely influenced by regional climatic factors, whereas non-epiphytic diversity was accounted for to similar degrees by regional and local factors.

The distribution of overall liverwort diversity was largely driven by regional climatic factors (temperature and precipitation, and their seasonality) and air humidity (as indicated by epiphytic bryophyte cover; Karger et al., 2012), so that ultimately temperature- and humidity-related variables had roughly equal contributions to total variance explained. The importance of these climatic factors in explaining geographical patterns of species richness in bryophytes is well known (Aranda et al., 2014; Medina et al., 2014; Chen et al., 2015), and can be linked to physiological limitations imposed by high temperatures and low humidity at low latitudes and elevations, and low temperatures at high latitudes and elevations (Rütten and Santarius, 1992; Leon-Vargas et al., 2006). The concordance of overall and epiphytic diversity patterns suggests that the overall pattern of species richness in liverworts in our study region was largely driven by the epiphytic species, which is unsurprising considering that they contributed 74.8% of the total species richness. Terrestrial bryophytes are scarce especially in tropical lowlands, where they are restricted to small patches of open soil (Richards, 1954; Pócs, 1982; Holz et al., 2002).

It is well known that especially epiphytic bryophyte communities are also influenced by other factors such as the characteristics of their host trees, for instance bark texture and pH or trunk diameter (Studlar, 1982; González-Mancebo et al., 2003; Gradstein and Culmsee, 2010). Our study did not assess these factors, but we do not consider that they strongly influence our results as our sampling strategy covered numerous trees at each site and elevation. As the forests studied by us are typically dominated by one or a few tree species (see the section "Materials and Methods"), our sampling covered these species. Since our study was focused on natural forests, the tree composition of the forests was determined by the combination of edaphic and climatic conditions of the study sites. Accordingly, even though tree-specific associations of the bryophytes were not included directly by us, their effect is to some degree indirectly covered by the environmental data included in the analyses.

In contrast to the overall pattern, non-epiphytic liverwort richness showed a quite distinct pattern, with a higher contribution of local environmental factors than epiphytes. This difference in epiphytic and non-epiphytic patterns likely reflects the different influences affecting their growth, in particular that epiphytic liverworts are more closely linked to climatic conditions (Medina et al., 2014), whereas non-epiphytic liverworts also depend on the local soil and microclimatic conditions. Both the impact of the amount of bare soil as well as site-specific climate conditions on non-epiphytic bryophyte species richness has been documented earlier (Pharo and Beattie, 2002; Mandl et al., 2009). Specifically, our analysis shows important effects of canopy height and slope inclination. The relationship to canopy height likely reflects the fact that tall forests at lower elevations provide highly unsuitable conditions for non-epiphytic bryophytes, with low light incidence at ground levels (Bazzaz and Pickett, 1980), and high accumulation of leaf litter (Richards, 1954). In the lowlands and lower montane forests, terrestrial species thus almost exclusively occur on disturbed soil surfaces like termite hills and earth walls (Richards, 1954; Pócs, 1982), which are formed by digging activities of animals or thrown up by uprooted trees (Jonsson and Esseen, 1990). Slope inclination is important for bryophytes because steeper slopes, depending on the soil properties, have more and larger canopy gaps created by uprooting events (Ohkubo et al., 2007; de Lima and de Moura, 2008), and hence, provide increased light exposure and patches of bare soil as well as dead wood.

Comparing the resulting RVI with the univariate regressions of the respective variables furthermore indicates that the balance of co-acting effects of all variables in the models are determinant for the richness patterns in diversity.

Focusing on the sampling strategy, we found that our standardized sampling strategy with small subplots included the majority of species present in each plot. Generally speaking, the results obtained by including only species recorded in the subplots versus including also the additionally recorded species in the entire plot were largely concordant. There was one minor, but potentially interesting difference, however: when we only included the species recorded in the subplots, the importance of local factors diminished in the GLM analyses. We interpret this as reflecting that the subplots predominantly covered ecologically generalist species, as 72.7% of the omitted species from the total plot collection were found in only one plot. Thus, the declining importance of local predictors for the subplot diversity probably reflects the good adaptability of the more widely distributed species and neglects the less widely distributed species with narrower environmental tolerance included in the plot-based analyses. It would appear that the latter species are more dependent on local microhabitat conditions.

Our study therefore confirms that the richness of epiphytic liverworts is closely linked to regional climatic conditions, whereas the occurrence of non-epiphytic liverworts depends much more on local microhabitat-specific factors. This implies, among other aspects, that species distribution models, which generally largely rely on macroclimatic factors (Lembrechts et al., 2019), would be expected to perform well for epiphytic bryophytes, but much more poorly for non-epiphytic and rare ones. For the latter, local-scale factors, which are generally not available in macroecological studies, would be needed to accurately predict their probabilities of occurrence. As already suggested by Zotz and Bader (2009) it may also be hypothesized that epiphytic bryophyte richness may react more directly to changes in regional climate in the course of global change than non-epiphytic richness, which may more strongly depend on the availability of small-scale microrefugia (Mandl et al., 2009; Raabe et al., 2010; Sun et al., 2013).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

KM, MK, and DQ designed this study. DQ, B-CH, KM, and MK conducted the field work. DQ, KM, and B-CH designed the set-up of the collection. B-CH, DQ, and KM collected the bryophyte specimens. MK recorded the local parameters of the plots. Y-MW identified the leafy liverworts. MN identified the thallose liverworts. B-CH processed the specimens and associated the labels. KM performed the analyses and wrote the manuscript. FL contributed to and revised the statistical analyses. All co-authors provided critical comments on the manuscript and have read and approved the submitted manuscript.

FUNDING

This work was funded by the German Research Foundation (DFG; Grant DQ 153/8) and the Swiss National Science Foundation (SNF; Grant 310030L_146906/1).

ACKNOWLEDGMENTS

We thank the national authorities in Uganda for granting the permits to carry out field work and collect voucher specimens. We are grateful to our counterpart Felly Mugizi Tusiime for facilitating fieldwork and the Herbarium of Makerere University for taking care of our specimens. We further thank D. N. Karger for statistical advice, and S. Noben and A. Weigand for their advice concerning organization and handling of the climate and plot-based data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00765/full#supplementary-material

REFERENCES

Akayezu, P., van Duren, I. C., Groen, T. A., Grueter, C. C., and Robbins, M. M. (2019). Abundance and spatial distribution of the main food species for mountain gorillas in the Virunga Massif. Rwanda. *Biodivers. Conserv.* 28, 3597–3620. doi: 10.1007/s10531-019-01838-0

Aranda, S. C., Gabriel, R., Borges, P. A. V., Santos, A. M. C., de Azevedo, E. B., Patiño, J., et al. (2014). Geographical, temporal and environmental determinants of bryophyte species richness in the Macaronesian Islands. *PLoS One* 9:e101786. doi: 10.1371/journal.pone.0101786

Araujo, M. B., and Rahbek, C. (2006). How does climate change affect biodiversity? *Science* 313, 1396–1397. doi: 10.1126/science.1131758

Barton, K. (2019). *MuMIn: Multi-Model Inference*. Available online at: https://CRAN.R-project.org/package=MuMIn (accessed May 21, 2019).

Bazzaz, F. A., and Pickett, S. T. A. (1980). Physiological ecology of tropical succession: A comparative review. *Ann. Rev. Ecol. Syst.* 11, 287–310. doi: 10.1146/annurev.es.11.110180.001443

Bruun, H. H., Moen, J., Virtanen, R., Grytnes, J., Oksanen, L., and Angerbjörn, A. (2006). Effects of altitude and topography on species richness of vascular plants, bryophytes and lichens in alpine communities. *J. Veg. Sci.* 17, 37–46. doi: 10.1111/j.1654-1103.2006.tb02421.x

Butynski, T. M., and Kalina, J. (1993). Three new mountain national parks for Uganda. *Oryx* 27, 214–224. doi: 10.1017/S003060530002812X

Cameron, C. A., and Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *J. Econ.* 77, 329–342. doi: 10.1016/S0304-4076(96)01818-0

Chalcraft, D. R. (2013). Changes in ecological stability across realistic biodiversity gradients depend on spatial scale. *Glob. Ecol. Biogeogr.* 22, 19–28. doi: 10.1111/j.1466-8238.2012.00779.x

Chen, H. (2018). *VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20*. Available online at: https://CRAN.R-project.org/package=VennDiagram (accessed September 11, 2019).

Chen, S., Ferry Slik, J. W., Mao, L., Zhang, J., Sa, R., Zhou, K., et al. (2015). Spatial patterns and environmental correlates of bryophyte richness: sampling effort matters. *Biodivers. Conserv.* 24, 593–607. doi: 10.1007/s10531-014-0838-8

Colwell, R. K., Gotelli, N. J., Ashton, L. A., Beck, J., Brehm, G., Fayle, T. M., et al. (2016). Midpoint attractors and species richness: modelling the interaction between environmental drivers and geometric constraints. *Ecol. Lett.* 19, 1009–1022. doi: 10.1111/ele.12640

Crawley, M. J., and Harral, J. E. (2001). Scale dependence in plant biodiversity. *Science* 291, 864–868. doi: 10.1126/science.291.5505.864

de Lima, R. A. F., and de Moura, L. C. (2008). Gap disturbance regime and composition in the Atlantic Montane Rain Forest: the influence of topography. *Plant Ecol.* 197, 239–253. doi: 10.1007/s11258-007-9374-x

Geffert, J. L., Frahm, J.-P., Barthlott, W., and Mutke, J. (2013). Global moss diversity: spatial and taxonomic patterns of species richness. *J. Bryol.* 35, 1–11. doi: 10.1179/1743282012Y.0000000038

Glime, J. M. (2017). *Bryophyte Ecology Volume 1: Physiological Ecology*. Michigan: Michigan Technological University and the International Association of Bryologists.

González-Mancebo, J. M., Losada-Lima, A., and McAlister, S. (2003). Host specificity of epiphytic bryophyte communities of a laurel forest on Tenerife (Canary Islands. Spain). *Bryologist* 106, 383–394. doi: 10.1639/04

Gradstein, S. R. (2008). "Epiphytes of tropical montane forests- impact of deforestation and climate change," in *The Tropical Mountain Forest- Patterns and Processes in a Biodiversity Hotspot Biodiversity and Ecology Series*, eds S. R. Gradstein, J. Homeier, and D. Gansert (Göttingen: Universitätsverlag Göttingen), 220.

Gradstein, S. R., and Culmsee, H. (2010). Bryophyte diversity on tree trunks in montane forests of Central Sulawesi, Indonesia. *Trop. Bryol.* 31, 95–105. doi: 10.11646/bde.31.1.16

Grau, O., Grytnes, J.-A., and Birks, H. J. B. (2007). A comparison of altitudinal species richness patterns of bryophytes with other plant groups in Nepal. Central Himalaya. *J. Biogeogr.* 34, 1907–1915. doi: 10.1111/j.1365-2699.2007.01745.x

Hawkins, B. A. (2001). Ecology's oldest pattern? *Trends Ecol. Evol.* 16:470. doi: 10.1016/S0169-5347(01)02197-8

Henriques, D. S. G., Borges, P. A. V., Ah-Peng, C., and Gabriel, R. (2016). Mosses and liverworts show contrasting elevational distribution patterns in an oceanic island (Terceira, Azores): the influence of climate and space. *J. Bryol.* 38, 183–194. doi: 10.1080/03736687.2016.1156360

Hitimana, J., Namara, A., Sengalama, T., and Nyirimana, J. (2006). *Community-Based Natural Resource Management (CBNRM) Plan*. Kinigi Area: CBNRM.

Holz, I., Gradstein, S. R., Heinrichs, J., and Kappelle, M. (2002). Bryophyte diversity, microhabitat differentiation, and distribution of life forms in Costa Rican upper montane Quercus forest. *Bryologist* 105, 334–348. doi: 10.1639/0007-2745(2002)105[0334:bdmdad]2.0.co;2

Howard, P. C. (1991). *Nature Conservation in Uganda's Tropical Forest Reserves*. Gland: IUCN.

Howard, P. C., Davenport, T. R. B., Kigenyi, F. W., Viskanic, P., Baltzer, M. C., Dickinson, C. J., et al. (2000). Protected area planning in the tropics: Uganda's national system of forest nature reserves. *Conserv. Biol.* 14, 858–875. doi: 10.1046/j.1523-1739.2000.99180.x

Hurvich, C. M., and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297–307. doi: 10.1093/biomet/76.2.297

Jonsson, B. G., and Esseen, P.-A. (1990). Treefall disturbance maintains high bryophyte diversity in a boreal spruce forest. *J. Ecol.* 78, 924–936.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2017). Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4:170122. doi: 10.1038/sdata.2017.122

Karger, D. N., Kluge, J., Abrahamczyk, S., Salazar, L., Homeier, J., Lehnert, M., et al. (2012). Bryophyte cover on trees as proxy for air humidity in the tropics. *Ecol. Indicat.* 20, 277–281. doi: 10.1016/j.ecolind.2012.02.026

Kasenene, J. (2001). "Lost logging– problems of tree regeneration in forest gaps in Kibale Forest, Uganda," in *African Rain Forest Ecology and Conservation. An Interdisciplinary Perspective*, eds W. Weber, L. J. T. White, A. Vedder, and L. Naughton-Treves (New Haven: Yale University Press), 588.

Kessler, M., Kluge, J., Hemp, A., and Ohlemüller, R. (2011). A global comparative analysis of elevational species richness patterns of ferns. *Glob. Ecol. Biogeogr.* 20, 868–880. doi: 10.1111/j.1466-8238.2011.00653.x

Kluge, J., and Kessler, M. (2011). Phylogenetic diversity, trait diversity and niches: species assembly of ferns along a tropical elevational gradient. *J. Biogeogr.* 38, 394–405. doi: 10.1111/j.1365-2699.2010.02433.x

Kluge, J., Worm, S., Lange, S., Long, D., Böhner, J., Yangzom, R., et al. (2017). Elevational seed plants richness patterns in Bhutan. Eastern Himalaya. *J. Biogeogr.* 44, 1711–1722. doi: 10.1111/jbi.12955

Körner, C. (2000). Why are there global gradients in species richness? Mountains might hold the answer. *Trends Ecol. Evol.* 15, 513–514. doi: 10.1016/s0169-5347(00)02004-8

Kraichak, E. (2012). Asexual propagules as an adaptive trait for epiphylly in tropical leafy liverworts (Lejeuneaceae). *Am. J. Bot.* 99, 1436–1444. doi: 10.3732/ajb.1200120

Kürschner, H. (2004). Life strategies and adaptations in bryophytes from the Near and Middle East. *Turki. J. Bot.* 28, 73–84.

Langdale-Brown, I., Ostmaston, H. A., and Wilson, J. G. (1964). *The Vegetation of Uganda and Its Bearing on Land-Use*. Entebbe: Government of Uganda.

Lembrechts, J. J., Nijs, I., and Lenoir, J. (2019). Incorporating microclimate into species distribution models. *Ecography* 42, 1267–1279. doi: 10.1111/ecog.03947

Leon-Vargas, Y., Engwald, S., and Proctor, M. C. F. (2006). Microclimate, light adaptation and desiccation tolerance of epiphytic bryophytes in two Venezuelan cloud forests. *J. Biogeogr.* 33, 901–913. doi: 10.1111/j.1365-2699.2006.01468.x

Longton, R. E. (1982). "Bryophyte vegetation in polar regions," in *Bryophyte Ecology*, ed. A. J. E. Smith (Dordrecht: Springer Netherlands), 123–165. doi: 10.1007/978-94-009-5891-3_5

Mandl, N. A., Kessler, M., and Gradstein, S. R. (2009). Effects of environmental heterogeneity on species diversity and composition of terrestrial bryophyte assemblages in tropical montane forests of southern Ecuador. *Plant Ecol. Divers.* 2, 313–321. doi: 10.1080/17550870903341877

McCain, C. M. (2005). Elevational gradients in diversity of small mammals. *Ecology* 86, 366–372. doi: 10.1890/03-3147

McCain, C. M., and Grytnes, J.-A. (2010). "Elevational gradients in species richness," in *Encyclopedia of Life Sciences (ELS)*, (Chichester: John Wiley & Sons, Ltd), doi: 10.1002/9780470015902.a0022548

Medina, N. G., Albertos, B., Lara, F., Mazimpaka, V., Garilleti, R., Draper, D., et al. (2014). Species richness of epiphytic bryophytes: drivers across scales on the edge of the Mediterranean. *Ecography* 37, 80–93. doi: 10.1111/j.1600-0587.2013.00095.x

Ohkubo, T., Tani, M., Noguchi, H., Yamakura, T., Itoh, A., Kanzaki, M., et al. (2007). Spatial and topographic patterns of canopy gap formation in a mixed dipterocarp forest in Sarawak. Malaysia. *Tropics* 16, 151–163. doi: 10.3759/tropics.16.151

Owiunji, I., Nkuutu, D., Kujirakwinja, D., Liengola, I., Plumptre, A., Nsanzurwimo, A., et al. (2005). *The biodiversity of the Virunga Volcanoes*. Available online at: https://programs.wcs.org/portals/49/media/file/volcanoes_biodiv_survey.pdf (accessed September 10, 2019).

Pharo, E. J., and Beattie, A. J. (2002). The association between substrate variability and bryophyte and lichen diversity in eastern Australian forests. *Bryologist* 105, 11–26. doi: 10.1639/0007-2745(2002)105[0011:tabsva]2.0.co;2

Pócs, T. (1982). "Tropical forest bryophytes," in *Bryophyte Ecology*, ed. A. J. E. Smith (Dordrecht: Springer Netherlands), 59–104. doi: 10.1007/978-94-009-5891-3_3

Proctor, M. C. F., Oliver, M. J., Wood, A. J., Alpert, P., Stark, L. R., Cleavitt, N. L., et al. (2007). Desiccation-tolerance in bryophytes: a review. *Bryologist* 110, 595–621.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raabe, S., Müller, J., Manthey, M., Dürhammer, O., Teuber, U., Göttlein, A., et al. (2010). Drivers of bryophyte diversity allow implications for forest management with a focus on climate change. *Forest Ecol. Manag.* 260, 1956–1964. doi: 10.1016/j.foreco.2010.08.042

Richards, P. W. (1954). Notes on the bryophyte communities of lowland tropical rain forest, with special reference to Moraballi Creek, British Guiana. *Veg. Acta Geobot* 5, 319–328. doi: 10.1007/BF00299586

Ring, U. (2008). Extreme uplift of the Rwenzori Mountains in the East African Rift, Uganda: Structural framework and possible role of glaciations. *Tectonics* 27, 1–19. doi: 10.1029/2007TC002176

Rütten, D., and Santarius, K. A. (1992). Relationship between frost tolerance and sugar concentration of various bryophytes in summer and winter. *Oecologia* 91, 260–265. doi: 10.1007/BF00317794

Schlüter, T. (2006). *Geological Atlas of Africa: With Notes on Stratigraphy, Tectonics, Economic Geology, Geohazards, and Geosites of Each Country*. Berlin: Springer-Verlag.

Seppelt, R. D., and Green, T. G. A. (1998). A bryophyte flora for Southern Victoria Land. Antarctica. *New Zealand J. Bot.* 36, 617–635. doi: 10.1080/0028825X.1998.9512599

Shaw, A. J., Cox, C. J., and Goffinet, B. (2005). Global patterns of moss diversity: taxonomic and molecular inferences. *Taxon* 54, 337–352. doi: 10.2307/25065362

Söderström, L., Hagborg, A., von Konrat, M., Bartholomew-Began, S., Bell, D., Briscoe, L., et al. (2016). World checklist of hornworts and liverworts. *PhytoKeys* 59, 1–828. doi: 10.3897/phytokeys.59.6261

Song, L., Ma, W.-Z., Yao, Y.-L., Liu, W.-Y., Li, S., Chen, K., et al. (2015). Bole bryophyte diversity and distribution patterns along three altitudinal gradients in Yunnan, China. *J. Veg. Sci.* 26, 576–587. doi: 10.1111/jvs.12263

Sonnleitner, M., Dullinger, S., Wanek, W., and Zechmeister, H. (2009). Microclimatic patterns correlate with the distribution of epiphyllous bryophytes in a tropical lowland rain forest in Costa Rica. *J. Trop. Ecol.* 25, 321–330. doi: 10.1017/S0266467409006002

Spitale, D. (2016). The interaction between elevational gradient and substratum reveals how bryophytes respond to the climate. *J. Veg. Sci.* 27, 844–853. doi: 10.1111/jvs.12403

Stewart, K. J., and Mallik, A. U. (2006). Bryophyte responses to microclimatic edge effects across riparian buffers. *Ecol. Appl.* 16, 1474–1486. doi: 10.1890/1051-0761(2006)016[1474:brtmee]2.0.co;2

Studlar, S. M. (1982). Host specificity of epiphytic bryophytes near Mountain Lake. Virginia. *Bryologist* 85, 37–50. doi: 10.2307/3243139

Sun, S.-Q., Wu, Y.-H., Wang, G.-X., Zhou, J., Yu, D., Bing, H.-J., et al. (2013). Bryophyte species richness and composition along an altitudinal gradient in Gongga Mountain. China. *PLoS One* 8:e58131. doi: 10.1371/journal.pone.0058131

Thomas, S. C., Liguori, D. A., and Halpern, C. B. (2001). Corticolous bryophytes in managed Douglas-fir forests: habitat differentiation and responses to thinning and fertilization. *Can. J. Bot.* 79, 886–896. doi: 10.1139/cjb-79-8-886

Tusiime, F. M., Byarujali, S. M., and Bates, J. W. (2007). Diversity and distribution of bryophytes in three forest types of Bwindi Impenetrable National Park. Uganda. *Afr. J. Ecol.* 45, 79–87. doi: 10.1111/j.1365-2028.2007.00862.x

Wang, J., Vanderpoorten, A., Hagborg, A., Goffinet, B., Laenen, B., and Patiño, J. (2017). Evidence for a latitudinal diversity gradient in liverworts and hornworts. *J. Biogeogr.* 44, 487–488. doi: 10.1111/jbi.12909

Weigand, A., Abrahamczyk, S., Aubin, I., Bita-Nicolae, C., Bruelheide, H., Carvajal-Hernández, C. I, et al. (2019). Global fern and lycophyte richness explained: How regional and local factors shape plot richness. *J. Biogeogr.* 60, 59–71. doi: 10.1111/jbi.13782

Wigginton, M. J. (2018). *Checklist and Distribution of the Liverworts and Hornworts of sub-Saharan Africa, including the East African Islands (Edition 4)*. London: Tropical Bryology Research.

Wilson, J. P., and Gallant, J. C. (eds) (2000). *Terrain Analysis: Principles and Applications*. New York, NY: John Wiley & Sons, Ltd.

Wolf, J. H. D. (1993). Diversity patterns and biomass of epiphytic bryophytes and lichens along an altitudinal gradient in the northern Andes. *Ann. Missouri Bot. Garden* 80, 928–960. doi: 10.2307/2399938

Zellweger, F., Braunisch, V., Morsdorf, F., Baltensweiler, A., Abegg, M., Roth, T., et al. (2015). Disentangling the effects of climate, topography, soil and vegetation on stand-scale species richness in temperate forests. *Forest Ecol. Manag.* 349, 36–44. doi: 10.1016/j.foreco.2015.04.008

Zhang, D. (2018). *rsq: R-Squared and Related Measures*. Available online at: https://CRAN.R-project.org/package=rsq (accessed July 15, 2019).

Zotz, G., and Bader, M. Y. (2009). "Epiphytic plants in a changing world-global change effects on vascular and non-vascular epiphytes," in *Progress in Botany*, eds U. Lüttge, W. Beyschlag, B. Büdel, and D. Francis (Berlin Heidelberg: Springer), 147–170. doi: 10.1007/978-3-540-68421-3_7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 26 June 2020
doi: 10.3389/fpls.2020.00829

Evolutionary History of the *Marchantia polymorpha Complex*

Anna-Malin Linde[1], Weerachon Sawangproh[2,3], Nils Cronberg[2], Péter Szövényi[4] and Ulf Lagercrantz[1]*

[1] Plant Ecology and Evolution, Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden, [2] Biodiversity, Department of Biology, Lund University, Lund, Sweden, [3] Division of Conservation Biology, School of Interdisciplinary Studies, Mahidol University, Kanchanaburi, Thailand, [4] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

The potential role of introgression in evolution has gained increased interest in recent years. Although some fascinating examples have been reported, more information is needed to generalize the importance of hybridization and introgression for adaptive divergence. As limited data exist on haploid dominant species, we analyzed genomes of three subspecies of the liverwort *Marchantia polymorpha*. We used available genomic data for subsp. *ruderalis* and carried out whole-genome (PacBio) sequencing for one individual each of subsp. *montivagans* and subsp. *polymorpha* as well as Illumina resequencing of additional genomes for all three subspecies. The three subspecies were compared against *M. paleacea* as outgroup. Our analyses revealed separation of the three taxa, but all three possible topologies were richly represented across the genomes, and the underlying divergence order less obvious. This uncertainty could be the result of the divergence of the three subspecies close in time, or that introgression has been frequent since divergence. In particular, we found that pseudo-chromosome 2 in subsp. *montivagans* was much more diverged than other parts of the genomes. This could either be explained by specific capture of chromosome 2 from an unknown related species through hybridization or by conservation of chromosome 2 despite intermittent or ongoing introgression affecting more permeable parts of the genomes. A higher degree of chromosomal rearrangements on pseudo-chromosome 2 support the second hypothesis. Species tree analyses recovered an overall topology where subsp. *montivagans* diverged first and subsp. *ruderalis* and subsp. *polymorpha* appeared as sister lineages. Each subspecies was associated with its own chloroplast and mitochondrial haplotype group. Our data suggest introgression but refute a previous hypothesis that subsp. *ruderalis* is a new stabilized hybrid between the other two subspecies.

Keywords: *Marchantia polymorpha*, hybridization, bryophytes, incomplete lineage sorting, whole-genome sequencing, phylogeny, introgression

OPEN ACCESS

Edited by:
Meenu Kapoor, Guru Gobind Singh Indraprastha University, India

Reviewed by:
Matthew Johnson, Texas Tech University, United States
Julissa Roncal, Memorial University of Newfoundland, Canada

****Correspondence:***
Nils Cronberg
nils.cronberg@biol.lu.se

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *21 January 2020*
Accepted: *22 May 2020*
Published: *26 June 2020*

Citation:
Linde A-M, Sawangproh W, Cronberg N, Szövényi P and Lagercrantz U (2020) Evolutionary History of the Marchantia polymorpha Complex. Front. Plant Sci. 11:829. doi: 10.3389/fpls.2020.00829

INTRODUCTION

Hybridization among diverging lineages is not uncommon in nature, especially in rapidly radiating groups (Seehausen, 2004; Grant et al., 2005; Mallet, 2005, 2007; Fontaine et al., 2015). This process may attenuate divergence, introduce adaptive divergence from another population, or even create a new hybrid species. Even though hybridization is widespread, hybrid speciation is probably

rare. When it happens, it is most often in the form of allopolyploid hybrid speciation (Soltis and Soltis, 2009) and documented instances of homoploid speciation are few (Rieseberg et al., 2003). A more frequent outcome of hybridization is introgression, the transfer of genetic material between species through hybridization and repeated backcrossing (Anderson and Hubricht, 1938). This may increase standing variation and adaptive divergence. Genetic recombination can quickly generate novel genotypes from existing nucleotide variation and may thus have an important role in adaptive evolution. Even if hybridization itself might be rare, introgression may provide new genetic variants at a higher frequency than *de novo* mutations (Ward and van Oosterhout, 2016; Martin and Jiggins, 2017). An increasing number of studies report evidence of introgression occurring across species boundaries as a consequence of hybridization (reviewed by Dowling and Secor, 1997) or horizontal gene transfer (Gogarten and Townsend, 2005; Galtier and Daubin, 2008) resulting in reticulate evolution with different parts of the genomes more or less exposed to gene transfer (Harrison and Larson, 2014).

Large scale genomic data provide an opportunity to characterize the history of hybridization and introgression (Payseur and Rieseberg, 2016). A complication for the analysis of such data is that incongruence between gene trees and the species tree could arise not only from hybridization/introgression but also from incomplete lineage sorting (ILS). In ILS ancestral polymorphisms persist over speciation events followed by chance fixation or persistent polymorphism in descendant lineages. Large population sizes and closely timed speciation events will increase the frequency of incongruent gene trees arising from ILS, and make the speciation and hybridization history more difficult to reveal (Rieseberg et al., 1999; Mossel and Roch, 2010).

Most genomic studies of hybridization and introgression have so far been conducted on organisms with a diploid dominant generation. It is thus of interest to study more cases having a dominant haploid generation and a short-lived sporophyte generation. In bryophytes, the diploid sporophyte is the actual hybrid combining the parental genomes (comparable to the F1 generation in a vascular plant). Meiosis takes place in the spore capsule, prior to spore formation, so the haploid spores represent recombinants of the parental genomes (comparable to the F2 generation in vascular plant). The spores are usually produced in massive amounts and wind-dispersed. After spore germination, the gametophyte phenotype is directly exposed to selection, no variation is masked by dominant alleles, so that favorable genes transferred to a new genomic background can potentially show a fast penetration in populations through clonal growth or secondary back-crossing (Shaw, 1994, 1998; Natcheva and Cronberg, 2007; reviewed by Natcheva and Cronberg, 2004).

Marchantia polymorpha L. is often treated as a complex of three subspecies which together has a cosmopolitan distribution, although introduced in some parts of the Southern Hemisphere (Paton, 1999). *Marchantia polymorpha* is one of the most studied species of liverworts but aspects of its phylogenetic relationships remain poorly resolved (Nishiyama et al., 2004; Qiu et al., 2006; Wickett et al., 2014), even after completion of whole-genome sequencing which was published in 2017 (Bowman et al., 2017). It has been used in botanical research for centuries, but has now been revived as a modern model plants to understand plant genetics and evolutionary processes (Shimamura, 2016). It is a thalloid liverwort, which can reproduce both sexually and asexually. Bryophytes (liverworts, mosses and hornworts) are the oldest of the extant lineages of land plants, and their position in the plant tree-of-life makes them interesting for studies concerning the evolution of land plants.

Following an early morphological taxonomic delimitation by Nees (1838) the *M. polymorpha*-complex was subdivided into three independent species. This subdivision was formalized by Burgeff (1943) as *M. polymorpha, M. alpestris* (Nees) Burgeff, and *M. aquatica* (Nees) Burgeff. Burgeff based this subdivision on restricted interfertility between the taxa in a reciprocal crossing experiment. The crosses between female *M. alpestris* and male *M. aquatica* rendered a relatively high frequency of viable spores (50–70% in all of five attempts with different accessions), whereas the other combinations were completely sterile (including the reciprocal cross, female *aquatica* x male *alpestris*) or nearly so. He was also able to repeatedly backcross female recombinants from the *alpestris* x *aquatica* cross with male *M. aquatica*. Later, observation of a recombinant in isozyme electrophoresis was taken as evidence of the occurrence of gene exchange between taxa (Bischler-Causse and Boisselier-Dubayle, 1991). Accordingly, the three taxa were instead recognized at the intraspecific level with the names commonly accepted today, subsp. *polymorpha*, subsp. *ruderalis* Bischl. & Boissel.-Dub. and subsp. *montivagans* Bischl. & Boissel.-Dub. Due to a typification error (the Linnean type for *M. polymorpha*, turned out to be the taxon found in aquatic environment), subsp. *ruderalis* refers to *M. polymorpha*, subsp. *montivagans* refers to *M. alpestris* and subsp. *polymorpha* refers to *M. aquatica, sensu* Burgeff. We use the names at subspecies level throughout this study, but we return to the question about taxonomic ranking in the discussion. The three subspecies of *M. polymorpha* are morphologically differentiated. *M. polymorpha* subsp. *polymorpha* shows thalli with distinct black continuous dark median line and appendage on innermost ventral scales with entire margin, whereas subsp. *montivagans* shows thalli without dark median line and appendage on innermost ventral scales with dentate margin. *M. polymorpha* subsp. *ruderalis* shows intermediate morphology between the other two subspecies by having thalli with discontinuous median line and appendage on innermost ventral scales with crenulated (projecting as low to sharp teeth) margin as shown in **Figure 1** (Paton, 1999; Atherton et al., 2010). The subspecies have been estimated to have diverged in the Late Miocene (ca. 5–7 MYA) (Villarreal et al., 2016). They are ecologically and partially geographically separated but can sometimes be found sympatrically (Schuster, 1992, reviewed by Shimamura, 2016) suggesting opportunities for hybridization. Among the three subspecies, subsp. *ruderalis* is the most common and it has been proposed to have originated as a relatively new stabilized hybrid between the two other subspecies, adapted to disturbed man-made habitats (Schuster, 1983, 1992). However, a limited electrophoretic study using four isozymes gave no support for this hypothesis (Boisselier-Dubayle and Bischler-Causse, 1989).

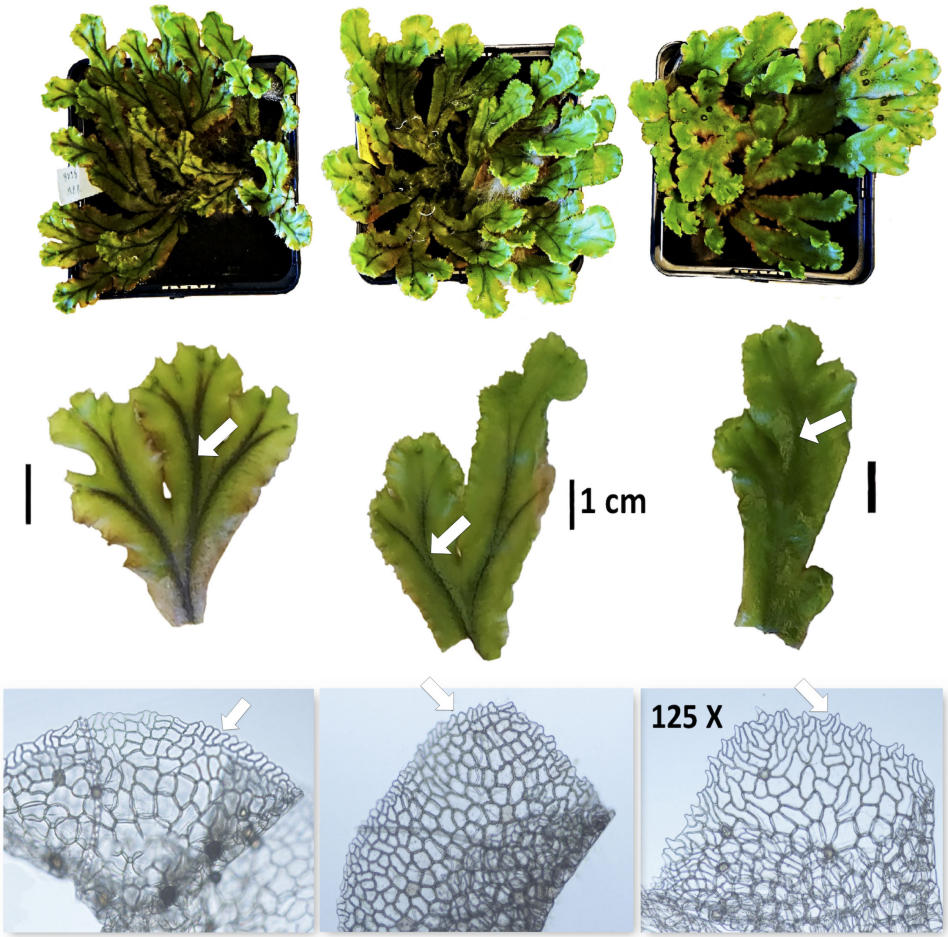


FIGURE 1 | Images of *M. polymorpha* subsp. *polymorpha* **(left)**, *M. polymorpha* subsp. *ruderalis* **(middle)** and *M. polymorpha* subsp. *montivagans* **(right)**. Note the differences in the size of thalli (first row), the thickness of the black midrib of thalli (second row, arrow) and morphology of innermost ventral scales with 125X magnification (third row, arrow).

MATERIALS AND METHODS

Plant Material and DNA Extraction

We sequenced 11 individuals of *M. polymorpha*; five representing subsp. *ruderalis*, three representing subsp. *polymorpha*, and three representing subsp. *montivagans*. All individuals were collected from locations in Sweden and Bulgaria (**Supplementary Table 1** and **Figure 2**). All living samples we used for DNA extraction were kept in culture at the Department of Biology (Lund University). DNA extraction was performed with Qiagen DNeasy Plant Minikit using young thallus tissues for Illumina sequencing, and with a modified CTAB protocol (Healey et al., 2014) for PacBio sequencing. The R package "ggmap" (Kahle and Wickham, 2013) was used to create **Figure 2**.

Genome Sequencing and Genome Assembly

One individual each of *M. polymorpha montivagans* (sample id MpmSA2) and *M. polymorpha polymorpha* (sample id MppBR5) were sequenced with Single-molecule real-time (SMRT) sequencing technology developed by Pacific BioSciences on a PacBio Sequel System with Sequel chemistry and sequence depth of 50X (Roberts et al., 2013). The reads were assembled using HGAP (Chin et al., 2013) and assembly statistics was assessed using QUAST (Gurevich et al., 2013) version 4.5.4, BUSCO (Simão et al., 2015) version 3.0.2 and CEGMA (Parra et al., 2007) version 2.5. BUSCO was used with the "Eukaryota odb9" dataset. Assembly quality and completeness are summarized in **Supplementary Table 2**.

For *M. polymorpha ruderalis* the publically available reference genome draft v.3.1 (Bowman et al., 2017) together with a chromosome-scale genome assembly (Diop et al., 2019) was used. The rest of the individuals in **Supplementary Table 1** were sequenced using Illumina HiSeq X sequencing platform with pair-end reads of 2 × 150 bp. The reads were mapped against the three *M. polymorpha* genome assemblies as described below. The genome of *M. paleacea* subsp. *diptera* was sequenced and assembled as described in details in Radhakrishnan (2017) and very short as following: Short-insert pair-end libraries were

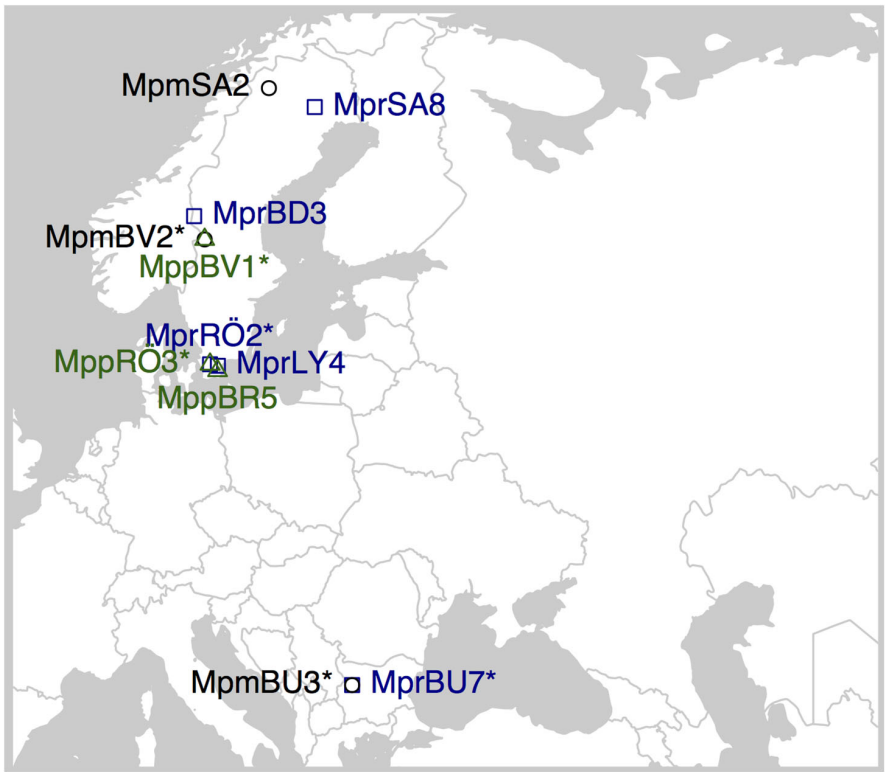


FIGURE 2 | Sampling locations for the *M. polymorpha* specimens used in this study. Overlapping symbols have an asterisk in the sample ID when sampled at the same location. MPM, *Marchantia polymorpha* subsp. *montivagans*; MPP, *M. polymorpha* subsp. *polymorpha*; MPR, *M. polymorpha* subsp. *ruderalis*.

produced using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA, United States) and long-insert mate-pair libraries were produced using Nextera Mate-Pair DNA library prep kit (Illumina, San Diego, CA, United States), following the manufacturer's protocol. The reads were trimmed using Trimmomatic v0.33 (Bolger et al., 2014) and SPAdes (Bankevich et al., 2012) were used to assemble the contigs. Scaffolding of the contigs was performed using the scaffolder of SOAPdenovo (Luo et al., 2012).

Preparation of Data: Alignment of Genomic Fragments (GFs)

The four genome assemblies (*M. polymorpha ruderalis, M. polymorpha montivagans, M. polymorpha polymorpha* and *M. paleacea diptera* were aligned using progressiveCactus (Paten et al., 2011a,b) version 2016-11-30 with default settings and with the following topology [*paleacea*, (*ruderalis*, *montivagans*, *polymorpha*)]. Cactus has been designed specifically to output HAL (hierarchical alignment) (Hickey et al., 2013). The resulting HAL-file was converted to MAF format using hal2maf from the HALtools utility (Hickey et al., 2013). The MAF-file was filtered with MafFilter (Dutheil et al., 2014) to extract the genomic fragment (GF) alignments. Only blocks where sequences from all four species occurred exactly once were kept. Additional filtering steps were carried out to match lengths and gaps. The GFs were concatenated (within scaffold borders) and fragmented into approximately 20,000 nt pieces. In total there were 2861 GFs of a length of approximately 20,000 nucleotides each and a total length of 60 MB, which corresponds to c. 25% of the total genome size. In order to assess effects of different GF lengths, shorter or longer GFs were tested, which gave the same results.

Sequences for the additional 2–5 short read-sequenced individuals from each subspecies were added to the GFs as follows. The four taxa alignments were split to generate a reference sequence set for each taxon. Illumina reads from nine additional genotypes were processed with BBDuk[1] to trim and filter reads. The resulting reads were mapped to their respective subspecies reference set with BBMap (see text footnote 1). Generated bam-files were then used to generate a vcf file with freebayes (Garrison and Marth, 2012), followed by BCFtools consensus (Narasimhan et al., 2016) to produce a consensus sequence for each genotype. These consensus sequences were then added to the original four taxa alignments and realigned with FSA (Bradley et al., 2009).

RNA Extraction and Genome Annotation

RNA was extracted from *M. polymorpha polymorpha* and *M. polymorpha montivagans* at two different time points with different light conditions, lightness and darkness. The RNA extraction was performed using RNeasy Plant Mini kit (Qiagen).

[1]https://jgi.doe.gov/data-and-tools/bbtools/

The sequenced RNA raw data were assembled in two ways: a *de novo* assembly using Trinity v.2.4.0 (Grabherr et al., 2011) and genome-guided assembly using a combination of Hisat2 v.2.1.0 (Kim et al., 2015) and StringTie v.1.3.4 (Pertea et al., 2015). In the latter case, the raw reads were first trimmed using Trimmomatic v.0.36 (Bolger et al., 2014). SamTools v.1.8 (Li et al., 2009) and Gffread, belonging to the Cufflinks v.2.2.1 package (Trapnell et al., 2013) were used for intermediate file sorting and format conversion steps. The final file formats for the *de novo* and the genome-guided assemblies are FASTA and GFF3, respectively. The two libraries (sampled in light/darkness) per sample were assembled separately.

Genome annotation was done using Maker version 3.01.2-beta (Holt and Yandell, 2011) in two runs. The NBIS annotation toolkit[2] was used for some of processing steps. In the first run the options *est2genome* and *protein2genome* were set to one, to obtain a first set of genes used to train the *ab initio* tools. The transcriptome assemblies were entered as *est,* respectively, *est_gff*. Swissprot (downloaded 2018-10-31 from https://www.uniprot.org/downloads) and the published proteome of *M. polymorpha ruderalis* were given as protein support. Augustus v. 3.2.3 (Stanke and Morgenstern, 2005) were trained using a non-redundant and AEDfiltered ($\geq$ 0.3) set of proteins without isoforms from the first Maker run. GeneMark-ES were trained using gmes_petap.pl – ES –training with the protein2genome Maker output file as – evidence. In the second run of Maker, *est2genome* and *protein2genome* were set to 0, and the parameter files from the training of the two *ab initio* tools were entered. For both runs, "*always_complete*" was set to 1, and *alt_splice* and *run_evm* were set to 0. Species-specific repeat libraries identified (described below) were entered as *rmlib*. BUSCO v. 2.5 (Simão et al., 2015) with the "Eukaryota odb9" dataset were used to check the completeness of transcriptomes and annotations. For *M. paleacea* the transcriptomes of *M. polymorpha* subspecies were used as *est* in the first data set, giving a smaller set of conserved genes used to train the *ab initio* tools. In the second round they were given as *alt_est.* Only scaffolds larger than 10,000 were considered. The transfer of gene models from the references of each subspecies to the additional samples was also done using Maker but with the reference transcriptome used as *est* and with *est2genome, always_complete* and the hidden option *est_forward* set to 1.

Preparation of Data: Alignment of Coding Sequences

The clustering of orthologs was done using OrthoVenn (Wang et al., 2015) based on the predicted output file from Maker, one file per species, with default settings (*e*-value cutoff: 10^{-5}, inflation value: 1.5) and 9957 single-copy gene clusters were extracted for further phylogenetic analyses. The orthologous proteins were aligned using MUSCLE (Edgar, 2004) and the coding sequences (CDSs) were aligned based on the protein alignments using trimAL (Capella-Gutiérrez et al., 2009) with the backtranslate option to keep the information about codon positions. Only CDSs longer than 300 bases were kept. For the concatenation of CDSs the *concat* command of the SeqKit (Shen et al., 2016) tool was used.

Repeat Annotation

RepeatModeler (Smit and Hubley, 2008-2015) (version 1.0.8_RM4.0.7) was used for *de novo* repeat family identification. The output was used as repeat library for RepeatMasker version 4.0.7 (Smit et al., 2013-2015).

Phylogenetic Analysis: Mitochondria and Chloroplast DNA

The organellar DNA sequences were treated as single loci. Three different phylogenetic methods were used i.e., MrBayes, neighbor joining (NJ), and RaxML methods. MrBayes version 3.2.6 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; Ronquist et al., 2012) was used to reconstruct the Bayesian phylogenetic tree using the best fitting substitution model of sequence evolution, selected using Modelgenerator version 85.1 (Keane et al., 2006). For chloroplast DNA this was GTR + G (lset nst = 6 rates = gamma) and for mitochondrial DNA GTR + G + I (lset nst = 6 rates = invgamma). Bootstrapped neighbor joining trees were also reconstructed using the "nj" and "boot.plylo" functions in the R package ape v. 5.2 (Paradis and Schliep, 2019). In addition, phylogenetic trees were reconstructed using RAxML with the nucleotide substitution model "–m GTRGAMMA" and 100 bootstraps using fast bootstrap (-x).

Phylogenetic Analysis: Genomic Fragments (GFs) and Transcripts

Individual ML-trees were reconstructed for all GF and transcript alignments using the maximum likelihood method based RAxML version 8.2.4 (Stamatakis, 2014). The nucleotide substitution model chosen for all trees was "–m GTRGAMMA". The GTR model of nucleotide substitution was chosen for all trees as it is the most general model, performing well for most real-world sequence data (RAxML manual), and 100 bootstraps using fast bootstrap (−x). The same methods were used with the concatenated versions of GFs and transcripts where all alignments for each type of data were joined together to one large alignment.

ASTRAL version 5.6.2 (Mirarab et al., 2014; Mirarab and Warnow, 2015) was used to estimate a species tree from the multiple GF/transcript trees. It takes a set of unrooted RAxML trees as input and gives as output an unrooted species tree, which is the tree that agrees with the largest number of quartet trees induced from the input tree set. It can handle ILS and is often more accurate than the concatenation method, except when the level of ILS is low (Mirarab et al., 2014; Mirarab and Warnow, 2015). If the bootstrap replicates for each alignment is included, ASTRAL performs a multi-locus bootstrapping. 100 bootstrapped replicates were done.

As a complement, to evaluate consistency, we also applied a bayesian approach using MrBayes for constructing phylogenetic trees and BUCKy for analyzing the complete set of trees. For MrBayes the Perl script "mb.pl" from the TICR pipeline[3] was used

[2]https://github.com/NBISweden/GAAS/

[3]http://crsl4.github.io/PhyloNetworks.jl/latest/man/ticr_howtogetQuartetCFs/

with default settings. This output was used as input for BUCKy version 1.4.4 (Larget et al., 2010; Mirarab et al., 2014) to estimate the dominant history of sampled individuals and how much of the genome that supports each relationship based on Bayesian concordance analysis. These concordance factors are given with a credibility interval taking into account the uncertainty in gene tree estimates. We first used the default prior (1) and then tested a second prior (0.01), with the same results.

Identification of Synteny and Other Chromosome-Level Comparisons

Chromosemble from the Satsuma2 packages (Grabherr et al., 2010) was used to order and orient the scaffolds of the *MpmSA2 and MprBR5* assemblies to 8 pseudo-chromosomes according to synteny with the *M. polymorpha ruderalis* genome. Only scaffolds larger than 100,000 were included. Satsuma2 package comprising SatsumaSynteny2, BlockDisplaySatsuma and MicroSyntenyPlot, was used to identify synteny matches, collect this information into synteny blocks, and visualize synteny as dotplots.

Gene order conservation was calculated using MCScanX_h (Wang et al., 2012) on the orthologous transcript data set. A collinear pair was defined as one orthologous pair lying directly adjacent to another orthologous pair, in both genomes compared and two or more adjacent genes are needed to be called a collinear block (-s 2 -m 0). Due to the fragmented nature of the genomes of subsp. *montivagans* and subsp. *polymorpha* the values are underestimations. The R package PopGenome (Pfeifer et al., 2014) was used to calculate pairwise nucleotide diversity between subspecies (d_{xy}).

Introgression Analyses

As a means to distinguish between ILS and introgression, we applied three variants of the ABBA-BABA test – Pattersons D statistic (Durand et al., 2011), Martins f statistic (Martin et al., 2013) and Bd-fraction (Pfeifer and Kapan, 2017) – using the R package "PopGenome" (Pfeifer et al., 2014). Local ancestry inference was conducted using Loter (Dias-Alves et al., 2018), with default settings and three ancestral populations (subsp. *ruderalis*, subsp. *polymorpha* and subsp. *montivagans*). When analyzing MpmBU3 or MppBV1 those individuals were excluded from their respective ancestral populations.

RESULTS

Phylogenetic Inference

In total, the genomes of six individuals of *M. polymorpha* subsp. *ruderalis*, three individuals of *M. polymorpha* subsp. *polymorpha* and three individuals of *M. polymorpha* subsp. *montivagans* are included in this study together with the genome of one individual of *M. paleacea* subsp. *diptera* used as outgroup. Only autosomes were included. Phylogenetic reconstructions in general displayed a clear separation of the three subspecies. However, the branching order of the three taxa was less obvious. Analyses based on complete nuclear DNA (irrespective of data set and phylogenetic method used) placed subsp. *polymorpha* and subsp. *ruderalis* as sister species. Subsp. *montivagans* is placed as sister lineage to subsp. *polymorpha* + subsp. *ruderalis* (**Figures 3A,B**). Using complete chloroplast DNA, subsp. *ruderalis* diverged first with subsp. *montivagans* and subsp. *polymorpha* as sister taxa (**Figure 3C**). Even though the support for these branching orders varied depending on tree reconstruction method, the length of the defining branch was always short. For complete mitochondrial DNA, the branching order was the same as that of the nuclear data, but with varying support and extremely short branches (**Figure 3D**).

These phylogenetic patterns could be the results of the divergence of all the three subspecies within a short period of time. This could lead to unresolved gene trees, or more than one supported topology due to ILS. Alternatively, recent hybridization might have obscured a previously clear branching order. To differentiate between these scenarios, we calculated concordance factors (CFs) for the three possible branching orders of the three subspecies. If, as previously suggested, subsp. *ruderalis* arose through a recent hybridization event between the other two subspecies we would expect gene trees clustering subsp. *ruderalis* and *polymorpha* (topology 1), and those clustering *ruderalis* and *montivagans* (topology 2), but not those grouping *montivagans* with *polymorpha* (topology 3).

The most abundant topology was the one favored in the nuclear species trees, topology 1 comprising 32/43% of all gene/GF trees and a CF of 0.49, followed by topology 2 (18/23% of all trees and CF = 0.28) (**Figure 4**). Even though topology 3 was the least abundant one, it constituted a considerable fraction of all supported individual trees (13/18%) and a CF of 0.21. In these analyses, trees with a bootstrap support of less than 70 were considered to be non-significant (21% for GFs and 32% for transcripts). These data contradict the proposed recent hybrid origin of subsp. *ruderalis*. Rather, the high frequency of supported trees for all three possible topologies suggests a similar age of the three subspecies and frequent ILS, possibly accompanied by more ancient hybridization and introgression.

In ABBA-BABA tests, an excess of ABBA sites over BABA sites indicative of introgression is signaled by a significant positive deviation from zero (see e.g., Heliconius Genome Consortium, 2012). Assuming a phylogeny according to the obtained species trees for nuclear data, all calculated statistics were close to zero and non-significant (see **Figures 5A–C**). This result support the conclusion that ILS is prevalent, and that subsp. *ruderalis* is of similar age as the other two subspecies, and not a recent hybrid between the two.

One Chromosome Has Experienced a Distinct Phylogenetic History

The genome-wide pattern seen in the phylogenetic analysis is not representative for all chromosomes. For the trees based on data representing chromosome 2 in subsp. *ruderalis* the concordance factors (CFs) for the primary concordance tree are 0.906 and 0.930 for GFs and transcripts, respectively. This is in contrast to all other chromosomes where CFs are less than 0.5. In agreement with this pattern, GF and transcript alignments corresponding to subsp. *ruderalis* chromosome 2 showed a strikingly higher

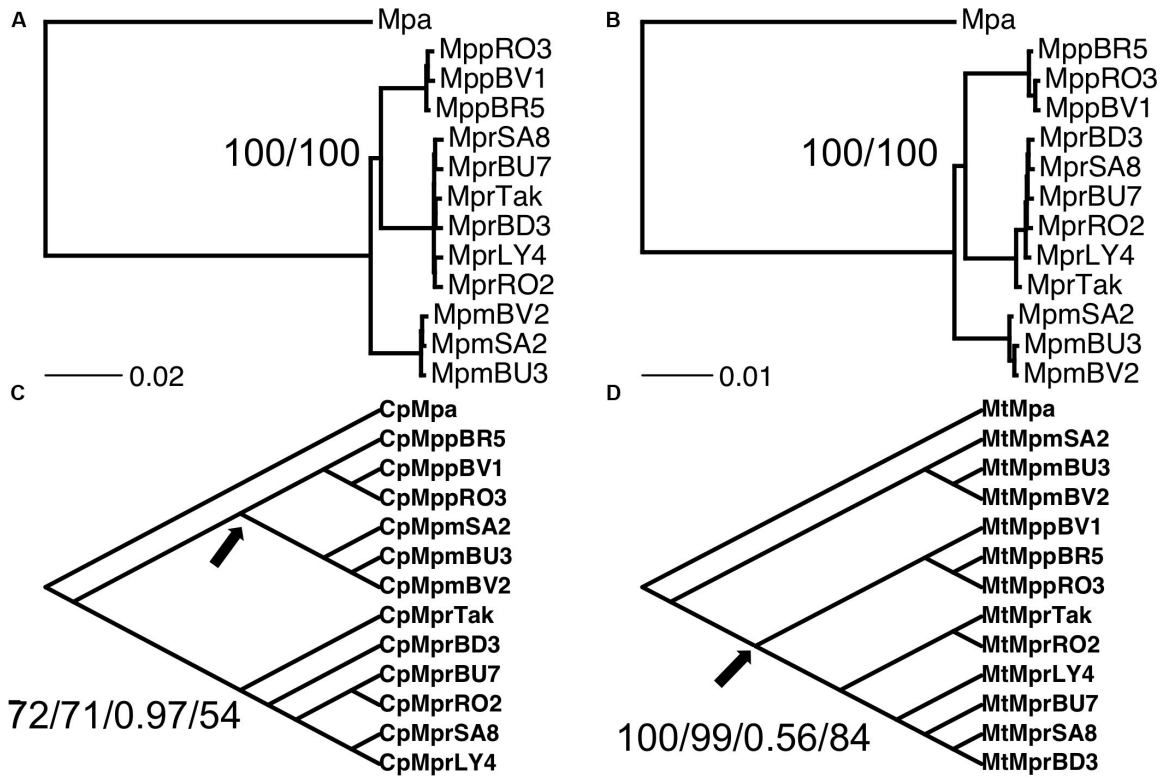


FIGURE 3 | Phylogenetic relationship between the three subspecies of *M. polymorpha.* **(A)** Phylogenetic tree obtained from a concatenated alignment of genomic fragments (GFs) with branch length corresponding to substitutions per site. The first node support value corresponds to the RAxML bootstrap value for the concatenated alignments. Included also, as a second support value, is the bootstrap value from the ASTRAL species tree analysis. **(B)** Phylogenetic tree obtained from a concatenated alignment of coding transcripts, with branch length corresponding to substitutions per site. Node support values corresponding to RAxML bootstrap value for concatenated alignments/bootstrap value from ASTRAL species tree analysis. **(C)** Phylogenetic tree obtained from the chloroplast genome, with RAxML bootstrap/NJ bootstrap/MrBayes support value/RAxML bootstrap values using only genes. **(D)** Phylogenetic tree obtained from the mitochondria genome, with RAxML bootstrap/NJ bootstrap/MrBayes support value/RAxML bootstrap values using only genes. For clarity, due to the very low number of substitutions, the trees in **(C,D)** are cladograms and the branches are not proportional to substitutions per site. The relationship within each subspecies clade might vary between the trees produced with the different methods. Node support values are always shown for the node where two of the subspecies cluster together, while support values were always high for the first two splits. Note also that the chloroplast and mitochondria genomes consistently congruent with the nuclear genome subspecies designation, no transfer of chloroplasts or mitochondria between supspecies is detected in this data set.

nucleotide divergence between subsp. *montivagans* and both the other subspecies (**Figures 6A,B**). Thus, all comparisons including subsp. *montivagans* and alignments comprising sequences from subsp. *ruderalis* chromosome 2 showed a divergence at least twice as high as any other comparison not including chromosome 2 and subsp. *montivagans*. The higher divergence is seen over a large part of the chromosome except at one end where it is at a comparable level to the rest of the genome (**Figure 6C**).

We might expect alignments corresponding to subsp. *ruderalis* chromosome 2 to have a large effect on phylogenetic reconstructions and the resulting species tree. To evaluate this, data were reanalyzed excluding alignments representing subsp. *ruderalis* chromosome 2. The analyses still support the same species tree with subsp. *polymorpha* together with subsp. *ruderalis* as sister to subsp. *montivagans*, but the branch lengths are now even shorter. We also calculated introgression-statistics (ABBA-BABA tests) for all three possible topologies. Excluding chromosome 2, all measures of introgression are non-significant and close to zero for all three possible topologies (**Figures 5D–F**). Thus, for analyses based on all chromosomes except number 2, there is no evidence of a deviation from a strict bifurcating evolutionary history. Rather, the analyses support an almost star like tree topology and frequent ILS.

As a means to better understand the distinct patterns observed for chromosome 2, we searched for other aspects of these sequences where they might deviate from the general pattern. The level of gene order conservation, as measured with MCscanX, comparison including both subsp. *montivagans* and chromosome 2, was significantly lower than other comparison (MPM vs MPP: χ^2 = 37.033; P = 1.162e-09, MPM vs MPR: χ^2 = 11.708; P = 6.224e-04) (**Figure 7A**). Even though assembly contiguity of the three subspecies differs, comparisons between chromosomes within subspecies are still valid.

Dot plots on chromosome-based alignments of genomic fragments also reveal more rearrangements for comparisons including chromosome 2 and subsp. *montivagans* (**Supplementary Figures 1**, **2**; also visualized in **Figure 7B**).

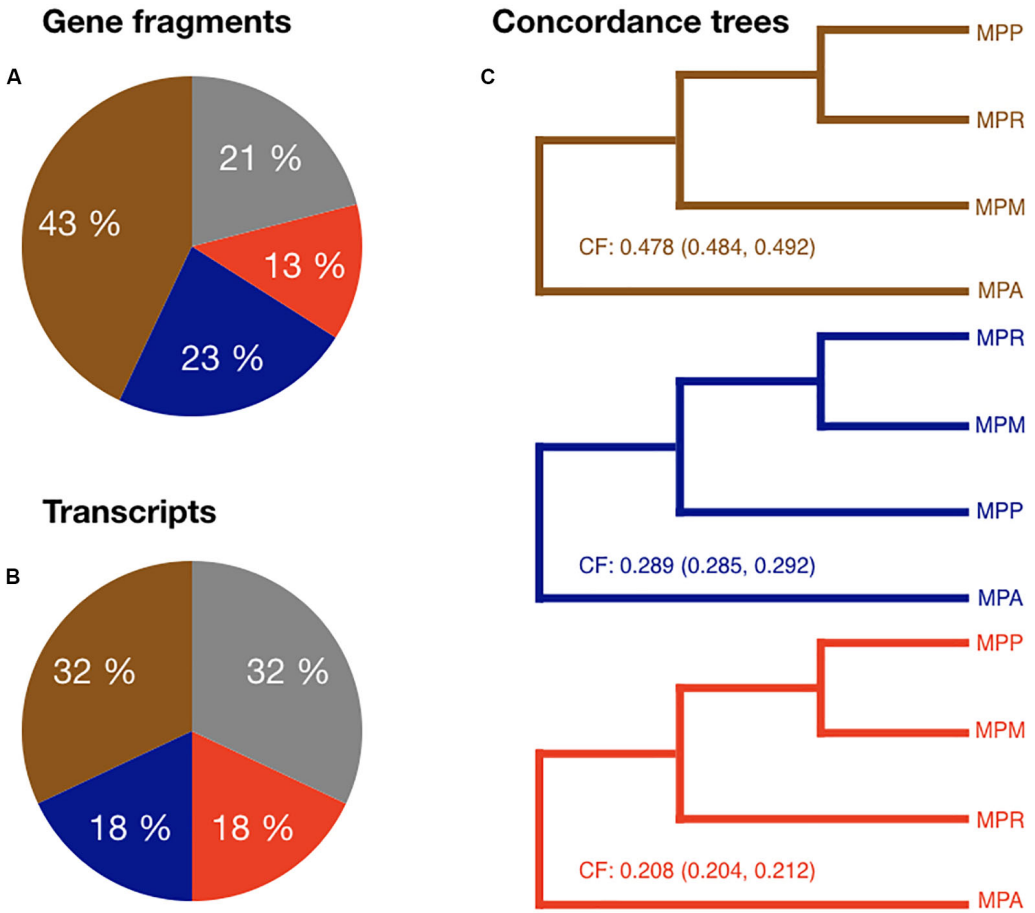


FIGURE 4 | (A) Pie chart showing the proportion of each topology among the gene fragment RAxML trees. Gray represent trees with a bootstrap value <70. **(B)** Pie chart showing the proportion of each topology among the transcript RAxML trees. Gray represent trees with a bootstrap value <70. **(C)** Brown tree: Primary concordance tree. Node values are the concordance factors (CFs) including confidence interval for CFs. Blue and red trees: Minor concordance trees. Node values are the concordance factors including confidence interval for CFs.

Furthermore, subsp. *montivagans* chromosome 2 sequences have a higher repeat content than those corresponding to other chromosomes. No such inflated content was observed for chromosome 2 sequences in the corresponding analysis of subsp. *ruderalis* and subsp. *polymorpha* (**Figures 7C–E**).

One Chromosome Shows Evidence of Recent Introgression

In an effort to detect more recent introgression events, all individuals from the three subspecies were investigated separately. At a genome-wide scale, no clear evidence of introgression was detected but analyzing 2861 GFs, 53 RAxML trees showed a phylogenetic topology where single individuals clustered with the wrong subspecies. For chromosome 1 approximately 8% of the input individual GF RAxML trees (23 out of 323 GFs) gave a topology indicating introgression between one individual and another subspecies. The corresponding values for the other chromosomes were lower and varied between 0 and 3.4%. Quartet CFs were calculated for the 54 possible quartet combinations (three individuals of subsp. *montivagans* x three individuals of subsp. *polymorpha* x six individuals of subsp. *ruderalis* × one individual of *Marchantia paleacea*). The CFs are expected to be independent on quartet combination in the absence of occasional introgression events in one of a few individuals. Analyzing chromosome 1 separately, two samples deviated from this expected pattern, MppBV1 and MpmBU3. In quartets including MppBV1 the CFs were higher than expected for the topology [(MPP, MPM), MPR] and while the CFs for quartets including MpmBU3 were higher than expected for [(MPR, MPM), MPP]. These data may indicate introgression between subsp. *polymorpha* and *montivagans* in the first case, and between subsp. *montivagans* and *ruderalis* in the second case.

To further explore possible introgression in individuals MppBV1 and MpmBU3, we performed local ancestry inference using the software Loter (Dias-Alves et al., 2018). Evidence of introgression was observed in both individuals on chromosome 1 (**Figure 8** and **Supplementary Figure 3**). For MpmBU3, one major area at the center of chromosome 1 and a few smaller regions were inferred as having subsp. *ruderalis* origin

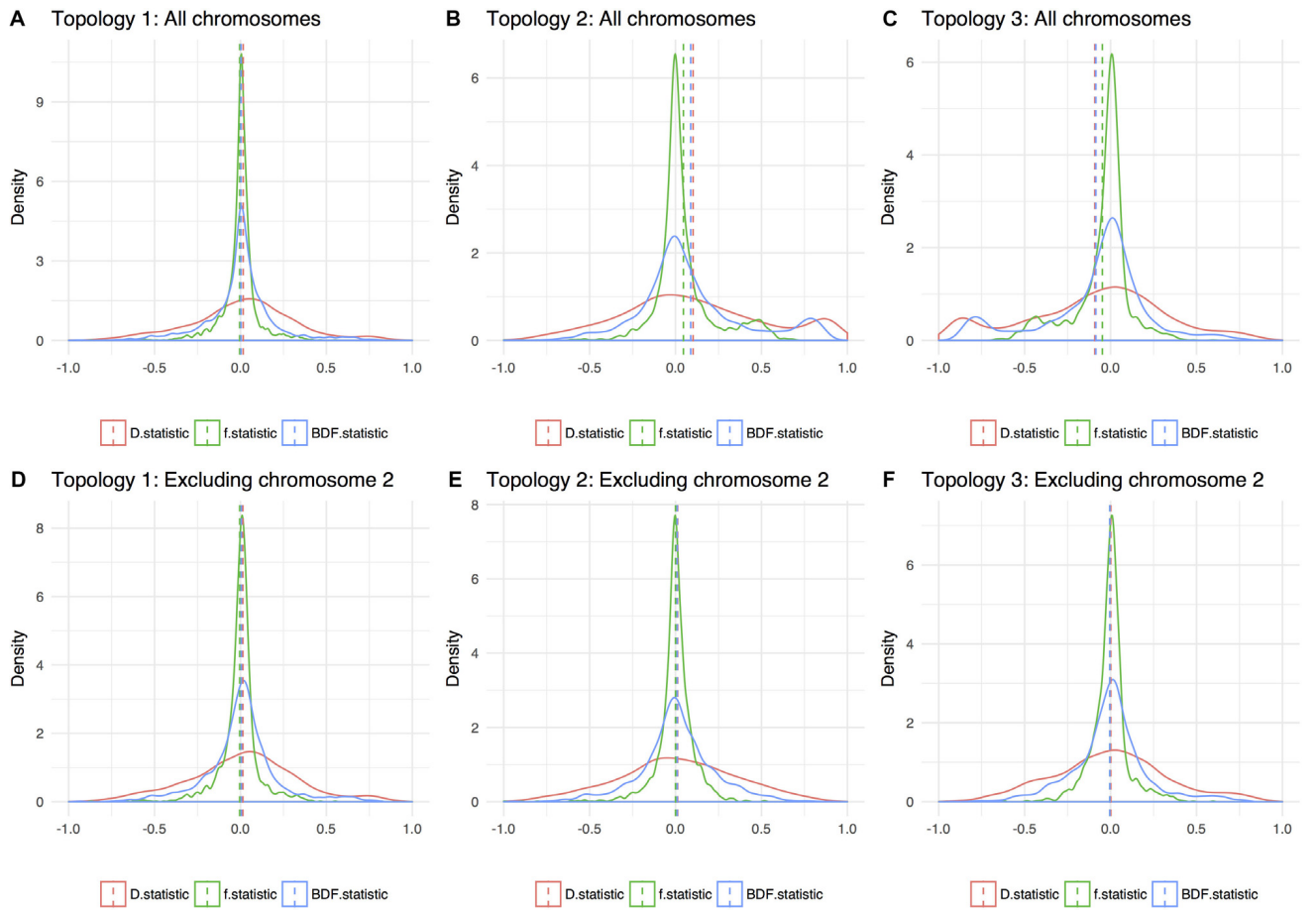


FIGURE 5 | Density plot of three variants of introgression statistics (ABBA-BABA tests): D statistics, f statistics, and BDF statistics. Dashed lines is the genome average for the different topologies and they are shown with **(A–C)** and without chromosome 2 **(D–F)**, respectively. Excluding chromosome 2, all three possible topologies give an averaged value of around 0, indicating a star tree. A value that deviates from 0 indicate hybridization or that the wrong phylogenetic tree topology is given.

(**Figure 8A**). The major area is separated into smaller regions separated by tracts of subsp. *montivagans* origin, suggesting that several generations with recombination have occurred.

An alternative interpretation to recent introgression for the inferred tracts of "foreign" origin could be ILS that for some reason is concentrated to a few areas on chromosome 1. An expectation from a recent introgression event in MpmBU3 is that introgressed tracts should be more similar to corresponding regions of subsp. *ruderalis* than to those of its own subsp. *montivagans*. This expectation does not hold for an ILS scenario. A plot of relative divergence along chromosome 1 (**Figure 8B**) clearly shows that in MpmBU3, areas with a concentration of tracts inferred as originating from subsp. *ruderalis,* show a lower divergence to pure *ruderalis,* than to the other two *montivagans* individuals (without evidence of introgression). Thus, our data support that introgression has occurred in BU3 some generations ago.

Similarly, for MppBV1 local ancestral inference identified one region on chromosome 1, with inferred ancestry from subsp. *montivagans* and one smaller region with inferred ancestry from subsp. *ruderalis* (**Supplementary Figure 3**). Again, the areas with a concentration of tracts inferred as subsp. *montivagans* show a lower divergence to pure *montivagans*, as compared to the pure *polymorpha* individuals. Thus, we find evidence for introgression resulting from hybridization a limited number of generations ago in a couple of individuals, but we find no evidence for any older fixed introgression events.

DISCUSSION

This paper reports the first large-scale phylogenomic analysis of the taxonomically controversial *M. polymorpha* complex in which three taxa of uncertain phylogenetic relationship have variably been treated at subspecies or species levels. The general phylogenetic pattern observed comprise three distinct taxa that diverged close in time. In line with this, the data also suggest frequent ILS resulting in a high degree of inconsistent gene trees. Still, species tree analyses recovered an overall topology where *M. polymorpha montivagans* diverged first

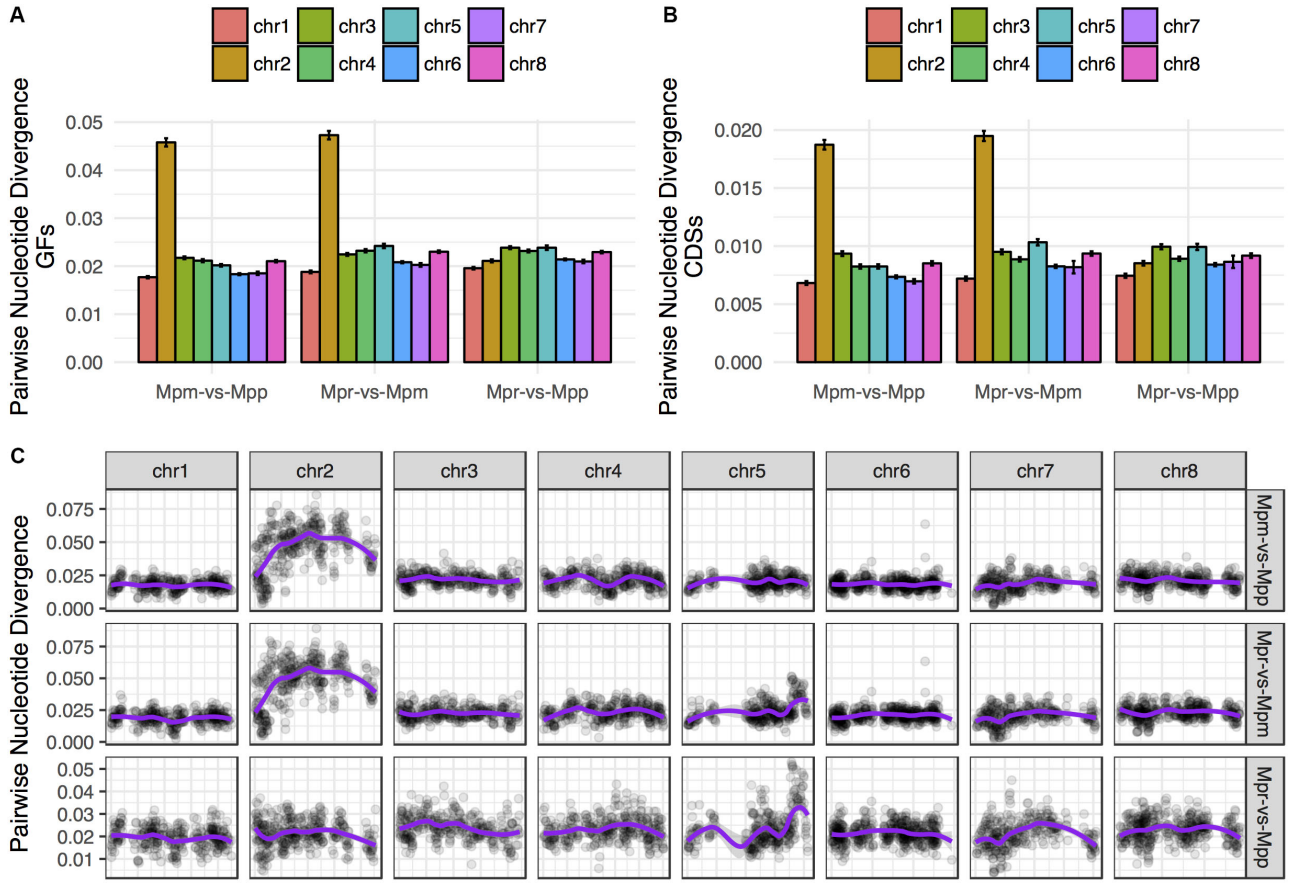


FIGURE 6 | The alignments including subsp. *montivagans* chromosome 2 show increased nucleotide divergence. **(A)** Pairwise nucleotide divergence averaged over all GFs, separated by chromosome. **(B)** Pairwise nucleotide divergence averaged over the orthologous transcripts data set, separated by chromosome. **(C)** Chromosome scans of pairwise divergence show that the increased divergence covers most of chromosome 2 and that there is a larger spread of values.

and *M. polymorpha ruderalis* and *M. polymorpha polymorpha* appeared as sister species.

Our data thus refute the hypothesis proposed by Burgeff (1943) and Schuster (1983, 1992) that subsp. *ruderalis* is a homoploid hybrid, formed by hybridization between subsp. *montivagans* and *polymorpha*. This hypothesis has been questioned by Boisselier-Dubayle et al. (1995) based on a limited data set, but our study is the first to test this hypothesis at the level of whole genomes.

In addition to the general phylogenetic patterns, our more detailed analyses revealed a more complex pattern with evidence suggesting hybridization and introgression between subspecies. One unexpected finding was that pseudo-chromosome 2 in subsp. *montivagans* showed several aberrant features. Most of this chromosome displayed more than twice the amount of genetic divergence to both subsp. *polymorpha* and *ruderalis,* as compared to other chromosomes. This increased divergence for chromosome 2 was also accompanied by higher degree of chromosomal rearrangements. Two scenarios that could explain this pattern include (1) hybridization with an unknown closely related species, and (2) extensive hybridization between *M. polymorpha* subspecies. For both scenarios additional factors must be included to explain that the effect is confined to a single chromosome. The first scenario requires genomic mixing with an unknown closely related species in the past or present. It also requires that one chromosome has been more resistant to elimination of foreign chromosomal material through repeated backcrossing to subsp. *montivagans*. The second scenario implies that hybridization between subspecies has been frequent in the past, and that a single chromosome has been more resistant to this hybridization. None of these scenarios seems very likely, but our analysis of potential hybridization in single individuals might favor the second scenario over the first. This is comparable with whole-genome studies of malaria parasite vectors belonging to the *Anopheles gambiae* complex (Fontaine et al., 2015). The mosquito species belonging to this complex show pervasive autosomal introgression, so that only a small part of the genome, mainly on the X-chromosome, has not crossed the species boundaries. The branching order determined from the X-chromosome was used to construct the true phylogeny, and then this topology was used to trace back the major introgression events. The authors concluded that their proposed historical branching order was represented by only 1.9% of 50-kb windows across the entire genome, but when this topology was

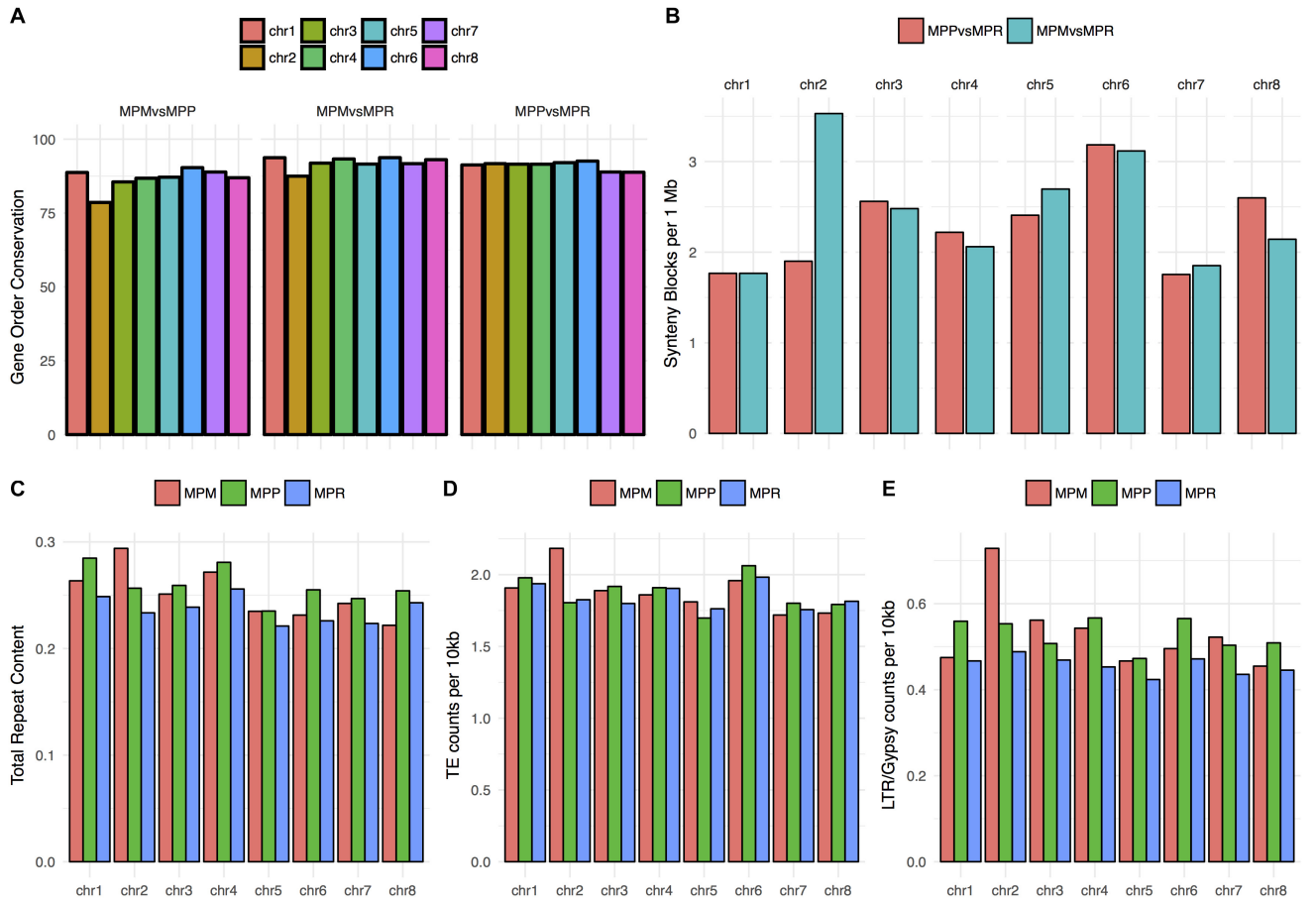


FIGURE 7 | (A) Percentage of gene order conservation in pairwise comparisons, using subsp. *ruderalis* chromosome coordinates as reference. The gene order conservation is significantly lower for chromosome 2 in the two comparisons including subsp. *montivagans* ($\chi 2 = 37.033$; $P = 1.162e\text{-}09$ resp. $\chi 2 = 11.708$; $P = 0.0006224$). **(B)** Number of synteny blocks per 1 Mbases for the comparisons between subsp.*ruderalis* chromosomes and subsp. *polymorpha* pseudochromosomes (red) and between subsp. *ruderalis* chromosomes and subsp. *montivagans* pseudochromosomes (cyan). A higher number of synteny blocks correlates with a higher number of synteny breaks and more rearrangements. The corresponding synteny dot plots are shown in **Supplementary Figures 1**, **2**. **(C)** Total repeat content per chromosome, represented as total repeat length per chromosome/pseudochromosome length. **(D)** Counts of transposable elements (Simple repeats, low complexity repeats and unknown are excluded) per 10,000 bases. **(E)** Counts per 10,000 bases of a specific transposable element, LTR/Gypsy, which is overrepresented in subsp. *montivagans* pseudochromosome 2.

recovered, the divergence times were consistently more distant, relative to the alternative topologies. With a similar scenario, the divergence of the aberrant pseudo-chromosome 2 in subsp. *montivagans* may represent the true phylogenetic separation from the other two subspecies.

For two individuals in our limited sample, MpmBU3 and MppBV1, we saw evidence of introgression with subsp. *ruderalis*, respectively, subsp. *montivagans* in restricted parts of their genomes. In both cases, we registered that the parental species pairs occurred in sympatry at the collection sites. For these pairs, Burgeff (1943) recorded low spore germination rate (9%) when crossing male subsp. *montivagans* and female subsp. *ruderalis* but considerably higher spore germination rate (50–70%) in the cross between female subsp. *montivagans* and male subsp. *polymorpha*. The reciprocal crosses did not render any viable spores. Burgeff also demonstrated that backcrossing with males of subsp. *polymorpha* was possible with progeny from the latter, more successful cross, which gives some experimental support for our suggestion that several generations with recombination may have occurred with regard to MpmBU3 and MppBV1.

In our cases, we registered that both the parental species pairs occurred in sympatry, and from our total population sampling it seems that sympatric populations are more common than generally recognized in the literature. For example, Damsholt (2002) states that the taxa rarely meet but mentions two exceptional sites in Scandinavia where subsp. *montivagans* and subsp. *ruderalis* (Sädvajaure, Pite Lappmark, Sweden), respectively, subsp. *montivagans* and subsp. *polymorpha* (hot springs by Landmannalaugar, Iceland) occur together. Only in the latter place, some intermediate plants possibly suggested introgression in that unusual habitat. If we were to generalize these observations, hybridization between subspecies may well have been frequent, but in most cases foreign DNA fragments are rapidly shortened through backcrossing and recombination. This

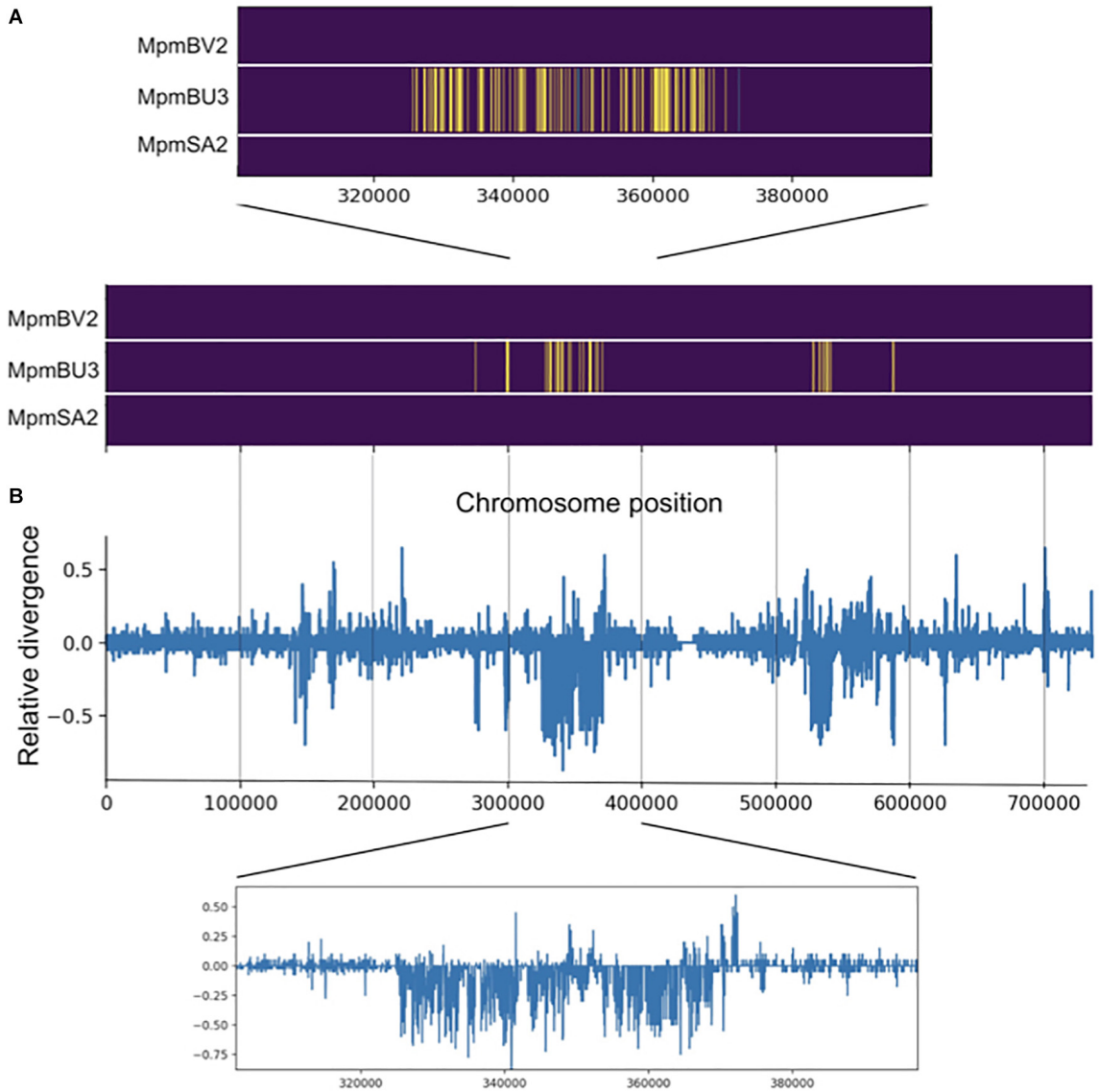


FIGURE 8 | Evidence for introgression is observed for one subsp. *montivagans* individual. **(A)** Local ancestry inference for three subsp. *montivagans* individuals derived from Loter software. Image plot of pseudo-chromosome 1 illustrating ancestral origin, subsp. *montivagans* (purple), *ruderalis* (yellow) and *polymorpha* (green). The *x*-axis shows SNP number along chromosome 1. **(B)** Plot of difference in genetic divergence between on the one hand MpmBU3 versus subsp. *ruderalis* individuals, and on the other hand MpmBU3 versus other subsp. *montivagans* individuals ($d_{MpmBU3\_ruderalis} - d_{MpmBU3\_montivagans}$), i.e., positive bars show genes that are diverged between MpmBU3 versus subsp. *ruderalis* (but not within subsp. *montivagans*), whereas negative bars reveal genes that differ to other accessions of subsp. *montivagans* (but not with subsp. *ruderalis*). Insets show magnifications of the data along the pseudo-chromosome.

could lead to pattern similar to the one obtained through frequent ILS and could thus fit with the observation of an almost star-like phylogeny of the three subspecies. Subspecies *montivagans* has a montane distribution, whereas both *polymorpha* and *ruderalis* occur in lowland areas, the former typically in wetlands and the latter in drier and more ruderal contexts including places subject to forest fire. It is therefore not unlikely that climatic oscillations during the Pleistocene has periodically brought the subspecies in closer contact than they normally are today and allowed for more frequent hybridization, so that the pattern we see may be a product of historic sympatry.

The haploid-dominant lifecycle might explain a spatially limited genetic exchange after hybridization. Interspecific hybridization in bryophytes results in diploid hybrid sporophytes formed after fertilization. The hybrid sporophyte is physically connected to the female plant and short-lived. The two parental genomes can recombine during meiosis, which takes place in the sporophyte, to form spores. The true hybrid is the sporophyte and the spores produced in the sporophyte are recombinants that have a mix of genes from both parents and can be referred to as hybrid segregates comparable to the F_2 generation of angiosperms with hybrid origin. Spores are

formed in tetrads and each tetrad is the result of a single meiotic event. The sporophytes produce not less than 300,000 spores in *M. polymorpha* (O'Hanlon, 1926) and thus 75,000 meiotic events are likely to take place during production of recombinant spores. If extensive amounts of genomic admixture in recombinants lead to incompatibilities, and therefore spore abortion or hampered spore germination, we can expect surviving individuals to show a strongly asymmetric representation of the parental genomes. A similar explanation has been suggested for high mortality and limited admixture observed in hybrid F_2 progeny in peat mosses (Natcheva and Cronberg, 2007). In sympatric populations of the pleurocarpous mosses *Homalothecium lutescens and sericeum*, mildly admixed individuals were relatively common and strongly admixed individuals were sometimes seen in both gametophyte and sporophyte generations (Sawangproh, 2019, Sawangproh et al., in press). This suggests that the admixed alleles were transmitted between generations and that sympatric populations behave as true hybrid zones.

Pseudo-chromosome 2 in subsp. *montivagans* did not only show higher nucleotide divergence and more chromosomal rearrangements. Is also showed a higher proportion of transposons. At present it is premature to speculate about a causal relationship between these observations, but it is possible that more rearrangements discriminating subsp. *montivagans* pseudo-chromosome 2 from its homologs attenuate chromosome pairing of this chromosome in hybrids. If so, such attenuated pairing and thus reduced introgression could explain why subsp. *montivagans* pseudo-chromosome 2 is still more diverged.

Our sampling is restricted to a limited part of the whole distribution range, and differentiation patterns could possibly deviate in other regions. However, it is worth to notice that the nuclear, mitochondrial and plastid genomes of the geographically remote accession Tak (from Japan) does not differ substantially from the Scandinavian accessions of subsp. *ruderalis*. The separation of the three taxa is also substantiated by the observation that each is associated with its own chloroplast and mitochondrial haplotype group. The present taxonomic treatment of *M. polymorpha ruderalis*, *M. polymorpha polymorpha* and *M. polymorpha montivagans* is intraspecific, as subspecies. This treatment has been questioned by Schuster (1992) and Damsholt (2002) arguing that the taxa are morphologically differentiated, having largely non-overlapping distribution areas. Kijak et al. (2018) came to the same conclusion based on data from two cpDNA regions. Whether to recognize the taxa at species or subspecies level depend to some extent on the species concept chosen and this will be discussed in another context. From a more practical point of view our study also raises a *memento* concerning phylogenetic inferences based on small sets of sequences. ILS and hybridization as well as differences in divergence among chromosomes, as evident from our study, may strongly affect the outcome of such analyses, especially for closely related species. Maybe, more so for bryophytes than for many other organisms if multiple hybridization events with transfer of small parts of the genome at a time is a widespread phenomenon (cf. Natcheva and Cronberg, 2007; Sawangproh et al., 2020).

CONCLUSION

In conclusion, our data shows that the *Marchantia polymorpha*-complex comprises three recently diverged lineages, with relatively frequent hybridization and introgression, at least in a longer time perspective. Only limited parts of the genomes appear to be transferred between lineages at each occasion and one chromosome is less porous to gene transfer than the others. Alleles transferred between the genomes could still lead to improved adaptation, as they are immediately exposed to selection in the dominant haploid phase of the bryophyte life cycle.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found at NCBI, https://www.ncbi.nlm.nih.gov under Bioproject PRJNA576577, and the raw data for *M. palacea* under Bioproject PRJNA362997.

AUTHOR CONTRIBUTIONS

UL and NC initiated the project. NC provided the plant material. UL and A-ML were responsible for DNA/RNA extractions. A-ML carried out the bioinformatics and statistics together with UL and WS, with input from NC. A-ML wrote the manuscript, in close collaboration with the other authors. PS provided chromosome assembly data and commented on the manuscript.

FUNDING

This work was supported by the Swedish Research Council (UL 2011-5609 and 2014-522). We thank the GENECO graduate research school at Lund University for a grant to WS to support travel costs and accommodation during a stay at Uppsala University. Support was also provided by Elly Olsson's fund and Ove Almborn's fund to WS and NC. Assembly of scaffolds into chromosome level was supported by the URPP *Evolution in Action*, grants of the Swiss National Science Foundation (PSZ 160004, 131726), the EU's Horizon 2020 Research and Innovation Program (PSZ PlantHUB-No. 722338), and the Georges and Antoine *Claraz Foundation.*

ACKNOWLEDGMENTS

We thank the National Genomics Infrastructure (NGI)/Uppsala Genome Center, SciLifeLab and Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for assistance in massive parallel sequencing and computational infrastructure and the National Bioinformatics Infrastructure Sweden for bioinformatics support (NBIS). Rayna Natcheva at Bulgarian Academy of Sciences and Johan Rydlöv are acknowledged for collection of plant material in Bulgaria,

respectively, northern Sweden. Dr. Guru Radhakrishnan and Dr. Pierre-Marc Delaux provided early access to genome data for *M. paleacea*. We also thank two reviewers for constructive input. This work was presented in part at IAB/iMoss conference Bryology 2019 in Madrid.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00829/full#supplementary-material

REFERENCES

Anderson, E., and Hubricht, L. (1938). Hybridization in Tradescantia. III. The evidence for introgressive hybridization. *Am. J. Bot.* 25, 396–402. doi: 10.1002/j.1537-2197.1938.tb09237.x

Atherton, I., Bosanquet, S., and Lawley, M. (2010). *Mosses and Liverworts of Britain and Ireland: A Field Guide*. London: British Bryological Society.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its appications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bischler-Causse, H., and Boisselier-Dubayle, M. C. (1991). Lectotypification of *Marchantia polymorpha* L. *J. Bryol.* 16, 361–365.

Boisselier-Dubayle, M. C., and Bischler-Causse, H. (1989). Electrophoretic studies in *Marchantia polymorpha* L. *J. Hattori Bot. Lab.* 67, 297–311.

Boisselier-Dubayle, M. C., Jubier, M. F., Lejeune, B., and Bischler, H. (1995). Genetic variability in the three subspecies of *Marchantia polymorpha* (Hepaticae): isozymes. RFLP and RAPD markers. *Taxon* 44, 363–376. doi: 10.2307/1223406

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017). Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell* 171:287–304.e15. doi: 10.1016/j.cell.2017.09.030

Bradley, R. K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., et al. (2009). Fast statistical alignment. *PLoS Comput. Biol.* 5:e1000392. doi: 10.1371/journal.pcbi.1000392

Burgeff, H. (1943). *Genetische Studien an Marchantia*. Jena: Gustav Fischer.

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Damsholt, K. (2002). *Illustrated Flora of Nordic Liverworts and Hornworts*. Lund: Nordic Bryological Society.

Dias-Alves, T., Mairal, J., and Blum, M. G. B. (2018). Loter: a software package to infer local ancestry for a wide range of species. *Mol. Biol. Evol.* 35, 2318–2326. doi: 10.1093/molbev/msy126

Diop, S. I., Subotic, O., Giraldo-Fonseca, A., Waller, M., Kirbis, A., Neubauer, A., et al. (2019). A pseudomolecule-scale genome assembly of the liverwort *Marchantia Polymorpha*. *Plant J.* 101, 1378–1396. doi: 10.1111/tpj.14602

Dowling, T. E., and Secor, C. L. (1997). The role of hybridization and introgression in the diversification of animals. *Ann. Rev. Ecol. Syst.* 28, 593–619. doi: 10.1146/annurev.ecolsys.28.1.593

Durand, E. Y., Patterson, N. J., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252. doi: 10.1093/molbev/msr048

Dutheil, J. Y., Gaillard, S., and Stukenbrock, E. H. (2014). MafFilter: a highly flexible and extensible multiple genome alignment files processor. *BMC Genomics* 15:53. doi: 10.1186/1471-2164-15-53

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., et al. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347, 1258524–1258524. doi: 10.1126/science.1258524

Galtier, N., and Daubin, V. (2008). Dealing with incongruence in phylogenomic. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 4023–4029. doi: 10.1098/rstb.2008.0144

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv [Preprint]*. Available online at: arXiv:1207.3907v2 (accessed 1 August 2019).

Gogarten, J. P., and Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687. doi: 10.1038/nrmicro1204

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Palma, F. D., et al. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26, 1145–1151. doi: 10.1093/bioinformatics/btq102

Grant, P. R., Grant, B. R., and Petren, K. (2005). Hybridization in the recent past. *Am. Nat.* 166, 56–67. doi: 10.1086/430331

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Harrison, R. G., and Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *J. Heredity* 105, 795–809. doi: 10.1093/jhered/esu033

Healey, A., Furtado, A., Cooper, T., and Henry, R. J. (2014). Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* 10:21. doi: 10.1186/1746-4811-10-21

Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487, 94–98. doi: 10.1038/nature11041

Hickey, G., Paten, B., Earl, D., Zerbino, D., and Haussler, D. (2013). HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* 29, 1341–1342. doi: 10.1093/bioinformatics/btt128

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. doi: 10.1186/1471-2105-12-491

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Kahle, D., and Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *R J.* 5, 144–161.

Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J., and Mclnerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* 6:29. doi: 10.1186/1471-2148-6-29

Kijak, H., Łodyga, W., and Odrzykoski, I. J. (2018). Sequence diversity of two chloroplast genes: rps4 and tRNAGly (UCC), in the liverwort *Marchantia polymorpha*, an emerging plant model system. *Acta Soc. Bot. Pol.* 87:3573.

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Larget, B. R., Kotha, S. K., Dewey, C. N., and Ané, C. (2010). BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26, 2910–2911. doi: 10.1093/bioinformatics/btq539

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. doi: 10.1186/2047-217X-1-18

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi: 10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature* 446, 279–283.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., et al. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res.* 23, 1817–1828.

Martin, S. H., and Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* 47, 69–74. doi: 10.1016/j.gde.2017.08.007

Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, i541–i548. doi: 10.1093/bioinformatics/btu462

Mirarab, S., and Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31, i44–i52. doi: 10.1093/bioinformatics/btv234

Mossel, E., and Roch, S. (2010). Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 7, 166–171. doi: 10.1109/tcbb.2008.66

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32, 1749–1751. doi: 10.1093/bioinformatics/btw044

Natcheva, R., and Cronberg, N. (2004). What do we know about hybridization among bryophytes in nature? *Can. J. Bot.* 82, 1687–1704. doi: 10.1139/b04-139

Natcheva, R., and Cronberg, N. (2007). Maternal transmission of cytoplasmic DNA in interspecific hybrids of peat mosses. *Sphagnum (Bryophyta). J. Evol. Biol.* 20, 1613–1616. doi: 10.1111/j.1420-9101.2007.01341.x

Nees, C. G. (1838). *Naturgeschichte der Europäischen Lebermoose mit besonderer Beziehung auf Schlesien und die Oertlichebkeiten des Riesengebirgs*, Vol. 4. Breslau: Grass, Barth & Co.

Nishiyama, T., Wolf, P. G., Kugita, M., Sinclair, R. B., Sugita, M., Sugiura, C., et al. (2004). Chloroplast phylogeny indicates that bryophytes are monophyletic. *Mol. Biol. Evol.* 21, 1813–1819. doi: 10.1093/molbev/msh203

O'Hanlon, M. E. (1926). Germination of spores and early stages in development of gametophyte of *Marchantia polymorpha*. *Bot. Gazette* 82, 215–222. doi: 10.1086/333650

Paradis, E., and Schliep, K. (2019). Ape5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071

Paten, B., Diekhans, M., Earl, D., John, J., Ma, J., Suh, B., et al. (2011a). Cactus graphs for genome comparisons. *J. Comput. Biol.* 18, 469–481. doi: 10.1089/cmb.2010.0252

Paten, B., Earl, D., Nguyen, N., Diekhans, M., Zerbino, D., and Haussler, D. (2011b). Cactus: algorithms for genome multiple sequence alignment. *Genome Res.* 21, 1512–1528. doi: 10.1101/gr.123356.111

Paton, J. A. (1999). *The Liverwort Flora of the British Isles*. Colchester: Harley Books.

Payseur, B. A., and Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Mol. Ecol.* 25, 2337–2360. doi: 10.1111/mec.13557

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Pfeifer, B., and Kapan, D. D. (2017). Estimates of introgression as a function of pairwise distances. *bioRxiv [Preprint]* doi: 10.1101/154377

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., and Lercher, M. J. (2014). PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31, 1929–1936. doi: 10.1093/molbev/msu136

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15511–15516. doi: 10.1073/pnas.0603335103

Radhakrishnan, G. (2017). *Tracing the Evolution of the Arbuscular Mycorrhizal Symbiosis in the Plant Lineage*. Doctoral thesis, University of East Anglia, Norwich.

Rieseberg, L. H., Archer, M. A., and Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity* 83, 363–372. doi: 10.1038/sj.hdy.6886170

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211–1216. doi: 10.1126/science.1086949

Roberts, R. J., Carneiro, M. O., and Schatz, M. (2013). The advantages of SMRT sequening. *Genome Biol.* 14:405.

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Sawangproh, W. (2019). *Gene transfer by interspecific hybridization in bryophytes*. Lund: Media-Tryck, Lund University, Sweden.

Sawangproh, W., Hedenäs, L., Lang, A. S., Hansson, B., and Cronberg, N. (2020). Gene transfer across species boundaries in bryophytes: evidence from major life cycle stages in Homalothecium lutescens and *H. sericeum*. *Ann. Bot.* 125, 565–579. doi: 10.1093/aob/mcz209

Schuster, R. M. (1983). "Phytogeography of the Bryophyta," in *New Manual of Bryology 1*, ed. R. M. Schuster (Nichinan: Hattori Botanical Laboratory), 463–626.

Schuster, R. M. (1992). *The Hepaticae and Anthocerotae of North America*. New York, NY: Columbia University Press.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19, 198–207. doi: 10.1016/j.tree.2004.01.003

Shaw, A. J. (1994). Systematics of Mielichhoferia (Bryaceae: Musci). III. Hybridization between *M. elongata* and *M. mielichhoferiana*. *Am. J. Bot.* 81, 782–790. doi: 10.1002/j.1537-2197.1994.tb15515.x

Shaw, A. J. (1998). "Genetic analysis of a hybrid zone in Mielichhoferia (Musci)," in *Bryology for the Twenty-First*, eds J. W. Bates, N. W. Ashton, and J. G. Duckett (Leeds: Century), 161–174. doi: 10.1201/9781315138626-12

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* 11:e0163962. doi: 10.1371/journal.pone.0163962

Shimamura, M. (2016). *Marchantia polymorpha*: taxonomy, phylogeny and morphology of a model system. *Plant Cell Physiol.* 57, 230–256. doi: 10.1093/pcp/pcv192

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Smit, A. F. A., and Hubley, R. (2008-2015). *RepeatModeler Open-1.0*. Availale online at: http://www.repeatmasker.org

Smit, A. F. A., Hubley, R., and Grenn, P. (2013-2015). *RepeatMasker Open-4.0*. Availale online at: http://www.repeatmasker.org

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, W465–W467.

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53. doi: 10.1038/nbt.2450

Villarreal, A. J. C., Crandall-Stotler, B. J., Hart, M. L., Long, D. G., and Forrest, L. L. (2016). Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. *New Phytol.* 209, 1734–1746. doi: 10.1111/nph.13716

Wang, Y., Coleman-Derr, D., Chen, G., and Gu, Y. Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 43, W78–W84. doi: 10.1093/nar/gkv487

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Ward, B. J., and van Oosterhout, C. (2016). HYBRIDCHECK?: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Mol. Ecol. Resour.* 16, 534–539. doi: 10.1111/1755-0998.12469

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 07 July 2020
doi: 10.3389/fpls.2020.00813

Single Nucleotide Polymorphism Charting of *P. patens* Reveals Accumulation of Somatic Mutations During *in vitro* Culture on the Scale of Natural Variation by Selfing

*Fabian B. Haas[1], Noe Fernandez-Pozo[1], Rabea Meyberg[1], Pierre-François Perroud[1], Marco Göttig[1], Nora Stingl[1], Denis Saint-Marcoux[2,3], Jane A. Langdale[2] and Stefan A. Rensing[1,4,5]**

[1] Plant Cell Biology, Department of Biology, University of Marburg, Marburg, Germany, [2] Department of Plant Sciences, University of Oxford, Oxford, United Kingdom, [3] Université de Lyon, UJM-Saint-Etienne, CNRS, Laboratoire BVpam - FRE 3727, Saint-Étienne, France, [4] BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany, [5] SYNMIKRO Center for Synthetic Microbiology, University of Marburg, Marburg, Germany

OPEN ACCESS

Edited by:
Michael R. McKain,
The University of Alabama,
United States

Reviewed by:
Matthew Johnson,
Texas Tech University, United States
Sara V. Good,
The University of Winnipeg, Canada

***Correspondence:**
Stefan A. Rensing
stefan.rensing@
biologie.uni-marburg.de

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 19 February 2020
Accepted: 20 May 2020
Published: 07 July 2020

Citation:
Haas FB, Fernandez-Pozo N, Meyberg R, Perroud P-F, Göttig M, Stingl N, Saint-Marcoux D, Langdale JA and Rensing SA (2020) Single Nucleotide Polymorphism Charting of P. patens Reveals Accumulation of Somatic Mutations During in vitro Culture on the Scale of Natural Variation by Selfing. Front. Plant Sci. 11:813. doi: 10.3389/fpls.2020.00813

Introduction: *Physcomitrium patens* (Hedw.) Mitten (previously known as *Physcomitrella patens*) was collected by H.L.K. Whitehouse in Gransden Wood (Huntingdonshire, United Kingdom) in 1962 and distributed across the globe starting in 1974. Hence, the Gransden accession has been cultured *in vitro* in laboratories for half a century. Today, there are more than 13 different pedigrees derived from the original accession. Additionally, accessions from other sites worldwide were collected during the last decades.

Methods and Results: In this study, 250 high throughput RNA sequencing (RNA-seq) samples and 25 gDNA samples were used to detect single nucleotide polymorphisms (SNPs). Analyses were performed using five different *P. patens* accessions and 13 different Gransden pedigrees. SNPs were overlaid with metadata and known phenotypic variations. Unique SNPs defining Gransden pedigrees and accessions were identified and experimentally confirmed. They can be successfully employed for PCR-based identification.

Conclusion: We show independent mutations in different Gransden laboratory pedigrees, demonstrating that somatic mutations occur and accumulate during *in vitro* culture. The frequency of such mutations is similar to those observed in naturally occurring populations. We present evidence that vegetative propagation leads to accumulation of deleterious mutations, and that sexual reproduction purges those. Unique SNP sets for five different *P. patens* accessions were isolated and can be used to determine individual accessions as well as Gransden pedigrees. Based on that, laboratory methods to easily determine *P. patens* accessions and Gransden pedigrees are presented.

Keywords: SNP, RNA-seq, *Physcomitrella patens*, *Physcomitrium*, ecotype, Gransden, Reute, RFLP

INTRODUCTION

Single Nucleotide Polymorphisms (SNPs) represent a major source of natural variation within any given species. In the plant kingdom, they are studied both in ecological and evolutionary context in order to understand population structure (Leaché and Oaks, 2017). They are also employed to study the genetic basis of variable natural traits such as resistance to flooding (Vashisht et al., 2011), or for the identification of genetic diversity in cultivars and admixed wild types through association mapping (Niu et al., 2019). SNP analysis is now successfully integrated in plant breeding for example in palm tree selection (Xia et al., 2019). For the moss model *Physcomitrium patens* (Hedw.) Mitten (previously known as *Physcomitrella patens*) (Beike et al., 2014; Medina et al., 2019; Rensing et al., 2020) whole genome SNP sets between the reference genome accession, Gransden (Gd) (Rensing et al., 2008; Lang et al., 2018), and the accessions Villersexel (Vx) (Kamisugi et al., 2008), Reute (Re) (Hiss et al., 2017) and Kaskaskia (Ka) (Perroud et al., 2011) have been reported (Hiss et al., 2017). Specifically, the genetic difference between Gd and Vx has been used to generate the first sequence-anchored genetic linkage map (Kamisugi et al., 2008) and recently the *P. patens* chromosome level genome assembly (Lang et al., 2018). Analysis of SNP segregation is a powerful tool that can be employed to analyze intra and inter accession fertility (Perroud et al., 2011, 2019; Meyberg et al., 2020), gene specific segregation patterns, and loci affected in segregants with specific traits. For example, the analysis of Gd and Vx segregants has been used to identify the *ANR* locus affected in mutants impaired in ABA hormone signaling (Stevenson et al., 2016), as well as loci involved in three-dimensional morphogenesis [*nog*1, (Moody et al., 2018)] and a novel microtubule depolymerizing-end-tracking protein [*CLoG*1, (Ding et al., 2018)]. Most recently, SNPs between Gd and Re were associated with the loss of fertility in the Gd background (Meyberg et al., 2020). However, there is no comparative study on a broad set of accessions, or within the different *P. patens* Gransden laboratory strains (Gd pedigrees).

Model organisms cultivated in the laboratory are usually considered to be genetically uniform due to their common origin. The original *P. patens* Gransden plant was collected by H.L.K. Whitehouse in Gransden Wood (Huntingdonshire, United Kingdom) in 1962. Engel cultured Whitehouse's sample (Engel, 1968) and derived the ancestor of all current *P. patens* Gransden strains from a single spore. In 1974 progeny of *P. patens* Gransden started to be distributed across the globe (Ashton and Cove, 1977; Cove, 2005). Since then, *P. patens* became an important model organism *inter alia* to study cell biology, evolutionary developmental biology and the water to land transition of plant life (Rensing, 2018; de Vries and Rensing, 2020). During its decades of *in vitro* cultivation, *P. patens* Gransden was predominantly propagated vegetatively (Ashton and Raju, 2000). While many labs vegetatively propagate the plants, others regularly let the plants go through the life cycle (sexual reproduction through selfing) and establish fresh cultures based on single spores. However, for most of the pedigrees the frequency and number of sexual reproduction events the plants went through is unknown. Phenotypic differences are documented between laboratory strains, for example Gransden strains have shown different levels of loss of fertility (Meyberg et al., 2020). This recently led to the introduction of the Reute accession for the study of sexual reproduction (Hiss et al., 2017). Mutations underlying such differences as well as potential silent mutations can occur during sexual as well as vegetative propagation in the lab. Such laboratory divergences haven been reported in both prokaryote (Smits, 2017) and eukaryote laboratory models, for example in *Chlamydomonas reinhardtii* (Flowers et al., 2015). Mutation and selection underlie the forces of evolution. However, under laboratory conditions natural selection usually is absent. Over time, somatic mutations can thus accumulate in laboratory strains that would not occur in natural populations. Indeed, repetitive vegetative propagation of *P. patens* in the laboratory loosens the selection pressure on genes required for sexual reproduction, apparently leading to deterioration of the latter (Ashton and Raju, 2000; Perroud et al., 2011; Hiss et al., 2017; Meyberg et al., 2020). It should be noted that *P. patens* is predominantly selfing in the (dominant) haploid stage, developing completely homozygous diploid sporophytes. Hence, spores result that are genetically identical to the parent even though they are the product of meiosis.

Previous *P. patens* SNP studies analyzed genomic DNA samples of different *P. patens* accessions (Hiss et al., 2017; Lang et al., 2018). However, *P. patens* gDNA samples are rare. Nevertheless, the recent publication of RNA-seq datasets (Demko et al., 2014; Frank and Scanlon, 2015; Stevenson et al., 2016; Szövényi et al., 2017; Perroud et al., 2018; Fernandez-Pozo et al., 2019) provides a source of information that can be used to detect SNPs. Due to the high number of RNA-seq samples analyzed, efficient pipeline processing is essential. A framework of a modular RNA-seq pipeline was previously published (Perroud et al., 2018). While adding to and modifying this pipeline, a powerful solution for the here presented SNP analysis was created. Due to the current lack of genomic DNA we analyzed whether the SNP analysis of RNA-seq samples leads to comparable results. Based on the called SNPs we determined the rate and nature of somatic mutations among the accessions and pedigrees.

To identify and track genetic variation in the laboratory, restriction fragment length polymorphisms (RFLP) can be employed. This technique is based on SNPs modifying restriction enzyme recognition sites, which are covered by polymerase chain reaction (PCR) amplicons to test for genetic variation in specific DNA regions (Botstein et al., 1980).

Here, we identified SNPs using recently published RNA-seq data as well as unpublished RNA-seq and gDNA-seq data for a range of *P. patens* accessions and Gd pedigrees, i.e., laboratory strains with a documented ancestry. We used the resulting data to separate accessions as well as pedigrees via SNP analysis, extracted unique SNP sets for all accessions and Gd pedigrees, and developed RFLP analyses that are useful in maintaining accession and Gd pedigree identification.

MATERIALS AND METHODS

Sequence Sources

This study used data of five different *P. patens* accessions: 171 Gransden (Gd), 20 Kaskaskia (Ka), 32 Reute (Re), 27 Villersexel (Vx), and 25 Wisconsin (Wi) samples. The dataset contains 206 previously published RNA-seq samples as well as 44 novel RNA-seq samples. In addition, 25 novel gDNA samples of *P. patens* accession Wisconsin (Wi) were analyzed. These 275 samples were used for SNP detection. In addition, the Wi gDNA samples were used to study variation in a naturally occurring population. All samples used in the present study are available at the NCBI SRA database and are detailed in **Supplementary Table S1**.

Plant Material, Nucleic Acid Extraction and Sequencing

Physcomitrella patens accession Villersexel was collected in 2003 by M. Lueth in Haute-Saone (France) on dry mud at a fish pond east of Villersexel, at the Villers la Ville junction (voucher 4296). The accession Kaskaskia was also collected in 2003 in Illinois (United States) on a periodically flooded drainage channel at a corn field by D. Vitt and M. Sargent. The voucher information for both accessions has previously been published (von Stackelberg et al., 2006; Beike et al., 2014). Accession Reute has also been collected by M. Lueth/M. von Stackelberg in 2006 close to Freiburg, Germany on an agriculturally used field. The exact location has previously been published (Hiss et al., 2017).

Reute Early Sporophyte 1 (ES1)

Physcomitrella patens accession Reute_2015 (Re_2015) (Hiss et al., 2017) was cultivated on 9 cm petri dishes on solid Knop's medium enclosed with parafilm under long day conditions (70 μmol m*−2 s*−1 white light, 16 h light, 8 h dark, 22°C) as described in Hiss et al. (2017). Re was regularly reproduced sexually once per year since 2011. Re_2015 is the culture derived from the sexual reproduction (selfing) performed in 2015. Gametangia induction was performed by transfer to short day conditions (see Hiss et al., 2017 for culture details). Sporophytes were harvested 6–9 days after watering and immediately put into 50 μl RNA-later (Qiagen, Hilden, Germany). RNA was extracted using 20 ES1 sporophytes (according to Hiss et al., 2017) using the RNeasy micro kit (Qiagen, Hilden, Germany), following the manufacturers' protocol. RNA concentration and quality were analyzed with the Agilent RNA 6000 Nano Kit on a Bioanalyzer 2100 (Agilent Technologies). Library preparation and subsequent sequencing was performed by the Max-Planck-Genome-Centre Cologne (mpgc.mpipz.mpg.de). A single library was prepared using the IVT-based low input RNA-seq protocol followed by sequencing with Illumina HiSeq 3000 (150 nt, single ended).

Kaskaskia RNA-seq

Physcomitrella patens accession Kaskaskia was isolated from seven days entrained protonemal culture under long day conditions (70 μmol m*−2 s*−1 white light, 16 h light, 8 h dark, 22°C), if not stated otherwise (**Supplementary Table S2**). Tissue was flash frozen in liquid nitrogen and the subsequent RNA extractions were performed as described in (Perroud et al., 2018). The library preparation and subsequent sequencing was processed using the TruSeq RNA kit (Illumina) according to the manufacturer's instructions. The libraries were sequenced with Illumina HiSeq (100 nt, paired-end).

Villersexel Laser Capture of Sexual Reproduction Stages

Physcomitrella patens Villersexel (Vx) plants were routinely grown under sterile conditions on ammonium supplemented medium under 20 μmol m*−2 s*−1 of continuous light at 24°C. Protonemata were obtained from ground tissue and cultivated on cellophane disks on the previous medium. After 2 weeks, small patches of protonemata were transferred to low nitrate medium and grown for about 2 months under 20 μmol m*−2 s*−1 of a 16:8 light:dark cycle at 24°C. Well-developed gametophores were then transferred to 16°C under the same light regime for 3 weeks to induce sexual organ differentiation. Fertilization was synchronized in all cultures by flooding growing pots with sterile deionised water for 30 h; flooded gametophores were transferred to 24°C under continuous light. 48 h after flooding, gametophore tips were examined under a hand dissection microscope for the presence of fertilized archegonia. Non-fertilized cultures were treated as previously except for flooding.

Fertilized and unfertilized archegonia were hand dissected and collected in 100% acetone. Tissue fixation was ensured by infiltrating archegonia under low pressure for 2 min followed by a 48 h incubation in 100% acetone. Acetone was then exchanged with HistoClear by incubating fixed tissues in 50% acetone/50% HistoClear for 2 h then 100% HistoClear for 2 h under continuous shaking. Tissues were embedded in wax using an automated Tissue Tek VIP 5 Vacuum Infiltration (Sakura) machine with the following sequence: 3 baths in HistoClear for 1, 1 and 2 h then 4 baths in wax for 1, 1, 2 and 2 h. Thick sections of 10 μm were prepared from the embedded tissues and deposited on Nuclease-free 1.0 polyethylene naphthalate (PEN) membrane slides (Carl Zeiss Microscopy, #415190-9081-000) in drops of 1 X ProtectRNA^{TM} RNase Inhibitor (SIGMA #R7397), air dried and stored at room temperature until further use. After wax removal in HistoClear and 100% ethanol baths, zygote/early embryos, egg cell and archegonium tissues were laser dissected from the sections using a PALM MicroBeam unit (Carl Zeiss) at a 40x magnification following the procedure described in Saint-Marcoux et al. (2015). About 200 sections were captured per sample and 3 biological replicates were prepared for each tissue.

RNA was extracted using the PicoPure RNA extraction kit from Life Technologies (#KIT0204) and amplified into cDNAs using the Ovation RNA-Seq System v2 kit from NuGEN (#7102-32) as in Saint-Marcoux et al. (2015). cDNA quantity was determined using a NanoDrop ND-1000 spectrophotometer. cDNA quality was analyzed on a 2100 BioAnalyzer (Agilent Technologies) using RNA nano chips (5067-1511, Agilent Technologies) following recommendations in the NuGEN kit.

1μg of cDNA was paired-end sequenced on an Illumina HiSeq 2000 platform at the Beijing Genomics Institute in China. At least 2 × 10 million 100 nt reads were obtained per sample. Samples containing "orphans" in the sample name contain reads where the mate did not pass the quality filter.

Wisconsin gDNA

Mature (brown) spore capsules of *Physcomitrium patens* were collected in September 2017 in Wisconsin, United States (original specimen in AUGIE herbarium) by Rafael Medina (Augustana College Illinois). The surface sterilization procedure was performed at a laminar flow bench with freshly prepared 1% sodium hypochlorite and autoclaved tap water for rinsing. Five Single spore capsule were sterilized separately. After the last rinsing step the water was kept in the tube and the spore capsule was squeezed by sterile forceps so that the spores were released into the water. This spore suspension was transferred (using a micro pipette and autoclaved filter-tips) to solidified (0.9% [w/v] agar) Knop's medium containing 1% glucose in 9 cm Petri dishes sealed using 3M Micropore tape or Parafilm. After 3–5 days, when spore germination starts, five single sporelings were isolated from each capsule batch and separately transferred to fresh plates. After eight weeks under long-day conditions juvenile gametophores (above agar) were harvested and immediately frozen in liquid nitrogen. Genomic DNA was isolated from frozen plant material as previously described (Lang et al., 2018). Library-preparation and sequencing was performed at the Max-Planck-Genome-Centre Cologne (mpgc.mpipz.mpg.de); 25 TPase-based DNA libraries were sequenced in 1 × 150 bp single reads on Illumina HiSeq 3000 Analyzers.

Wisconsin experiment 2 was contaminated by prokaryotic sequences. The read contamination removal was done as described in Lang et al. (2018) and Nguyen et al. (2019). The leftover reads were used for further analysis.

Read Analysis

For easier manageability of the data, all original sample names were converted to a new nomenclature. Separator is always an underscore; the first two characters identify the accession (Gransden [Gd], Reute [Re], Kaskaskia [Ka], Villersexel [Vx] and Wisconsin [Wi]), the next one the origin/pedigree of the sample (e.g., MR-WT11), followed by the experiment defined by roman numbers (e.g., XX). Sample replicates (1-5), library type (SE or PE) and experiment type (mutant [MUT] or wild type [WTY]) are the last parts (Supplementary script "rename and extraction", **Supplementary File 1**). An example sample name is Gd_MR-WT11_XX_1_PE_WTY. Each RNA-seq sample went through a modified pipeline, build on top of the RNA-seq pipeline previously described (Perroud et al., 2018). The pipeline was modified by updating all software versions, enabling single-end (SE) read processing and adding SNP calling and post processing parts (**Figure 1**).

Read Quality

For read quality filtering and adapter removal, Trimmomatic (Bolger et al., 2014) version 0.39, was used. Adapter trimming of appropriate adapters (SE.fna or PE.fna; standard sequences included in the Trimmomatic package) was performed with a seed mismatch of 2, a palindrome clip threshold of 30, and a simple clip threshold of 10 for the paired-end reads (PE.fna:2:30:10). Base pairs with a quality score less than three were removed from the start (LEADING:3) and end (TRAILING:3) of the reads. Reads were further filtered using a sliding window of four base pairs with a minimum average quality score of 15 (SLIDINGWINDOW:4:15), removal of the first 10 base pairs (HEADCROP:10), and kept reads of 30 base pairs or more (MINLEN:30).

Poly-A clipping was performed by Prinseq-lite (Schmieder and Edwards, 2011) version 0.20.4. A minimum length of five poly-A/T nucleotides at the 5′- or 3′-end were required to remove the poly-A/T tails (TRIM_TAIL_LEFT 5; TRIM_TAIL_RIGHT 5). Only reads longer than 30 nt were kept (min_len 30).

Reference Genome Mapping

All filtered RNA-seq samples were mapped to the *P. patens* reference genome V3 (Lang et al., 2018) by GMAP-GSNAP (Wu and Nacu, 2010) version 2018-7-04. SAM and BAM file processing was performed by samtools (Li et al., 2009) version 1.9. Only uniquely mapped reads were used for further analysis.

Removing Duplicate Reads

De-duplication based on the unique mapped BAM files was done using samtools package markdup with the remove duplicate reads option (r).

Variant Detection

The SNP calling pipeline (**Figures 1B,C**) uses GATK version 4.0.9.0 (McKenna et al., 2010). The workflow was setup according to the classic GATK best practices workflow for RNA-seq[1,2] by modification of the approach published earlier (Hiss et al., 2017).

SNP Calling

GATK HaplotypeCaller was performed in default mode. To account for *P. patens* being haploid, the option "ploidy 1" was used.

The python script GetHighQualVcfs.py (Wang et al., 2012) was used for quality score recalibration. The option for haploid genomes (ploidy 1) was chosen. In addition, the alternative nucleotide quality (ALTQ) needed to be higher than 90% (percentile 90) and the genotype quality (GQ) value had to be greater than 90 (GQ 90).

The GATK tools BaseRecalibrator, ApplyBQSR and PrintReads were used in default mode.

Ploidy test

To test the samples' ploidy, GATK HaplotypeCaller was performed in default mode for diploid genomes (ploidy 2).

The python script GetHighQualVcfs.py was used for quality score recalibration. The option for diploid genomes (ploidy 2) was chosen. In addition, the alternative nucleotide quality needed to be higher than 90% (percentile 90) and the genotype quality (GQ) value had to be greater than 90 (GQ 90).

The results of both ploidy runs (1n and 2n) were compared. The results were interpreted taking into account the knowledge of previously haploid tested samples (**Supplementary Table S10**; cf. Results). We observed that the differences in the defined genotypes (GATK output 0/0, 0/1, 1/1, and 1/2) correspond to the differences in the number of called SNPs. Therefore, we chose the number of called SNPs to compare the two ploidy runs.

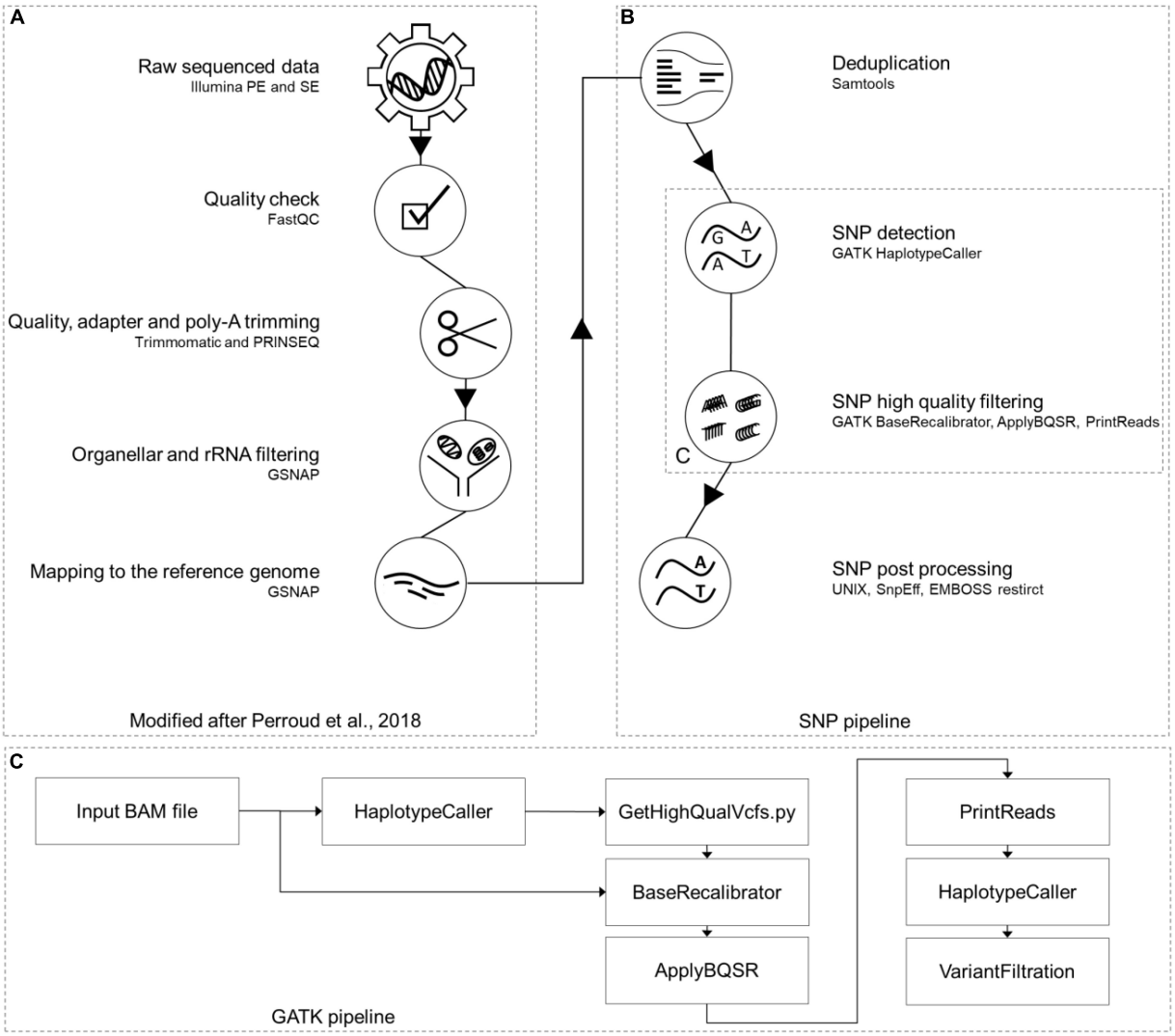


FIGURE 1 | RNA-seq SNP calling pipeline. Part **(A)** of this pipeline was previously published (Perroud et al., 2018). The additional SNP calling branch **(B)** on the right side starts with removing read duplications, using Samtools package "markdup" and continues with the GATK toolbox for SNP calling **(C)**. The last steps of this pipeline are post processing steps like SnpEff and EMBOSS restrict together with UNIX shell scripts. This figure has been modified based on a figure published in The Plant Journal (Perroud et al., 2018; https://onlinelibrary.wiley.com/doi/full/10.1111/tpj.13940).

Filtering Wisconsin gDNA SNPs

Single nucleotide polymorphisms called from the Wisconsin accession gDNA were filtered by using only reads uniquely mapping to the *P. patens* v3.3 gene annotation (to make the data comparable to the RNA-seq data). Bedtools intersect (Quinlan and Hall, 2010) version 2.29.0 was used, with the option (u) to write the original entry only once if multiple overlaps are found, to extract all gene models intersecting SNPs (Supplementary script "rename_and_extraction").

Post SNP Calling Filter

The JGI gene atlas samples contain spike-in RNAs, which should not harbor SNPs. Hence, based on SNPs detected in these reads, filters were adjusted so that none of the RNA-seq spike-in base changes (sequencing errors) pass it. Filter values were allelic depths for the reference and alternative alleles (AD), mapped read depth (DP) as well as their fold change (FC) plus a minimum of three samples per SNP.

The above described values were adjusted through three consecutive filter steps. (i) The first filter was the read coverage filter with a minimum read depth of nine reads and a minimum of seven reads supporting the SNP. FC of AD and DP has to be greater than 0.77 (Supplementary script "SNP_filtering"). (ii) The second filter step removes all SNPs not present in at least three samples. This filter ensures the use of SNPs found by all technical triplicates of an experiment. (iii) While the third filter removed all indel positions.

The GO bias analyses were conducted as described previously (Widiez et al., 2014) to contrast gene sets affected by SNPs vs. the background of all genes. Visualization of the GO terms

was implemented using word clouds generated by https://www.wortwolken.com. Word size is proportional to the –log10 (q-value), and over−represented GO terms were colored dark green if –log10 (q-value) ≤ 4 and light green if –log10 (q-value) > 4.

Plots were done by using R version 3.6.2 and ggplot2 version 3.2.1. Upset plot for the SNP intersection was performed with the R package UpSetR (Conway et al., 2017). All regression lines and confidence intervals were calculated by the R package ggplot2, method "lm" and the R package ggpubr version 0.2.5 to calculate $R^{1,2}$.

SNP Normalization

Several plots (**Supplementary Figures S3–S6**) were generated to check for potential normalization methods. The number of read covered base pairs (coverage), the number of reads per sample (reads), and the number of genes, respectively their accumulated length (genes) were taken into account.

Coverage method

The dependency of called SNPs based on the number of read covered based pairs was determined with the following method.

To find all read covered base pairs, the mapping output (BAM format) was analyzed by samtools package depth. All sequence positions, including unused reference sequence positions, were printed (aa). The output was filtered for depth ≥ 9 (similar to the SNP DP value). The number of filtered SNPs were divided by the number of read covered base pairs. To compare the values directly with the results found in the division was done vice versa to derive the format "one SNP per X bp".

To plot the values, the number of SNPs were corrected by the maximum number of read covered base pairs (**Supplementary Figure S4**).

Reads method

To detect the relation between the number of filtered SNPs and the number of sequenced reads, the values were plotted using the R packages described in section "Post SNP Calling Filter."

Genes methods

To answer the question whether SNPs accumulated at specific chromosomes and to observe the relation between the number of genes or their length with the number of detected SNPs, gene information extracted from the *P. patens* v3.3 annotation GFF file (Lang et al., 2018) was used. Both, the number of genes and the gene length, were summarized per chromosome. The extracted gene values were divided by the number of filtered SNPs to derive relation in the gene number and gene length plots, respectively. To test for significance Fisher's exact test was performed. The number of base pairs w/o SNPs for each of the 27 individual chromosomes (and for all unassigned, merged scaffolds) was compared. All p-values were corrected using the R method p.adjust using the method (Benjamini and Hochberg, 1995).

[1]https://gatkforums.broadinstitute.org/gatk/discussion/3892/the-gatk-best-practices-for-variant-calling-on-rnaseq-in-full-detail

[2]https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels

Extracting Exclusive SNPs

In the context of SNPs found only in a specific accession or Gd pedigree, the terms unique and exclusive are used synonymously. Exclusive SNPs were extracted for each accession and for each Gransden pedigree, using bash/awk scripts (Supplementary scripts "rename_and_extraction", "SNP_clustering" and "SNP_filtering"). First, all SNPs found in all GATK VCF files were grouped into a single file. Subsequently, the groups were inspected for SNPs exclusive for a specific accession or Gd pedigree (Supplementary script "SNP_filtering"). For further accession analysis, the SNPs were sorted by the number of supporting samples. SNPs supported by $> 90\%$ of the samples of one accession, and not found in others, were defined as exclusive. The read coverage filter was not applied for the accession exclusive SNP selection. For the Gd pedigrees, the Gd exclusive SNPs were ranked by the number of supporting samples. The SNP with most sample support received the highest rank, the five SNPs with the most sample support were chosen and defined as exclusive.

Accession Clustering

Detected nucleotide variation was clustered by two different methods. The first method was an artificial FASTA alignment (**Supplementary File 4**). This method clusters only SNPs, no InDels. Only SNPs that passed all filter steps were used. Each SNP is a single column in the alignment. If the sample contains a SNP at a specific position, the SNP nucleotide was added to the FASTA sequence of the sample, otherwise the reference nucleotide was used.

The second method was chosen to cluster SNPs and InDels. Instead of nucleotides, numbers were chosen to represent a SNP, InDel or the reference. A matrix was created by substitution of reference and variant nucleotides: reference 0; SNP 1; indel 2. This converted numbers were added to the matrix similar to the nucleotides in the above described FASTA file. Each row is a single sample and each column a unique SNP/indel position (Supplementary script "SNP_clustering").

The artificial FASTA alignment was imported to SplitsTree (Huson and Bryant, 2005) version 4.14.8. A network was calculated using default parameters. The tree was generated by the NJ option and stored in NEXUS format. FigTree (Bouckaert et al., 2014) version 1.4.4 was used to draw a circular tree based on the SplitsTree NEXUS file.

The SNP/indels 0-1-2 matrix was loaded into R version 3.6.2 using the function dist with the method euclidean. To get a three dimensional PCA plot, the results were transferred to the R package rgl version 0.100.30.

SNP Effects

Synonymous and non-synonymous SNPs for each sample were detected by SnpEff (Cingolani et al., 2012) version 4.3T in default mode. SnpEff used a database created of the *P. patens* genome annotation v3.3 to locate SNP positions at gene regions. Only SNPs that passed all three filter steps (minimum nine reads have to cover the SNP position and minimum seven reads have to support the SNP, at least three samples have to support the SNP, indels are removed) were used.

Synonymous and non-synonymous SNPs were extracted from the SnpEff CSV file output and all involved genes were extracted from the SnpEff gene.TXT file. Functional analyses were done via GO-bias analysis, described in chapter "Post SNP Calling Filter."

Identification of Restriction Sites Overlapping With SNPs

EMBOSS restrict[3] was used to detect SNPs in putative restriction endonuclease recognition regions. The enzyme database, containing all necessary information about the recognition sites, was loaded with the tool EMBOSS rebaseextract[4]. The rebase restriction endonucleases databases, withrefm.907 and proto.907, were downloaded at ftp://ftp.neb.com/pub/rebase. EMBOSS restrict was performed with a minimum length of the restriction enzyme recognition site of five base pairs (sitelen 5) and all enzyme at the database were used (enzymes all).

SNP Verification via PCR and RFLP (Restriction Fragment Length Polymorphism)

Exclusive SNPs for each *P. patens* accession overlapping with a restriction enzyme recognition site were selected as described above. SNPs affecting six or eight nt long recognition sites were chosen. Additionally, enzyme requirements for easy usability and frequency of cuts in ± 2 kbp around the SNP were analyzed to ensure an interpretable gel band pattern. Primers were designed to result in an amplicon of 700-1,400 bp and similar annealing temperatures (∼ 59°C, **Supplementary Tables S7**, **S8**).

Plant Material and gDNA Extraction

To analyze SNPs located within restriction enzyme sites (comparison of accessions) and SNPs without restriction enzyme site (comparison of Gd pedigrees) the *P. patens* accessions and Gd pedigrees Gransden DE Marburg 2015 (Gd_DE_MR), Gransden Japan (Gd_JP, Gd_JP_Okazaki and Gd_JP_St.Louis), Gransden Grenoble (Gd_CH), Reute 2015 (Re), Kaskaskia (Ka) and Villersexel (Vx) were cultivated as described above. Genomic DNA for PCR amplification was isolated, using a fast protocol using one to two gametophores as published in (Cove et al., 2009).

PCR Analysis and Sequencing

Polymerase chain reaction was carried out with OneTaq polymerase (NEB) following the manufacturers' protocol. Annealing was carried out between 55°C and at 59°C and elongation time was adjusted to the longest fragment chosen (95 s). For primer sequences see **Supplementary Tables S7, S8**. 5 µl PCR product, 2.5 µl of the forward primer (10 µM) and 2.5 µl water were Sanger sequenced (Macrogen, Germany) (**Supplementary Table S9** and **Supplementary File 6**). PCR products and all subsequent fragment analyses were visualized via gel electrophoresis (0.7% agarose, Roth, Germany) using peqGREEN (VWR, Germany) as dye. The 1 kbp size standard was purchased from NEB.

[3]http://emboss.sourceforge.net/apps/cvs/emboss/apps/restrict.html

[4]http://emboss.sourceforge.net/apps/cvs/emboss/apps/rebaseextract.html

Restriction Analysis

For each tested SNP, 15 µl PCR product of all accessions were used as input for the enzymatic digestion. Restriction was carried out for the SNPs Re_c3_17747483_A-T, Vx_c3_2712099_A-G and Ka_c01_25061888_C-A using 2U of the corresponding enzyme (**Supplementary Table S7**, NEB) for 3 h at 25°C for *Swa*I and at 37°C for *Nde*I and *Xba*I. Fragments resulting from the restriction were visualized via gel electrophoresis as described before (PCR analysis and sequencing).

Natural Population Diversity

To determine variation within a naturally occurring *P. patens* population, the accession Wisconsin gDNA SNP results were used. Because of bacterial contamination, sample Wi_2 was excluded from this study. The experiment was designed with four capsules and five spores each. Each spore represents one sample. The number of exclusive SNPs for each sample (spore) within a spore capsule were detected as well as the number of exclusive SNPs for each spore capsule. The results were compared with the results of exclusive SNPs found in laboratory accessions and pedigrees of Gransden, Gd_DE 2011, 2012 and 2015, and Reute 2007, 2012 and 2015. To highlight the results Venn diagrams were created by venny[5].

All samples described above were used to generate an artificial FASTA alignment (for methods see section "Extracting exclusive SNPs") which was analyzed by Splitstree. Here, only gDNA SNPs which intersected with the *P. patens* v3.3 annotation file were kept. The branch lengths were adjusted by coverage normalization (see section "Coverage method").

RESULTS

Read Analysis and SNP Discovery

The analysis was conducted with a total of 4.7 billion RNA-seq reads (**Supplementary Table S3**). 68% of all reads are from Gransden, Reute reads account for 18%, Kaskaskia for 12% and Villersexel for 2% (**Supplementary Table S4**). After pre-processing and mapping to the reference genome (**Figure 1A**) 81% of all reads remained (**Supplementary Table S3**). De-duplication (to account for potential PCR bias) further reduced the amount of reads by 20%, leaving 3.0 billion reads as input for the GATK SNP pipeline (**Figures 1B,C**). The unfiltered Wisconsin gDNA samples amounted to 1.0 billion reads. Processing, mapping to the reference and deduplication discarded more than half of the raw reads; 473 million reads were used for the SNP pipeline (**Supplementary Table S3**).

Funariaceae are known for naturally occurring polyploidization (Rensing et al., 2013; Beike et al., 2014), this has also been demonstrated during *P. patens* mutant generation using protoplasts (Schween et al., 2005). We performed a ploidy test using GATK with $n = 1$ vs. $n = 2$ and generally detect a lower number of SNPs when assuming haploidy ($n = 1$), on average 65.4% of $n = 2$. The percentage range of samples confirmed to be haploid (36.2 – 92.2%) approximately

[5]https://bioinfogp.cnb.csic.es/tools/venny/index.html

coincides with the percentage range of all samples (30.7 – 92.9%) (**Supplementary Table S10** and **Supplementary File 7**). Moreover, manual inspection of the VCF files for the Wi gDNA SNP calls showed very minor differences, that are smaller than those of the RNA-seq data of confirmed haploid plants. Taken together, we do not find evidence for polyploid plants among the samples used.

For the Wisconsin gDNA samples 2,473,107 SNPs were called by the GATK pipeline (**Figure 1C**). After intersecting the gDNA SNPs with the gene coordinates of the *P. patens* v3.3 annotation, 140,832 SNPs were kept that represent the transcriptome, to be comparable to the RNA-seq SNPs. Merging the Wi v3.3 SNPs with the results of the RNA-seq accessions ended up in a total number of 1,233,585 transcribed gene space SNPs. Gd has the lowest number of SNPs relative to the reference assembly. This fits the expectation, since the reference genome was derived from a Gd pedigree. The accessions Wi and Ka have the highest number of SNPs per sample (**Supplementary Figure S1**). The highest SNP reduction can be observed after the (i) read coverage filter, which was, together with the (ii) sample support filter, adjusted using spike-ins (see section "Materials and Methods" for details). (i) Read coverage and (ii) sample support filter, together with the (iii) indels removal, were reducing the SNP set by 88% (146,816 SNPs shared by five accessions, **Supplementary File 5**). A comparison of SNP intersection between SNPs called in this study and SNPs previously published (Lang et al., 2018) demonstrates a large overlap of 89% of the previously called Vx SNPs (as compared to those that were detected in this study) and minor overlaps for Re (26%) and Ka (28%) (**Supplementary Table S5**).

TABLE 1 | Average number of SNPs affecting gene coding sequences per sample.

	Gd	Re	Vx	Ka	Wi	all
Start changes[a]	3	8	8	18	29	8
Stop changes[b]	6	23	22	97	237	41
Sequence changes[c]	774	2,978	2,942	11,794	13,090	3,168
Synonymous	411	1,232	1,272	4,698	4,737	1,300
Non-synonymous	363	1,746	1,670	7,096	8,353	1,868

[a]*Start changes include start codon gains and losses.* [b]*Stop changes include gains and losses of stop codons.* [c]*Sequence changes include non-synonymous changes affecting the encoded amino acids, synonymous sequence changes, and insertions or deletions that do not change the sequence frame.*

SNP Comparison Between Accessions

Most SNPs can be observed in the intergenic regions (up- and downstream of the gene bodies according to the v3.3 annotation). The SNP distribution for all accessions is around 40:60 (gene regions/intergenic regions). The accessions Wi and Vx have almost no SNPs flanking the two base pairs next to the splice site (splice site region).

Most of the SNPs shown in **Figure 2** are accumulated in non-coding regions. Exonic SNPs can be synonymous, not affecting the coding sequence, or non-synonymous, leading to a change in the amino acid sequence of the protein encoded by the gene (for average number of SNPs per sample see **Table 1** and for total numbers of SNPs see **Supplementary Table S11**). The two accessions from North America, being geographically most far away from the reference sample, are the ones with the most changes affecting the coding sequence. Individual SNP effects in the exclusive accession SNPs list can be found in **Supplementary File 3**.

Less than 12% of all SNPs called by the GATK pipeline passed all three filter steps: Gd has 39,614 and Re has 42,094 SNPs left, Vx has 52,960 SNPs and Wi has 63,597 SNPs. The highest number of SNPs are found in Ka with 76,076 SNPs (**Supplementary Figure S1** and **Figure 3**, left horizontal bars). The number of SNPs coincides with the geographical distance to the reference Gransden (**Figure 3**, horizontal bars; **Supplementary Figure S2**). After applying four different normalization methods (see section "Materials and Methods" for details), Gransden and Reute exhibit always the lowest SNP rate (**Supplementary Table S6**), mirroring previous results (Beike et al., 2014; Lang et al., 2018). The approximate linear relationship between number of reads and called SNPs (**Supplementary Figure S3**) led to the normalization by read number. The coefficient of determination (R^2) is found to be 0.6 – 0.93 (**Supplementary Figure S3**). To compensate for unequal distribution of reads we also normalized by the fraction of the sequence space that carries enough read support to allow SNP calling (see section "Coverage method," **Supplementary Figure S4**). By applying the SNPs to read covered base pairs, instead of the raw read number, the R^2 values increased. Wi, Ka and Vx reach almost 1, Re and Ge 0.77 and 0.85. Based on the coverage normalization (**Supplementary Figure S5**), Gd has 1 SNP per 4,666 bp, Reute has 1 SNP per 1,912 bp followed by Ka (1 SNP per 630 bp), Wi (1 SNP per 206 bp) and Vx (1 SNP per 143 bp). The gene normalization methods (**Supplementary Figure S6**) indicate that chromosome 19 and chromosome 26 exhibit significantly ($q \leq 0.05$) more SNPs than the other chromosomes.

The SNP intersection shows 1,541 SNPs are shared by all accessions (**Figure 3**). There are accession specific SNPs (exclusive SNPs) as well. Most exclusive SNPs are present in Vx (31,818), followed by Wi (22,014), Ka (7,905), Re (4,793) and Gd (2,184) (**Figure 3**, vertical black bars). Gd is sharing 94% of its SNPs with other accessions, Ka and Re share > 87%, Wi shares 65% and Vx 40% SNPs with all other accessions.

Applying a filter to extract exclusive SNPs supported by ≥ 90% of the samples, Wi and Ka have most exclusive SNPs/InDels, Wi has 4,007 unique SNPs, Ka 3,393. 890 SNPs are only present in the Re accession while in the Vx accession 21 exclusive SNPs were found (**Supplementary File 3**).

100 kbp SNP hotspot regions were detected to survey the *P. patens* accessions (**Supplementary Figure S7** and **Supplementary File 2**). On Chr26, starting at 300,000 bp, a hotspot region is present in all accessions. All accessions but Gd share one region on Chr19. Gd, Re and Ka share 100 kbp hotspot regions on Chr03 and one on Chr06. Ka, Wi, and Vx share regions on Chr04, 07 and 13 (**Supplementary Figure S7** and **Supplementary File 2**). Biased GO terms of the described regions are shown in **Supplementary Figure S8**. Most 100 kbp SNP hotspot

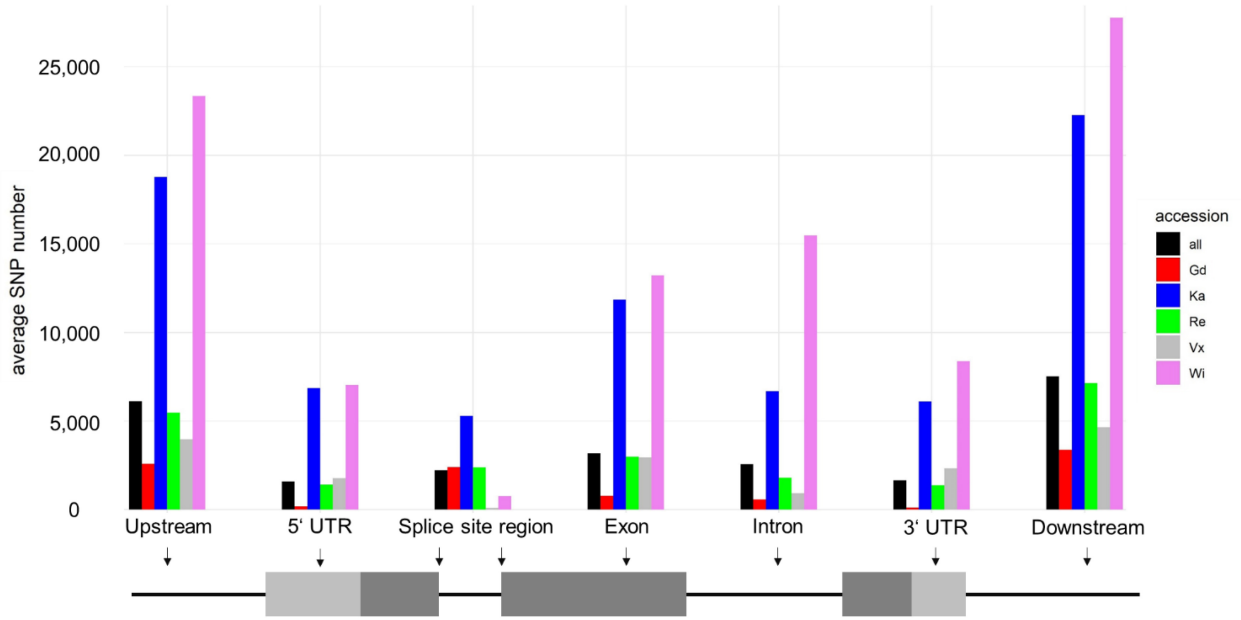


FIGURE 2 | Average snpEff output for each accession. Shown are average numbers of SNPs affecting specific regions, highlighted in a schematic gene structure shown below the corresponding grouped columns. Most SNPs are up- and downstream of genes (intergenic). SNPs at splice site regions are intron SNPs, located on the first and last two intron base pairs.

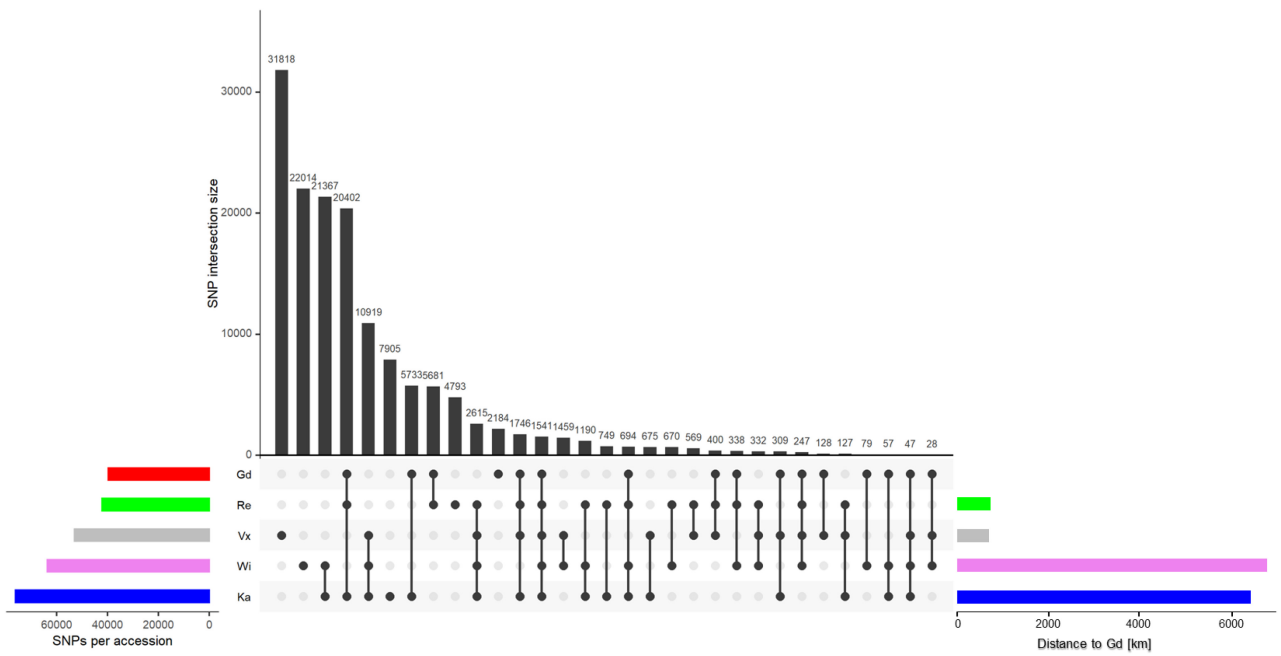


FIGURE 3 | SNP intersection of the five accessions. The horizontal colored bars on the left show the total number of SNPs per accession after applying all three filter steps. The bars to the right show the geographic distance to the reference Gd. The colors represent the five accessions throughout the text. The vertical black bars show the number of intersecting SNPs, marked by the dots below.

regions are overlapping with the SNP hotspots found by (Lang et al. (2018); **Supplementary File 2**, Table B). However, there are also a few hotspot regions only found in the present study.

Using an artificial FASTA alignment of all SNPs, we performed a clustering analysis (**Figure 4**). Samples of the accessions Gd, Re, Ka, Vx and Wi are clustering with each other, respectively, indicating that our approach is able to detect the respective

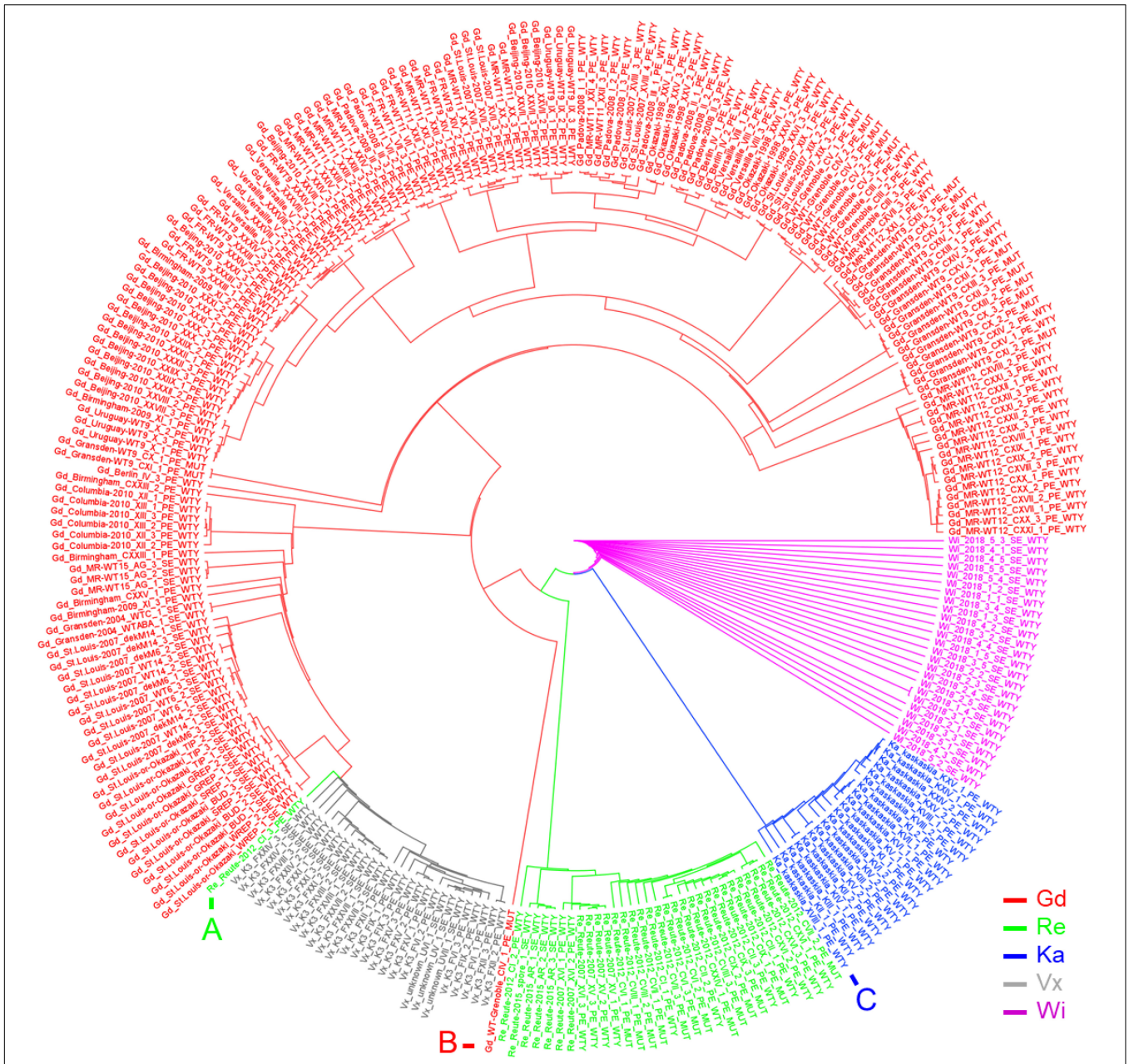


FIGURE 4 | Circularized SplitsTree network based on an artificial FASTA SNP alignment file. The Neighbor-Joining tree of five *P. patens* accessions is shown. All libraries cluster within their accession and applied treatment, except for the marked libraries: **(A)** Sample Re_REUTE-2012_CI_3 has 100 x lower read coverage than the other Re samples. It clusters next to the low read coverage Vx samples. **(B)** Sample Gd_WT-Grenoble_CIV_1 is a Gd outgroup. **(C)** Ka sample which was falsely annotated as Gd at the NCBI SRA (XVIII_1_PE_WTY), determined by exclusive SNP analysis (**Supplementary File 3**, Sheet Ka_exclusive_SNPs).

genetic background. The three European accessions form a clade to which Ka and Wi are sister. One Re sample, belonging to the experiment CI_3 (NCBI BioProject PRJNA411193), does not cluster with the other Reute samples (**Figure 4A**). The number of reads in this sample is 100 x lower than in the other samples of experiment CI, potentially causing biased SNP calling and hence incorrect clustering. The Gd sample CIV_1 (**Figure 4B**) possesses an outlier position with regard to the other European samples. The sample of the NCBI BioProject PRJNA411163 is annotated as Gransden accession. However, it could be shown by clustering (**Figure 4C**) and exclusive SNP analysis that the sample belongs to the accession Kaskaskia. Principal component analysis (PCA) of SNPs as well as InDels recapitulates the SNP clustering results (**Supplementary Figure S9**). The samples from Szövényi et al. (2017) went into the SNP calling pipeline as a blind test. The sample origin was originally marked as unknown. Both clustering methods assigned the samples to Vx, with corresponds to the origin confirmed by the authors.

SNP Comparison of Gransden Pedigrees

Gransden is more widely used in laboratories than any of the other *P. patens* accessions. Based on information retrieved from the laboratories involved, the Gransden accession was classified into four pedigrees, Germany (DE), United Kingdom (UK), Switzerland (CH) and Japan (JP) (**Figure 5**). The original Gransden accession from the United Kingdom made it first to Hamburg, Germany (founding the DE pedigree), before it was sent to Lausanne, Switzerland (CH) and Okazaki, Japan (JP). The Lausanne strain was sent to Versailles, France and further distributed to Padova, Italy and Grenoble, France. In 1998, Gransden DE arrived in Freiburg, Germany. In Freiburg the Gd plants went through sexual reproduction (selfing) once per year. Starting 1999 the Freiburg pedigree went through nine rounds of selfing leading to WT9. The offspring were labeled by consecutive numbers or the year of sexual propagation. Gransden Freiburg (WT9) was sent to Uruguay, Beijing (China) and Marburg, Germany. Gransden Marburg started in 2011 and went through selfing each year except 2013. The Gd United Kingdom 2004 sample was sent to St. Louis, United States (**Figure 5a**) for gDNA isolation and used to sequence the *P. patens* reference genome (Rensing et al., 2008). However, the Gd UK 2004 reference sample was not broadly distributed. In 2007, another Gd sample was sent to St. Louis, USA from Okazaki, Japan. These plants were used for further analysis and also sent to Columbia. It should be noted that most papers that cite the reference genome paper with its Gd 2004 sample are actually using different pedigrees.

Our analyses show that Gransden accumulated different mutations in different laboratories during prolonged *in vitro* culture. To eliminate misleading SNP background noise, the exclusive SNPs for the Gd pedigrees were detected after applying read coverage and sample support filters. The intersection of the four Gd pedigrees (**Supplementary Figure S11**) shows that Gransden Germany (DE) has 1,112 exclusive SNPs while Gd_CH has 67 exclusive SNPs, Gd_JP 187 and Gd_UK features four (**Figure 5**). Because there is no SNP supported by at least 90% of all samples of a specific pedigree, the extraction of exclusive SNPs was done by getting the best supported SNPs. SNP ranking by the number of samples that support it was used to select the five most supported SNPs for a given pedigree. The Gd_DE top five SNPs are supported by 76–77 samples, Gd_CH between 12 and 18 samples, Gd_JP by 12 to 29 samples. For Gd_UK three samples support the top five list (**Supplementary File 3**). A clear clustering based on the FASTA alignment file, as for the accessions, is not possible (**Supplementary Figure S10**). In some cases, the samples grouped by experiments instead of Gd pedigree, which could be due to the low number of SNPs, and similar genes being expressed, biasing the number of available SNPs for the comparisons. If samples are highly specific for a single tissue (e.g., antheridia bundles or spores), not all genes are covered by the extracted transcripts and consequently SNPs cannot be detected.

Since some of the samples have a documented sexual propagation history (i.e., we know how many years/cycles of sexual reproduction lie between samples) we used the opportunity to determine whether SNPs were generally lost or gained in these samples. We find that for samples that were subject to regular sexual reproduction, SNP numbers generally decreased along the timeline (**Supplementary Table S12** and **Supplementary Figure S12**). The observed mutation rate was found to be similar across the different pedigrees (**Supplementary Table S13**).

Experimental Confirmation of Selected SNPs via Sequencing and RFLP Analysis

For all primer pairs (**Supplementary Tables S7, S8**) covering SNPs specific for different accessions, PCR amplicons could be generated. Sequencing analysis of the PCR products showed in all tested positions (9/9 positions, **Supplementary Tables S7, S8**) the presence of the predicted SNP in the corresponding accessions' and Gd pedigree background (**Supplementary Figures S13–S17**). To provide an easy and cheap tool to distinguish the different accessions, RFLP analysis (**Figure 6**) was successfully established for the SNPs Re_c3_17747483_A-T, Vx_c3_2712099_A-G and Ka_c01_25061888_C-G (**Supplementary Figures S13–S15**). The Re_c3_17747483_A-T amplicon (1,255 nt) was digested with *Nde*I resulting in two fragments (990 nt and 265 nt) for the accessions Gd, Ka and Vx, and absence of digestion in Re (**Supplementary Figure S13**). For Vx_c3_2712099_A-G, the amplicon of 1,366 nt was digested with *Swa*I leading to two fragments (1,063 nt and 303 nt) in Gd, Ka and Re but not in Vx (**Supplementary Figure S14**). For Ka_c01_25061888_C-G, the 1,342 nt amplicon was digested with *Xba*I resulting in two fragments (984 nt and 358 nt) in Gd, Re and Vx, but no digestion in Ka (**Supplementary Figure S15**). Results for SNPs not tested by RFLP (**Supplementary Table S8**) for two accession primer pairs (Re_c04_21933417 and Vx_c13_4764050) and five Gd pedigree primer pairs (Gd_DE_c02_12750876, Gd_DE_c05_3105395, Gd_DE_c12_2095061, Gd_JP_c20_868 8243, Gd_CH_c23_11248087), show the presence of the predicted accession and Gd pedigree SNPs on the sequence level (**Supplementary Figures S16**, **S17**).

Natural Population Variation and Selection

Samples of pedigrees with known propagation history were chosen to estimate the annual number of mutations per base pair (observed mutation rate). The time period covered is six years for Gd and eight years for Re. The number of SNPs called for all pedigrees generally decreases under regular sexual propagation. The same is true for the estimated mutation rate (**Supplementary Table S13**). The lowest annual mutation rate with 2E-07 was detected for the Freiburg WT11 (FR_WT11) pedigree, the highest rate for Reute-2012 with 4E-06.

The diversity of genome-wide SNPs found within the Wisconsin natural population single spore isolates is lower compared with three selfed generations (pedigrees) of laboratory accessions. The lower numbers can be observed both on sample/spore and on pedigree/capsule level (**Supplementary Figure S18**). However, on the level of the artificial FASTA alignment of the gene body SNPs, represented

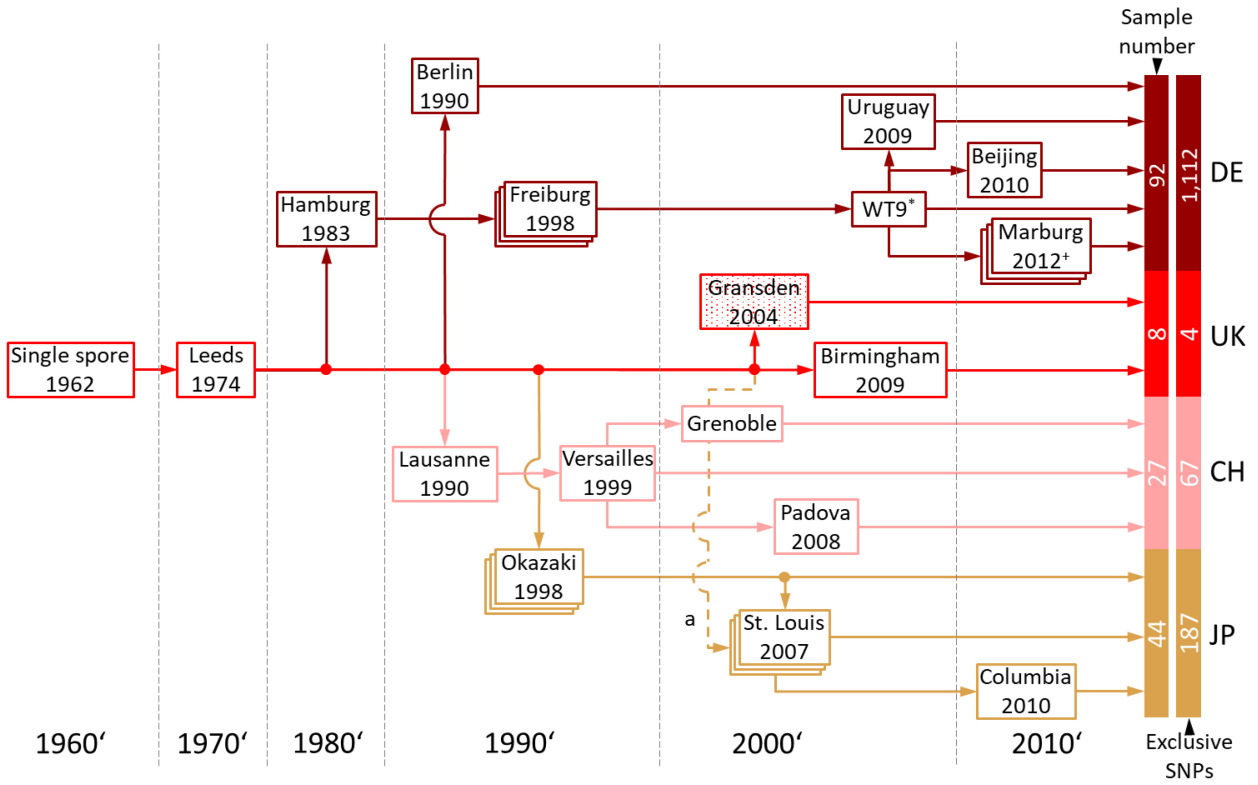


FIGURE 5 | Gransden pedigree. The pedigree diagram shows Gransden strains of 13 different labs used in the present study. The Gransden accession was arranged in four different pedigrees, Germany (DE), United Kingdom (UK), Switzerland (CH) and Japan (JP). The United Kingdom pedigree was sent to St. Louis, United States in 2004 and used to sequence the reference genome (a). However, this strain was not used or broadly distributed afterwards. The plants analyzed in St. Louis are derived from the Japan pedigree (2007). Pedigrees shown in stacked boxes went through yearly selfing. + Since 2011 yearly selfing except 2013. * Since 1999 Gransden Freiburg went through nine generations leading to WT9. The numbers of samples and exclusive SNPs for each of the four pedigrees are shown to the right (also shown in **Supplementary Figure S11**).

by a Splitstree tree (**Figure 7**), similar normalized branch lengths for Wi samples and most Re and Gd pedigrees can be observed.

The ploidy test using GATK with $n = 1$ and $n = 2$ resulted in a high rate of congruence for Wi. The $n = 1$ explained 84.3% – 95.6% (average 88.1%) of the SNPs called in the $n = 2$ run (**Supplementary File 7**). Approximately 18% of the Wi SNPs are heterozygous, a lower number than for any of the other accessions/pedigrees (**Supplementary Table S14**). Hence, the naturally occurring heterozygosity of the Wi population is lower than that observed in cultured samples. Much of what is detected as heterozygous is probably due to very closely related (identical and near-identical) paralogs that are known to be present in the *P. patens* genome (Rensing et al., 2008). Yet, the low apparent Wi heterozygosity reinforces that *P. patens* is a predominantly selfing species (Perroud et al., 2019; Meyberg et al., 2020; Rensing et al., 2020).

We calculated the rate between non-synonymous nucleotide changes (Ka) and synonymous changes (Ks) per sample and accession (**Supplementary Table S11** and **Supplementary File 7**). Over all samples, the Ka/Ks rates follow a clear linear trend ($R^2_{adj} = 0.98$, **Supplementary Figure S19**), suggesting neutral evolution (no global selective pressure). However, most individual samples deviate from the 99% confidence interval of the linear regression and hence putatively show evidence of negative selection (Ks ≫ Ka), or positive (Darwinian) selection (Ka ≫ Ks). The accession Gd, which represents the genome reference, apparently is under negative selection, all the other four accessions show evidence of positive selection (**Supplementary Table S11** and **Supplementary File 8**). The GO bias of genes affected by non-synonymous changes was calculated and visualized via word clouds (**Supplementary Figure S20**).

DISCUSSION

Read Analysis and SNP Discovery

Here, we analyzed sequence variants in *P. patens* accessions and Gransden pedigrees using mainly sequences from gene expression (RNA-seq) experiments. Therefore, this study is limited to the gene space, lacking information of most of the intergenic regions, where the selection pressure is lower and more changes accumulate (Krasovec et al., 2017). On the other hand, the advantage of using RNA-seq data is the much higher availability of data. Very few genomic data sets, and with low sequencing depth, are currently available

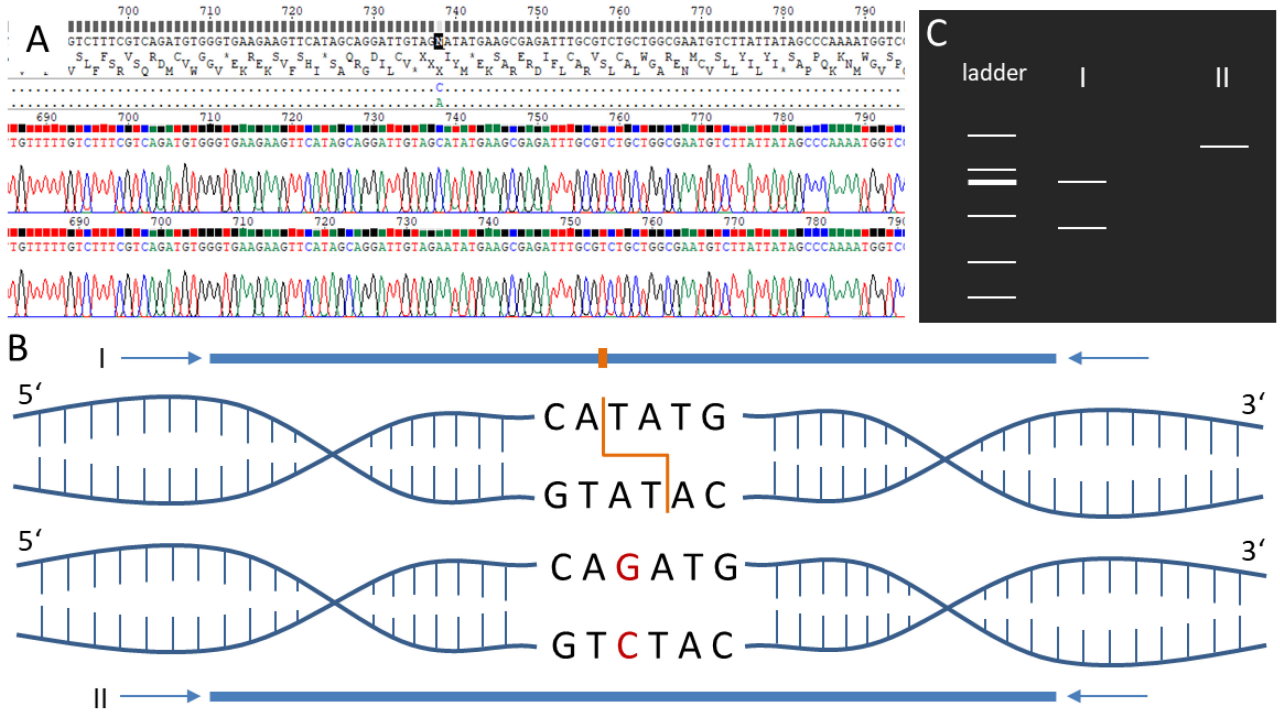


FIGURE 6 | Schematic visualization of the restriction fragment length polymorphism (RFLP) analysis. **(A)** Electropherogram of the sequenced amplicons generated via PCR using forward and reverse primers. **(B)** PCR amplicons of the samples I and II covering the same genomic position in two different *P. patens* accessions. Sample I sequence includes a restriction enzyme site for *Nde*I (orange). Sample II contains a single nucleotide polymorphism (SNP, red) resulting in the loss of the restriction enzyme site. **(C)** If amplicons are digested via the corresponding restriction enzyme Nde*I*, sample I results in two bands when separated via gel electrophoresis, whereas sample II results in one band. See **Supplementary Figures S13**–**S15** for experimental verification of the accession-specific RFLP regions.

for *P. patens* accessions and Gransden pedigrees. However, hundreds of RNA-seq experiments could be used in this study, allowing much higher resolution to detect sequence variants in genes. To ensure the quality of the SNPs found, several filters were applied. Finding a feasible filter for the called SNPs is a major step during the analysis due to risk of over- or underestimation. The presence of RNA spike-ins in some of the samples, which mimic natural eukaryotic mRNAs, gave us the opportunity to distinguish sequencing/mapping errors from actual sequence variants.

Single nucleotide polymorphisms filtering is required to reduce the false-positive rate of SNP detection. Amplification errors during sample preparation and sequencing (Ma et al., 2019) can lead to incorrectly called SNPs as well as software issues while mapping and SNP calling (Ribeiro et al., 2015). We used RNA spike-ins to detect such false-positive SNPs. Spike-ins do not exhibit SNPs. Hence, all called SNPs in spike-in mRNAs represent sequencing or computation errors. The read depth filter was adjusted to remove spike-in SNPs without losing too much sensitivity. GATK output VCF files contain a lot of information about the background data of the SNP, *inter alia*, read coverage at the SNP position. By extracting all spike-in SNPs and evaluating different parameters, the read coverage parameter [DP] and the parameter of how many reads at that position were supporting the SNP [AP], seemed to be the most feasible parameters to filter out spike-in SNPs. The number for DP of nine reads was chosen because only 4/381 spike-in SNPs were left after applying that filter (equaling 1% false positives; at DP = 10 the sensitivity breaks down). Another observation led to the sample support filter. SNP variation of more than 30% between replicate RNA-seq samples could be observed (**Supplementary Figure S18,A**). Using only SNPs found in at least three samples removed the last four false positive spike-in SNPs and makes the remaining SNPs more reliable. The improvement of filtering can also be observed by comparing the results with previously detected SNPs. The intersection of SNPs called in this study and SNPs found by Lang et al. (2018) shows an increasing number of intersection by applying the three filtering steps (**Supplementary Table S5**). The SNPs found for Re and Ka maybe have been under-estimated by Lang et al. (2018). The accessions Re and Ka have a 10 x lower number of SNPs compared to the accession Vx (Lang et al., 2018). Here, the number of intersecting SNPs between (Lang et al., 2018) and our results shows an almost 90% intersection of Vx SNPs at the strictest filter step. For Re and Ka, the intersection is less than 30% (**Supplementary Table S5**). Potentially, the absence of Re and Ka SNPs in the previous study is a result of sub optimally adjusted filter parameters or it could be an effect of low read coverage. Sufficient read depth at library level, large number of read mapping/coverage and high sequencing quality are major foundations for high quality SNP calling results. In some cases, it is possible that some SNPs were not found in one

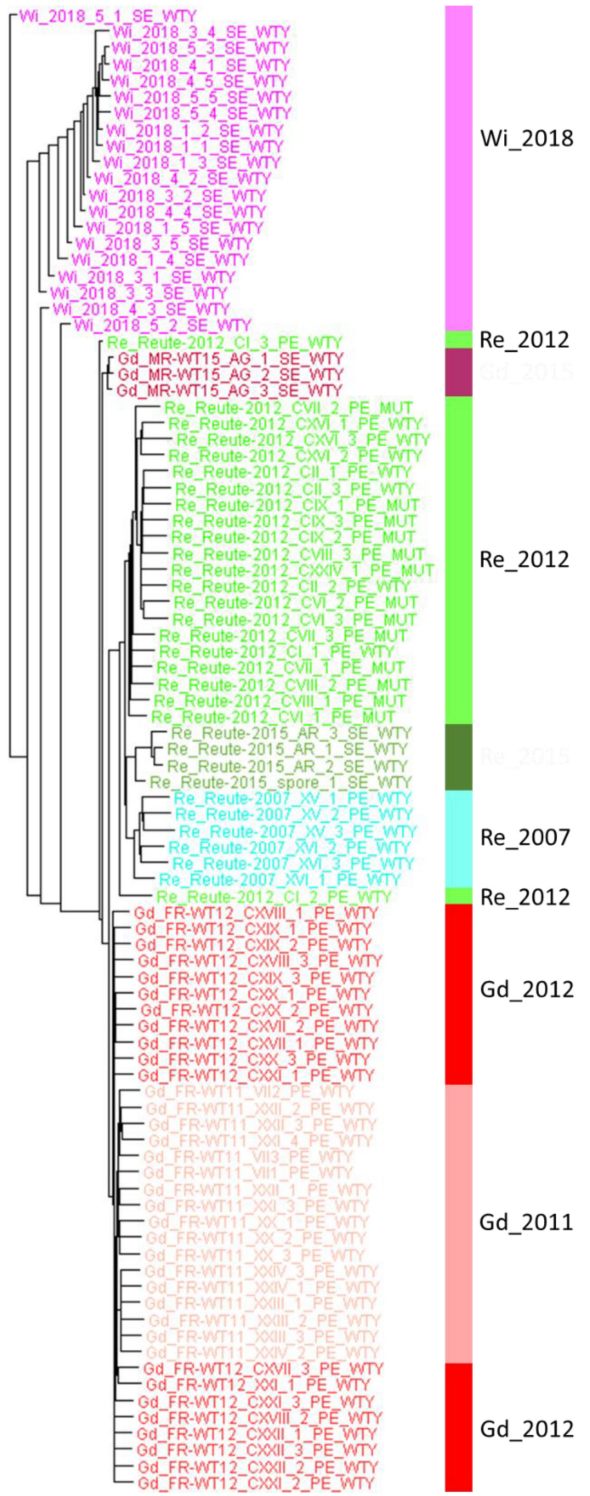


FIGURE 7 | Splitstree tree of Wisconsin natural population and three generations of Gd and Re. The tree is based on part of the artificial SNP FASTA alignment containing Wi samples (without the bacterial contaminated spore capsule experiment 2) and three generations of Re (2007, 2012 and 2015) and Gd (2011, 2012 and 2015). The Splitstree network tree was branch length-corrected by the maximum number of covered base pairs (see coverage normalization in section "Materials and Methods").

accession or strain because the data available for that position and accession was not enough to detect it in a reliable way. Samples with low read coverage show inconsistency in SNP-to-read correlation (**Supplementary Figures S3–S5**). A reason for this behavior could be non-linear relation between number of SNPs and number of reads for very high and very low read numbers. Samples with a low number of reads can lead to incoherent SNP calling results due to stochastic coverage fluctuation. The high variability in such low read coverage samples can be observed in **Supplementary Figure S5**: the data range of Wi and Vx are wider than all the others. The low number of reads available for the Vx laser capture experiment (BioSample PRJNA602303) is probably related to the RNA-seq extraction technique, yielding small amounts of RNA that might be prone to bias before and/or after amplification.

To reduce the SNP per read effect, we normalized the SNPs by the coverage method, resulting in an observable increase of linear relationship (**Supplementary Figures S3**, **S4**). The number of SNPs called for each sample became more reliable in terms of comparability and reflect well previous studies and expectation of genetic distance coinciding with geographic distance (**Supplementary Figure S2**). The RNA-seq based SNP pipeline described here can in future be applied to stringently call SNPs for *P. patens* accessions and pedigrees, or can be adjusted to suit data sets from other model organisms for which a reference genome or transcriptome is available.

SNP Comparison Between Accessions

When locating the position of the SNPs in the genome, most of them were found in non-coding regions upstream and downstream the gene body (UTRs), as well as in introns and splicing sites within the introns. Many changes were observed in the coding sequences of the five accessions. These changes may lead to alterations in the protein sequence of the final gene product, by changing start or stop codons, or producing frame changes (**Table 1**).

The total number of filtered raw SNPs per accession (**Figure 3**) in comparison to the Gd genome reference shows (as expected) the Gd accessions as the one with the smallest number of changes followed by Re, Vx, Wi and Ka. This order agrees with the distance to the Gd geographical location in the Southeast of England (**Supplementary Figure S2**): Re (Hiss et al., 2017) and Vx (Kamisugi et al., 2008) in close vicinity to each other at the border of France and Southwestern Germany, and Wi and Ka (Perroud et al., 2011) in North America.

Results from Lang et al. (2018), where variance at genomic level was detected using the accessions Re, Vx and Ka, showed a SNP rate of one SNP per 1,783 bp for Re, per 644 bp for Ka and 188 bp for Vx while another study found a SNP rate of one SNP per 207 bp for Vx (Ding et al., 2018). Similar results for the number of base pairs per SNP can be found for the RNA-seq analysis in this study (Re 1 SNP each 1,912 bp, Ka 630 bp and Vx 143 bp) (**Supplementary Figure S5**). The SNP density based on RNA-seq (this study) and gDNA (Lang et al., 2018) is similar, although more SNPs are expected to be detected based on gDNA due to the presence of intergenic regions that are not under

selection. This could be another indication of an underestimated SNP number as discussed above. In any case, our method using RNA-seq data for gene space SNP calling yields appropriate results allowing to estimate differences in accessions by SNPs.

We have chosen two different methods to cluster the SNPs related to each sample. An artificial FASTA alignment with all SNPs as well as a matrix including SNPs and indels. Both methods show similar results (**Figure 4** and **Supplementary Figure S9**). The outlier sample Re_CI_3 has a very small read number, probably yielding misleading results. Sample Gd_CIV_1 also appears as an outlier (**Figure 4**). However, in the PCA 3D plot, the sample clusters according to expectation (**Supplementary Figure S9**). Our SNP pipeline had proven its functionality by blind tests as well as by pointing out unexpected metadata errors. The sample Ka_XVIII_1 was re-sequenced to replace a previous Gd experiment in which one of the triplicates failed (Perroud et al., 2018). For this sample, our SNP clustering (**Figure 4**) shows clear evidence for the accession being Ka, not Gd. Indeed, manual checking exclusive SNPs there is no doubt that it is Ka (**Supplementary File 3**, Sheet Ka_exclusive_SNPs). Most probably, the plant material was accidentally mislabeled.

The extraction of exclusive SNP sets for each of the five accessions helps to identify unknown *P. patens* sequences. Here we provide a set of SNPs for all examined accessions that will be useful for molecular identification of accessions. The low number of exclusive Vx SNPs are based on the uniqueness of the single Vx samples. Each Vx sample provided a big list of SNPs, but a high number of these SNPs were only available in one or two other Vx samples. A higher read coverage or more standardized mRNA could solve this issue. For low coverage reasons, we were not using the read coverage filter for the detection of exclusive SNPs. High sample support was chosen as an alternative and promoted exclusive SNP selection in a reasonable way, yielding confirmable molecular identification.

Observed approximate linearity between number of called SNPs and reads per sample (**Supplementary Figure S3**) lead to the read normalization method. When applying the coverage method that takes into account the fraction of the gene space covered by enough reads to allow SNP calling, linearity increased even further (**Supplementary Figure S4**). While both raw and normalized counts lead to the same conclusions in terms of genetic distance, we suggest the coverage normalization to most accurately describe the data.

SNP Comparison of Gransden Pedigrees

Gd is the current reference accession for *P. patens*, and was used to generate the genome sequence (Rensing et al., 2008; Lang et al., 2018). However, over the years of cultivation in the lab, it has shown an accumulation of somatic mutations which was confirmed in this study and observed before, culminating in observable phenotypic changes (Meyberg et al., 2020). One of the characteristics of laboratory models is the capacity to maintain the organism cultivated in the lab for multiple generations, being able to progress through the complete life cycle. The reduction of fertility of Gd accessions in the lab limits experimental design, especially when studying sexual reproduction or when the generation of off-spring is required for the experiments. For this reason, the accession Reute, which shows the lowest number of differences with the Gd genome reference, and which has a much higher fertility than Gd (Meyberg et al., 2020) has been proposed as an alternative to study sexual reproduction (Hiss et al., 2017; Meyberg et al., 2020).

Due to changes in land use, at the original Gransden collection site no *P. patens* can be found anymore. However, phenotypic data suggest that Gransden was not always infertile, because Gd_JP shows intermediate fertility between Re and extant Gd_DE pedigrees (Hiss et al., 2017; Meyberg et al., 2020). Our data show that, as expected, Gd_UK shows the lowest number of SNPs as compared to the reference genome that was derived from Gd_2004 (UK). All other pedigrees show substantial and unique SNPs (**Figure 5** and **Supplementary Figure S11**), demonstrating that during *in vitro* culture somatic mutations occur and accumulate in independent fashion. The practice of regular sexual reproduction of the cultured strains has the advantage that by this procedure it is ensured that the full life cycle can be followed. On top of that there is evidence that even during selfing *P. patens* is able to effectively purge deleterious mutations (Szövényi et al., 2017).

By comparing the normalized gene space SNP count of the Wi natural population samples with those of selfed progressions of Re and Gd laboratory strains we can estimate the genetic variability occurring in natural vs. laboratory samples (**Figure 7** and **Supplementary Figure S18**). Interestingly, the variation of three generations of homozygous (selfed) Re and Gd offspring is similar to that observed in naturally occurring Wi samples (representing the same generation but four spore capsules and five spores each). Based on the normalized data, the three generations of selfed laboratory cultures might even have acquired and retained slightly more mutations than visible in the single Wi natural population. We conclude that a substantial amount of genetic variation occurs both through somatic mutation during vegetative propagation (Meyberg et al., 2020) as well as during sexual propagation by selfing. However, since the practice of regular selfing selects for fertility it seems preferable to follow that practice over exclusive vegetative propagation.

Like for the accessions, specific SNPs for each pedigree were extracted. The diversity of Gd pedigrees is lower than that of the accessions and hence there were not enough samples supporting the same SNP. To detect exclusive SNPs for each pedigree ranking the SNPs by sample support gave us the opportunity to extract the SNPs supported by most of the samples. Obligatory for this method is a correct metadata grouping of the samples. If samples would be described to be the wrong pedigree, exclusive SNPs cannot be accurately determined. Another issue is the sub-clustering of samples. We can observe this for the Gd_JP pedigree as well as for Gd_UK. There are SNPs in the Japan pedigree that occurred in St. Louis, after it was brought to the USA. Our Gd_JP sample set is mostly represented by samples from the USA. Extracted exclusive SNPs with high sample support can thus be scored for the JP- > USA pedigree, but maybe not for the

full Gd_JP pedigree. Nevertheless, our provided exclusive SNP list can be used to classify the origin of unknown samples (**Supplementary Figure S17**).

Experimental Confirmation of Selected SNPs

In large experiments that handle many samples, mistakes might occur during the management of the samples in the lab, in the sequencing facility or during later data analysis. The identification of exclusive SNPs in the *P. patens* accessions allows the detection and correction of mistakes in experimental metadata, such as the ones mentioned earlier (**Figure 3**), *in silico*. Moreover, the exclusive SNPs found in the different accessions were used to identify unique targets for restriction enzymes, allowing the development of RFLP assays to differentiate between the *P. patens* accessions. The presence of the predicted SNPs in all tested sequences confirms the successful and stringent SNP selection presented here. The successful establishment of the RFLP analysis for the *P. patens* accessions provides a fast and cheap tool to test the accession background of laboratory strains as well as newly collected *P. patens* accessions. With regard to the Gd pedigrees so far, no SNPs within a restriction enzyme site with enough sample coverage could be identified. However, differentiation between Gd_DE, Gd_JP and Gd_CH could be performed successfully based on the sequencing data (**Supplementary Figure S17**). Thus, SNPs between the Gd pedigrees need to be analyzed via sequencing so far, but including more Gd data sets in the presented approach and/or analyzing a small subset of Gd pedigrees could help to improve and identify SNPs, which could be used within a future RFLP approach to differentiate Gd pedigrees.

Independent of the RFLP method, the origin of *P. patens* plant material can be discovered by using the presented primers (**Supplementary Tables S7, S8**) and sequencing the amplicon. If sequencing data is already available (single fragments, RNA-seq or gDNA sample[s]), our pipeline and the exclusive SNP sets can be used to easily identify plant origins.

Natural Population Variation and Selection

The number of observable mutations on the level of a naturally occurring population (Wi single spore isolates) is in the approximate same range as the mutations occurring in culture undergoing annual sexual reproduction (**Figure 7** and **Supplementary Figure S18**). For samples mainly propagated vegetatively, observed mutations are somatic in nature. For samples that regularly go through sexual reproduction, changes introduced via meiotic recombination cannot be distinguished from somatic changes. Intriguingly, the number of detected SNPs was found to decline over time in samples with a known heritage of regular sexual reproduction (**Supplementary Tables S12, S13**). We take this as evidence that sexual reproduction, even in a haploid, selfing species is able to efficiently purge deleterious mutations, as previously shown (Szövényi et al., 2017).

Consequently, the majority of the observed mutations probably are somatic. The observed mutation rates (changes per year and site) are in the range of 7E-07 to 4E-06 (**Supplementary Table S13**). Studies in other plants found rates in the E-08 range (Hanlon et al., 2019; Schoen and Schultz, 2019). The observed *P. patens* mutation rates are approximately two orders of magnitude higher than the estimated rate of synonymous substitutions per synonymous site per year, 9E-09 (Rensing et al., 2007). Hence, *in vitro* propagation of *P. patens* apparently leads to the fixation of a higher number of mutations than occur naturally, and maybe more than described in other plant propagation systems. Many labs perform regular shredding of protonemal tissue for propagation. This mode of propagation might increase the number of fixed somatic mutations via induction of the DNA repair system through cell damage, potentially resulting in higher mutational load.

The Ka/Ks ratio of the Gransden pedigree generally is below 1, suggesting potential negative (purifying) selection on many loci (**Supplementary Table S11**). All other accessions, to the contrary, exhibit ratios larger than 1, suggesting potential positive (Darwinian) selection. The latter is regardless of whether they are naturally occurring (Wi) or cultured (Ka, Re, Vx). Potentially, the decades-long vegetative culture of Gd, most of it vegetatively, led to the expression of negative selection. All other accessions are much more recent isolates and in particular all Re samples studied went through annual sexual reproduction, which apparently effectively purges deleterious mutations. Interestingly, the GO terms over-represented among those genes affected by non-synonymous changes (**Supplementary Figure S20**) include microtubule-based movement (Re) and reproduction (Vx), fitting recently published data that show these terms contrasted between male infertile Gd and fertile Re (Meyberg et al., 2020). It appears probable that the artificial environment of vegetative *in vitro* Gd propagation led to a loss of fertility due to loss of selection pressure on genes required for sexual reproduction.

CONCLUSION

Our study of sequence variants in *P. patens* laboratory strains revealed the accumulation of somatic mutations over years of cultivation, some of which can be detrimental e.g., with regard to fertility. It appears to be good practice to regularly let the lab cultures reproduce sexually, in order to keep selective pressure and to purge deleterious mutations. Since the original Gd accession is not available any more, and Gd JP shows less fertility than Re, it appears sensible to use Re (with its low number of SNPs as compared to Gd) for any studies that shall involve the life cycle. The identification of exclusive sets of SNPs for *P. patens* laboratory strains and accessions allowed the development of RFLP tests to identify the different accessions. Similarly, Gd pedigrees can be identified by sequencing of PCR products based on the pedigree-exclusive SNPs determined in this study. The variation of selfed laboratory strains is on the same order of magnitude as that of a natural population analyzed.

DATA AVAILABILITY STATEMENT

All RNA-seq samples used in this study are available via the NCBI SRA. Please see **Supplementary Table S1** in Supplementary.pdf for more details.

AUTHOR CONTRIBUTIONS

FH analyzed the raw read data and performed SNP calling. DS-M, JL, P-FP, and RM contributed RNA-seq data. FH, NS, and RM setup and performed RFLP as well as sequencing analyses. SR conceived of the study and supervised it together with NF-P and P-FP. FH, NF-P, RM, and SR wrote the manuscript with the help of all authors.

FUNDING

The JGI Plant Gene Atlas project conducted by the US Department of Energy Joint Genome Institute was supported by the Office of Science of the US Department of Energy under contract no. DE−AC02−05CH11231. SR is grateful for funding by Deutsche Forschungsgemeinschaft (Grant Number DFG RE 1697/15-1). DS-M was supported by ERC Advanced Grant (EDIP) and BBSRC grant BB/M020517/1 to JL.

ACKNOWLEDGMENTS

We are grateful to Ralph S. Quatrano and to Monsanto (St. Louis, MO, United States) for sequencing RNA-seq data of *P. patens* accession Kaskaskia, to Rafael Medina for samples of the *P. patens* accession Wisconsin, and to Stuart McDaniel for helpful comments. We also thank A. G. Goesmann, Justus-Liebig-University, Giessen for access to the de.NBI infrastructure.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00813/full#supplementary-material

REFERENCES

Ashton, N. W., and Cove, D. J. (1977). The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, *Physcomitrella patens*. *Mol. Gen. Genet.* 154, 87–95. doi: 10.1007/bf00265581

Ashton, N. W., and Raju, M. V. S. (2000). The distribution of gametangia on gametophores of *Physcomitrella* (Aphanoregma) patens in culture. *J. Bryol.* 22, 9–12. doi: 10.1179/jbr.2000.22.1.9

Beike, A. K., von Stackelberg, M., Schallenberg-Rüdinger, M., Hanke, S. T., Follo, M., Quandt, D., et al. (2014). Molecular evidence for convergent evolution and allopolyploid speciation within the *Physcomitrium*-Physcomitrellaspecies complex. *BMC Evol. Biol.* 14:158. doi: 10.1186/1471-2148-14-158

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodological).* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England).* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila* melanogaster strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics (Oxford, England).* 33, 2938–2940. doi: 10.1093/bioinformatics/btx364

Cove, D. (2005). The moss *Physcomitrella patens*. *Annu. Rev. Genet.* 39, 339–358.

Cove, D. J., Perroud, P. F., Charron, A. J., McDaniel, S. F., Khandelwal, A., and Quatrano, R. S. (2009). Isolation of DNA, RNA, and protein from the moss *Physcomitrella patens* gametophytes. *Cold Spring Harb. Protoc.* 2009:db.rot5146.

de Vries, J., and Rensing, S. A. (2020). Gene gains paved the path to land. *Nat. Plants* 6, 7–8. doi: 10.1038/s41477-019-0579-5

Demko, V., Perroud, P.-F., Johansen, W., Delwiche, C. F., Cooper, E. D., Remme, P., et al. (2014). Genetic analysis of DEFECTIVE KERNEL1 loop function in three-dimensional body patterning in *Physcomitrella patens*. *Plant Physiol.* 166, 903–919. doi: 10.1104/pp.114.243758

Ding, X., Pervere, L. M., Bascom, C. Jr., Bibeau, J. P., Khurana, S., Butt, A. M., et al. (2018). Conditional genetic screen in *Physcomitrella patens* reveals a novel microtubule depolymerizing-end-tracking protein. *PLoS Genet.* 14:e1007221. doi: 10.1371/journal.pgen.1007221

Engel, P. P. (1968). The induction of biochemical and morphological mutants in the moss *Physcomitrella patens*. *Am. J. Bot.* 55, 438–446. doi: 10.1002/j.1537-2197.1968.tb07397.x

Fernandez-Pozo, N., Haas, F. B., Meyberg, R., Ullrich, K. K., Hiss, M., Perroud, P.-F., et al. (2019). PEATmoss (*Physcomitrella* expression atlas tool): a unified gene expression atlas for the model plant *Physcomitrella patens*. *Plant J.* 102, 165–177. doi: 10.1111/tpj.14607

Flowers, J. M., Hazzouri, K. M., Pham, G. M., Rosas, U., Bahmani, T., Khraiwesh, B., et al. (2015). Whole-genome resequencing reveals extensive natural variation in the model green Alga *Chlamydomonas reinhardtii*. *Plant Cell* 27, 2353–2369. doi: 10.1105/tpc.15.00492

Frank, M. H., and Scanlon, M. J. (2015). Cell-specific transcriptomic analyses of three-dimensional shoot development in the moss *Physcomitrella patens*. *Plant J.* 83, 743–751. doi: 10.1111/tpj.12928

Hanlon, V. C. T., Otto, S. P., and Aitken, S. N. (2019). Somatic mutations substantially increase the per-generation mutation rate in the conifer *Picea sitchensis*. *Evol. Lett.* 3, 348–358. doi: 10.1002/evl3.121

Hiss, M., Meyberg, R., Westermann, J., Haas, F. B., Schneider, L., Schallenberg-Rdinger, M., et al. (2017). Sexual reproduction, sporophyte development and molecular variation in the model moss *Physcomitrella patens*: introducing the ecotype Reute. *Plant J. Cell Mol. Biol.* 90, 606–620. doi: 10.1111/tpj.13501

Huson, D. H., and Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Kamisugi, Y., Von Stackelberg, M., Lang, D., Care, M., Reski, R., Rensing, S. A., et al. (2008). A sequence-anchored genetic linkage map for the moss, *Physcomitrella patens*. *Plant J.* 56, 855–866. doi: 10.1111/j.1365-313x.2008.03637.x

Krasovec, M., Eyre-Walker, A., Sanchez-Ferandin, S., and Piganeau, G. (2017). Spontaneous mutation rate in the smallest photosynthetic eukaryotes. *Mol. Biol. Evol.* 34, 1770–1779. doi: 10.1093/molbev/msx119

Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J. Cell Mol. Biol.* 93, 515—-533.

Leaché, A. D., and Oaks, J. R. (2017). The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 48, 69–84. doi: 10.1146/annurev-ecolsys-110316-022645

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Ma, X., Shao, Y., Tian, L., Flasch, D. A., Mulder, H. L., Edmonson, M. N., et al. (2019). Analysis of error profiles in deep next-generation sequencing data. *Genome Biol.* 20:50.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Medina, R., Johnson, M. G., Liu, Y., Wickett, N. J., Shaw, A. J., and Goffinet, B. (2019). Phylogenomic delineation of *Physcomitrium* (Bryophyta: Funariaceae) based on targeted sequencing of nuclear exons and their flanking regions rejects the retention of *Physcomitrella*, *Physcomitridium* and *Aphanorrhegma*. *J. Syst. Evol.* 57, 404–417. doi: 10.1111/jse.12516

Meyberg, R., Perroud, P.-F., Haas, F. B., Schneider, L., Heimerl, T., Renzaglia, K. S., et al. (2020). Characterization of evolutionarily conserved key players affecting eukaryotic flagellar motility and fertility using a moss model. *New Phytol.* doi: 10.1111/nph.16486 [Epub ahead of print].

Moody, L. A., Kelly, S., Rabbinowitsch, E., and Langdale, J. A. (2018). Genetic regulation of the 2D to 3D growth transition in the moss *Physcomitrella patens*. *Curr. Biol.* 28, 473–478.e5. doi: 10.1016/j.cub.2017.12.052

Nguyen, T.-P., Muhlich, C., Mohammadin, S., van den Bergh, E., Platts, A. E., Haas, F. B., et al. (2019). Genome improvement and genetic map construction for aethionema arabicum, the first divergent branch in the Brassicaceae family. *G3 (Bethesda, Md).* 9, 3521–3530. doi: 10.1534/g3.119.400657

Niu, S., Song, Q., Koiwa, H., Qiao, D., Zhao, D., Chen, Z., et al. (2019). Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biol.* 19:328. doi: 10.1186/s12870-019-1917-5

Perroud, P.-F., Cove, D. J., Quatrano, R. S., and McDaniel, S. F. (2011). An experimental method to facilitate the identification of hybrid sporophytes in the moss *Physcomitrella patens* using fluorescent tagged lines. *New Phytol.* 191, 301–306. doi: 10.1111/j.1469-8137.2011.03668.x

Perroud, P.-F., Haas, F. B., Hiss, M., Ullrich, K. K., Alboresi, A., Amirebrahimi, M., et al. (2018). The *Physcomitrella patens* gene atlas project: large scale RNA-seq based expression data. *Plant J.* 95, 168–182. doi: 10.1111/tpj.13940

Perroud, P.-F., Meyberg, R., and Rensing, S. A. (2019). *Physcomitrella patens* reute mCherry as a tool for efficient crossing within and between ecotypes. *Plant Biol.* 21, 143–149. doi: 10.1111/plb.12840

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England).* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rensing, S. A. (2018). Great moments in evolution: the conquest of land by plants. *Curr. Opin. Plant Biol.* 42, 49–54. doi: 10.1016/j.pbi.2018.02.006

Rensing, S. A., Beike, A. K., and Lang, D. (2013). "Evolutionary importance of generative polyploidy for genome evolution of haploid-dominant land plants," in *Plant Genome Diversity : Physical Structure, Behaviour and Evolution of Plant Genomes*, Vol. 2, eds I. J. Leitch, J. Greilhuber, D. Jaroslav, and W. Jonathan (Vienna: Springer-Verlag), 295–305. doi: 10.1007/978-3-7091-1160-4_18

Rensing, S. A., Goffinet, B., Meyberg, R., Wu, S.-Z., and Bezanilla, M. (2020). The moss *Physcomitrium* (*Physcomitrella*) *patens*: a model organism for non-seed plants. *Plant Cell* 32, 1361–1376. doi: 10.1105/tpc.19.00828

Rensing, S. A., Ick, J., Fawcett, J. A., Lang, D., Zimmer, A., Van de Peer, Y., et al. (2007). An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol. Biol.* 7:130. doi: 10.1186/1471-2148-7-130

Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69.

Ribeiro, A., Golicz, A., Hackett, C. A., Milne, I., Stephen, G., Marshall, D., et al. (2015). An investigation of causes of false positive single nucleotide polymorphisms using simulated reads from a small eukaryote genome. *BMC Bioinformatics* 16:382. doi: 10.1186/s12859-015-0801-z

Saint-Marcoux, D., Billoud, B., Langdale, J. A., and Charrier, B. (2015). Laser capture microdissection in *Ectocarpus siliculosus*: the pathway to cell-specific transcriptomics in brown algae. *Front. Plant Sci.* 6:54. doi: 10.3389/fpls.2015.00054

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026

Schoen, D. J., and Schultz, S. T. (2019). Somatic mutation and evolution in plants. *Annu. Rev. Ecol. Evol. Syst.* 50, 49–73. doi: 10.1146/annurev-ecolsys-110218-024955

Schween, G., Schulte, J., Reski, R., and Hohe, A. (2005). Effect of ploidy level on growth, differentiation, and morphology in *Physcomitrella patens*. *Bryologist.* 108, 27–35. doi: 10.1639/0007-2745(2005)108[27:eoplog]2.0.co;2

Smits, W. K. (2017). SNP-ing out the differences: investigating differences between *Clostridium* difficile lab strains. *Virulence* 8, 613–617. doi: 10.1080/21505594.2016.1250998

Stevenson, S. R., Kamisugi, Y., Trinh, C. H., Schmutz, J., Jenkins, J. W., Grimwood, J., et al. (2016). Genetic analysis of *Physcomitrella patens* identifies Abscisic acid non-responsive, a Regulator of ABA responses unique to basal land plants and required for desiccation tolerance. *Plant Cell* 28, 1310–1327.

Szövényi, P., Ullrich, K. K., Rensing, S. A., Lang, D., van Gessel, N., Stenøien, H. K., et al. (2017). Selfing in haploid plants and efficacy of selection: codon usage bias in the model moss *Physcomitrella patens*. *Genome Biol. Evol.* 9, 1528–1546. doi: 10.1093/gbe/evx098

Vashisht, D., Hesselink, A., Pierik, R., Ammerlaan, J. M., Bailey-Serres, J., Visser, E. J., et al. (2011). Natural variation of submergence tolerance among *Arabidopsis thaliana* accessions. *New Phytol.* 190, 299–310. doi: 10.1111/j.1469-8137.2010.03552.x

von Stackelberg, M., Rensing, S. A., and Reski, R. (2006). Identification of genic moss SSR markers and a comparative analysis of twenty-four algal and plant gene indices reveal species-specific rather than group-specific characteristics of microsatellites. *BMC Plant Biol.* 6:9. doi: 10.1186/1471-2229-6-9

Wang, S., Meyer, E., McKay, J. K., and Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat. Methods* 9, 808–810. doi: 10.1038/nmeth.2023

Widiez, T., Symeonidi, A., Luo, C., Lam, E., Lawton, M., and Rensing, S. A. (2014). The chromatin landscape of the moss *Physcomitrella patens* and its dynamics during development and drought stress. *Plant J.* 79, 67–81. doi: 10.1111/tpj.12542

Wu, T. D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26, 873–881. doi: 10.1093/bioinformatics/btq057

Xia, W., Luo, T., Zhang, W., Mason, A. S., Huang, D., Huang, X., et al. (2019). Development of high-density SNP markers and their application in evaluating genetic diversity and population structure in *Elaeis guineensis*. *Front. Plant Sci.* 10:130. doi: 10.3389/fpls.2019.00130

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 July 2020
doi: 10.3389/fpls.2020.01062

The Chloroplast Land Plant Phylogeny: Analyses Employing Better-Fitting Tree- and Site-Heterogeneous Composition Models

Filipe Sousa[1], *Peter Cíváň*[1,2], *Peter G. Foster*[3] *and Cymon J. Cox*[1]*

[1] Centro de Ciências do Mar, Universidade do Algarve, Faro, Portugal, [2] INRA, Université Clermont-Auvergne, Clermont-Ferrand, France, [3] Department of Life Sciences, Natural History Museum, London, United Kingdom

The colonization of land by descendants of charophyte green algae marked a turning point in Earth history that enabled the development of the diverse terrestrial ecosystems we see today. Early land plants diversified into three gametophyte-dominant lineages, namely the hornworts, liverworts, and mosses, collectively known as bryophytes, and a sporophyte-dominant lineage, the vascular plants, or tracheophytes. In recent decades, the prevailing view of evolutionary relationships among these four lineages has been that the tracheophytes were derived from a bryophyte ancestor. However, recent phylogenetic evidence has suggested that bryophytes are monophyletic, and thus that the first split among land plants gave rise to the lineages that today we recognize as the bryophytes and tracheophytes. We present a phylogenetic analysis of chloroplast protein-coding data that also supports the monophyly of bryophytes. This newly compiled data set consists of 83 chloroplast genes sampled across 30 taxa that include chlorophytes and charophytes, including four members of the Zygnematophyceae, and land plants, that were sampled following a balanced representation of the main bryophyte and tracheophyte lineages. Analyses of non-synonymous site nucleotide data and amino acid translation data result in congruent phylogenetic trees showing the monophyly of bryophytes, with the Zygnematophyceae as the charophyte group most closely related to land plants. Analyses showing that bryophytes and tracheophytes evolved separately from a common terrestrial ancestor have profound implications for the way we understand the evolution of plant life cycles on land and how we interpret the early land plant fossil record.

Keywords: land plants, phylogeny, bryophytes, chloroplast, composition heterogeneity

OPEN ACCESS

Edited by:
Stefan A. Rensing,
University of Marburg, Germany

Reviewed by:
Sònia Garcia,
Botanical Institute of Barcelona, Spain
Domingos Cardoso,
Universidade Federal da Bahia, Brazil
Mark Puttick,
University of Bath, United Kingdom

***Correspondence:**
Cymon J. Cox
cymon.cox@googlemail.com

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *19 February 2020*
Accepted: *26 June 2020*
Published: *10 July 2020*

Citation:
Sousa F, Civáň P, Foster PG and Cox CJ (2020) The Chloroplast Land Plant Phylogeny: Analyses Employing Better-Fitting Tree- and Site-Heterogeneous Composition Models. Front. Plant Sci. 11:1062. doi: 10.3389/fpls.2020.01062

INTRODUCTION

It is widely accepted that land plants, or embryophytes, descend from an aquatic green algal ancestor (Karol, 2001; McCourt et al., 2004) that colonized land over 450 Mya (Magallón et al., 2013; Morris et al., 2018), however, reconstructing the relationships among the bryophytes (liverworts, hornworts, and mosses) and tracheophytes (lycopods, ferns, and seed plants), and identifying the algal lineage that is most closely related to the embryophytes, has been challenging and controversial (Cox, 2018). These six

major land plant lineages, as well as the six major streptophyte algal groups (Klebsormidales, Chlorokybales, Mesostigmales, Coleochaetales, Charales, and Zygnematales) are each typically well-supported clades and considered monophyletic natural groups. Relationships among the streptophyte algae have been determined with increasing congruence and statistical confidence, converging on a phylogeny that places the conjugating algae of the Zygnematophyceae as the sister-group of land plants (Wickett et al., 2014; Puttick et al., 2018). Among the land plants, the monophyly of the tracheophytes is well supported by molecular evidence and has been assumed partly due to their common possession of an elaborate vascular system, although it is now known that the water-conducting cells of bryophytes are homologous to those of tracheophytes and governed by a similar developmental system (Xu et al., 2014). By contrast, a common origin of the three bryophyte groups, independent of the tracheophytes, has not previously been considered likely, with the majority of studies showing that the tracheophytes evolved from the bryophytes after their initial diversification. Indeed, phylogenetic inferences of sequence data from the nuclear, plastid, and mitochondrial genomes have resulted in conflicting yet statistically well-supported topologies of land plant relationships showing that either the liverworts (e.g. Lewis et al., 1997; Gao et al., 2010), the hornworts (e.g. Hedderson et al., 1996; Wickett et al., 2014), or the clade Setaphyta (Puttick et al., 2018), that contains mosses plus liverworts (e.g. Nishiyama and Kato, 1999; Karol et al., 2010), were the first lineage to split from the remaining land plants. However, in recent years, several studies have supported a hypothesis whereby the first divergence of land plants was between bryophytes and tracheophytes, ruling out a direct descendance of the tracheophytes from bryophytes, and having profound implications for how we view the evolution of plants on land. These newer studies have used better-fitting models that more accurately account for heterogeneity in the data, and therefore suggest that previous hypotheses were based on overly simplistic analyses (Cox et al., 2014; Puttick et al., 2018; Sousa et al., 2019).

Incongruence among phylogenetic tree topologies can be attributed to biological processes, such as incomplete lineage sorting (ILS) and hybridization, and methodological issues, such as inappropriate choice of substitution models. In the case of the land plant phylogeny, however, two main evolutionary processes underlie the observed inconsistency of phylogenetic inferences. Firstly, given the large geological timescale over which land plants have evolved, nucleotide data are subject to substitutional "saturation" at synonymous codon sites, that are under low selective pressure since they do not change the amino acid sequence. Over time, multiple substitutions can occur on synonymous sites, to an extent that they no longer carry reliable phylogenetic signal (Jeffroy et al., 2006). In such cases, the exclusion or recoding of synonymous sites is necessary to remove the non-phylogenetic signal (Cox et al., 2014; Sousa et al., 2019). Secondly, sequence data from highly divergent lineages often display compositional heterogeneity, meaning that the long-term probability of change to a particular nucleotide or amino-acid is different among sites or lineages. Consequently, commonly used substitution models, that assume a fixed nucleotide or amino acid composition among all sites and lineages, may lead to erroneous phylogenetic inference if the data are composition heterogeneous (Foster, 2004). Both composition site- and tree-heterogeneity are the result of varying mutational pressures or selection (for example, for high GC content) and may result in a high level of homoplasy. Composition site-heterogeneity can be modeled using mixture models such as the CAT model (Lartillot and Philippe, 2004), whereas composition tree-heterogeneity can be modeled with non-stationary models such as the NDCH model (Foster, 2004; Foster et al., 2009).

In this study, we reassess the support for land plant relationships based on a newly compiled data set of 83 chloroplast protein-coding genes. Chloroplast sequence data typically represents a single linkage group, since chloroplasts are usually inherited uniparentally as a circular non-recombining chromosome, resulting in reduced opportunities for recombination between different chloroplast lineages (Birky, 2001). There are also no documented cases of lateral gene transfer between chloroplast genomes (Bock, 2010). Thus, there is a reasonable expectation that all genes in the chloroplast genome should carry phylogenetic signal supporting the same tree, i.e. the whole chloroplast genome tree is effectively a gene tree which may or may not be congruent with the species tree, and incongruence among trees inferred from individual chloroplast genes is likely the result of systematic error, rather than ILS. The concatenation of chloroplast genes for phylogenetic analyses is therefore justified, and the resulting tree is analogous to a tree reconstructed from a single non-recombining nuclear DNA sequence. However, as in nuclear genomes, chloroplast protein-coding genes are also subject to composition biases due to drift and different mutational pressures, and thus appropriate modeling of composition site- and tree-heterogeneity is warranted for phylogenetic reconstruction from highly-divergent chloroplast sequences.

Our reconstruction of the land plant phylogeny based on codon-degenerated (non-synonymous) nucleotide data and amino acid data, under better-fitting composition tree-heterogeneous (non-stationary) models, result in trees where bryophytes are monophyletic, strengthening the hypothesis presented by Cox et al. (2014). These new analyses, together with published analyses of nuclear protein coding data (Puttick et al., 2018; Sousa et al., 2019) support the hypothesis whereby the first evolutionary split among land plants occurred between the bryophytes and the tracheophytes, and suggests a need for a re-interpretation of the fossil evidence and the nature of the ancestral embryophyte.

MATERIALS AND METHODS

The thirty taxa selected for analyses include four chlorophyte algae, nine charophyte algae, of which four are members of the Zygnematophyceae, six bryophytes, sampled evenly among liverworts, mosses, and hornworts, and 11 tracheophytes, including representatives of lycopods, ferns, and seed plants (**Table 1**). Protein-coding genes which were annotated in at least 15 of the sampled taxa were selected for analysis, resulting in a data set of 83 genes (**Supplementary Information Table**

S1). Individual nucleotide alignments and the respective amino acid translation were constructed using TranslatorX (Abascal et al., 2010), and poorly aligned regions were identified using GBlocks (vers. 0.91b; Talavera and Castrasana, 2007). Alignments were inspected manually, and regions of low coverage, i.e., at the beginning and ends of sequences, or with ambiguous alignment, were identified and removed by codon triplet position, to maintain a full correspondence between codon triplets of the nucleotide sequences and their amino acid translation. Concatenated data matrices were constructed from the combined protein-coding genes (48861 sites) and their corresponding combined protein translations (16287 sites). The proportion of missing characters among ingroup taxa were very low, with a mean of 4.38% per taxon (median 2.36%), suggesting that the results were unlikely to be biased by ambiguous data (Lemmon et al., 2009). In addition to standard DNA coding, all synonymous substitutions of the protein-coding gene data were eliminated by codon-degenerate recoding with IUPAC ambiguity codes (Cox et al., 2014). Thus, three concatenated data sets were generated: 1) nucleotides, 2) codon-degenerate recoded nucleotides, and 3) the translated amino acid sequences.

Three tree-independent tests of model process homogeneity were performed using pairwise sequence comparisons in each of the three data sets to assess whether the data were homogeneous with respect to among-lineage composition (i.e. stationarity) and instantaneous substitution rate, and process reversibility. Bowker's Test (Ababneh et al., 2006; Jermiin et al., 2017) is a general test of model process homogeneity between sequences, whereas Stuart's and Ababneh's tests indicate deviation from stationarity and rate homogeneity, respectively (Ababneh et al., 2006; Jermiin et al., 2020; Jermiin and Misof, 2020). All tests were performed using P4 (vers. 0.89 - Foster, 2004).

Optimal sets of partitions among genes (11 partitions) and among codon-positions in genes (21 partitions) were determined using PartitionFinder (Lanfear et al., 2014), using a general time-reversible (GTR) model with a discrete (4 categories) gamma-distribution of rates among sites (Γ_4), with empirical base frequencies (F_{emp}), and with the best partitioning schemes chosen using the Bayesian Information Criterion (BIC). To test whether the optimal gene partitioning scheme estimated by PartitionFinder was dependent on the estimated neighbor-joining starting tree, which by default resulted in a tree in which hornworts were nested in the tracheophytes and is likely incorrect, an alternative optimal gene partitioning scheme, contingent on a fixed tree showing monophyletic bryophytes, was determined and analyzed by ML bootstrap.

Best-fitting substitution models were determined using Modelgenerator (Keane et al., 2006). In addition, the green-plant specific empirical amino-acid substitution model, gcpREV, was used for analyses of amino acid sequence data (Cox and

TABLE 1 | Taxon sampling.

Taxon name	Classification[1]	GenBank Accession	No. of genes[2]	% Missing characters[3]	% G-C
Chlorella vulgaris	Chlorophyta, Trebouxiophyceae	NC_001865	65	21.28	38.24
Chlamydomonas reinhardtii	Chlorophyta, Chlorophyceae	NC_005353	57	32.39	36.30
Ostreococcus tauri	Chlorophyta, prasinophytes	NC_008289	54	35.56	42.02
Nephroselmis olivacea	Chlorophyta, prasinophytes	NC_000927	74	4.81	43.13
Mesostigma viride	Streptophyta, Mesostigmales	NC_002186	79	3.55	33.57
Chlorokybus atmophyticus	Streptophyta, Chlorokybales	NC_008822	81	1.66	37.96
Klebsormidium flaccidum	Streptophyta, Klebsormidiales	NC_024167	73	7.23	43.33
Chara vulgaris	Streptophyta, Charales	NC_008097	81	1.34	34.63
Chaetosphaeridium globosum	Streptophyta, Coleochaetales	NC_004115	83	0.09	33.77
Staurastrum punctulatum	Streptophyta, Desmidiales	NC_008116	81	1.22	35.77
Zygnema circumcarinatum	Streptophyta, Zygnematales	NC_008117	81	0.90	37.93
Mesotaenium endlicherianum	Streptophyta, Zygnematales	NC_024169	81	0.74	44.29
Roya anglica	Streptophyta, Zygnematales	NC_024168	81	0.47	36.63
Pellia endiviifolia	Streptophyta, Marchantiophyta	NC_019628	82	0.50	38.24
Ptilidium pulcherrimum	Streptophyta, Marchantiophyta	NC_015402	77	10.7	35.72
Physcomitrella patens	Streptophyta, Bryophyta	NC_005087	80	2.84	33.46
Syntrichia ruralis	Streptophyta, Bryophyta	NC_012052	77	9.90	33.21
Nothoceros aenigmaticus	Streptophyta, Anthocerotophyta	NC_020259	81	2.80	39.10
Anthoceros formosae	Streptophyta, Anthocerotophyta	NC_004543	81	1.92	37.31
Isoetes flaccida	Streptophyta, Lycopodiophyta	NC_014675	79	3.83	40.75
Huperzia lucidula	Streptophyta, Lycopodiophyta	NC_006861	83	0.03	38.98
Selaginella moellendorffii	Streptophyta, Lycopodiophyta	NC_013086	66	10.12	50.77
Equisetum hyemale	Streptophyta, Monililformopses	NC_020146	81	0.52	36.02
Psilotum nudum	Streptophyta, Monililformopses	KC117179	79	7.68	38.57
Angiopteris evecta	Streptophyta, Monililformopses	NC_008829	83	0.01	38.01
Adiantum capillus-veneris	Streptophyta, Monililformopses	NC_004766	79	7.31	43.42
Pinus thunbergii	Streptophyta, Spermatophyta	NC_001631	69	20.61	40.65
Cycas revolute	Streptophyta, Spermatophyta	JN867588	81	1.08	40.72
Arabidopsis thaliana	Streptophyta, Spermatophyta	NC_000932	76	8.70	39.04
Nymphaea alba	Streptophyta, Spermatophyta	NC_006050	77	8.20	41.01

Taxon names, classification, NCBI GenBank accession numbers, and numbers of genes present in data set. [1]Classification follows NCBI GenBank. The given ranks are not equal but the highest available while distinguishing among the six major lineage of land plants and the major groups of algae. [2]Number of gene present out of a total of 83. [3]Percentage of missing characters in gene sequence and protein data matrices.

Foster, 2013). Maximum-likelihood (ML) bootstrap analyses were conducted using an MPI-compiled version of RAxML (vers. 7.0.4–7.8.4–8.0.26; Stamatakis, 2006). RAxML analyses consisted of 300 or 400 bootstrap replicates with default settings for parameter estimation accuracy, a discrete gamma-distribution of among-site rate heterogeneity (4 categories; Γ_4) and estimated composition frequencies (F_{est}).

Bayesian Markov Chain Monte-Carlo (MCMC) analyses were performed using P4 with the NDCH and NDCH2 non-stationary composition models (Cox et al., 2008). Homogeneous (stationary) analyses were performed by defining a single composition vector on the NDCH model (CV1). Composition tree-heterogeneous analyses on the protein data were performed using the NDCH2 model which includes a separate composition vector for each node of the tree. Fit of the model composition to the data was determined using posterior predictive simulations of the χ^2 statistic of composition homogeneity as implemented in P4 (Foster, 2004). Indicators of poor MCMC performance — low acceptance rates, poor mixing between hot and cold chains, excessively long branch lengths (Brown et al., 2010) were noted. MCMC analyses were also performed using Phylobayes MPI (vers. 1.2f — Lartillot and Philippe, 2004) with the CAT infinite profile mixture model (F_{CAT}), which specifically handles composition site-heterogeneity. Posterior predictive tests were applied to Phylobayes analyses to assess model-fit.

Stationarity of MCMC chains was assessed by observing the likelihood of samples (and other parameters) over time, and convergence to the correct posterior probability distribution was determined by running multiple MCMC chains in parallel and calculating the average standard deviation of split support (asdoss) between independent chains. Posterior probabilities (PP) < 0.95 and bootstrap values (BS) < 90% were considered low and indicative of weak support of nodes, whereas larger values were considered strong indicators of clade support. Details of individual analyses, the specific models used, and the diagnostic statistics are included in the legends of the figures in the Supporting Information. The combined nucleotide, codon-degenerate and protein matrices, all in nexus format and with characters sets, were deposited on Zenodo (DOI: 10.5281/zenodo.3886964).

RESULTS

Matched-Pairs Tests of Process Homogeneity

In **Figure 1**, the plotted p-values for the matched-pairs tests of homogeneity for each of the nucleotide, codon-degenerated, and amino acid data sets are shown. All three data sets fail all three tests, although the assumption of model homogeneity is violated more severely in the nucleotide data sets than in the amino acids data. These tests indicate that the data are neither stationary with respect to composition (composition varies across lineages) or homogeneous with respect to instantaneous rates (rates vary across lineages).

Nucleotide Data

All ML bootstrap analyses of the protein-coding nucleotide data ($GTR+\Gamma_4+F_{est}$) strongly support the placement of the moss lineage as sister-group to all other plants (BS>90%), with the hornworts fully supported as the sister-group to the tracheophytes (**Figure 2A**;

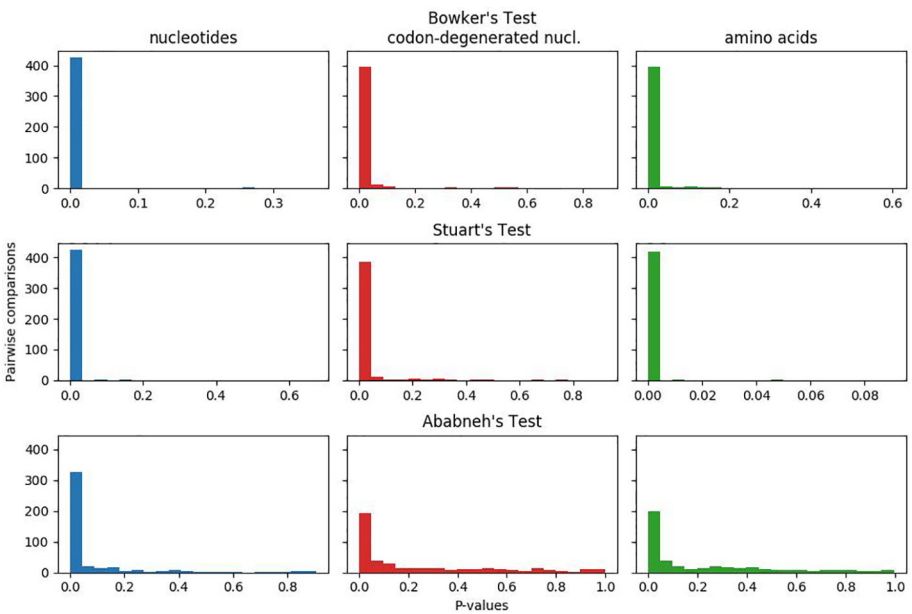


FIGURE 1 | Plots of p-values for Bowker's, Stuart's, and Ababneh's matched-pairs tests of model homogeneity for each of the three data sets. Numbers of rejected (p-values < 0.05) tests: Bowker's test nucleotides 427 (98%), codon-degenerated nucl. 398 (91%), amino acids 401 (92%); Stuart's test nucleotides 424 (97%), codon-degenerated nucl. 387 (89%), amino acids 401 (98%); Ababneh's test nucleotides 331 (76%), codon-degenerated nucl. 193 (44%), amino acids 126 (29%).

Figures S1–S3). ML bootstrap analyses with optimal numbers of gene partitions (11 partitions with separate models; **Figure S1**) did not result in topological differences compared to the non-partitioned ML bootstrap analysis (**Figure 2A**), and the use of an alternative starting tree for estimating the optimal gene partitioning scheme resulted in a slightly altered partitioning scheme but ultimately had no substantial effect on the statistical support regarding the placement of bryophyte lineages (**Figure S2**). The ML bootstrap analyses with optimal numbers of codon-position partitions (21 partitions; **Figure S3**) was also congruent with other analyses regarding the placement of bryophytes, but resulted in a different arrangement among tracheophyte lineages. Whereas the non-partitioned and gene-partitioned analyses placed ferns as sister-group to other tracheophytes in the codon-position and partitioned analyses, the lycopods appear as sister-group to other tracheophytes in the codon-partitioned analyses (**Figure S3**).

Bayesian MCMC analyses of the nucleotide data using a tree-homogeneous composition model shows full support (PP = 1.0) for the placement of mosses as sister-group to other land plants and hornworts as the sister-group to tracheophytes (**Figure S4**). Similarly, tree-heterogeneous composition model analyses (F_{NDCH2}) resulted in a similar topology and support values, but with the lycopods as the sister-group to the remaining tracheophytes (**Figure S5**). Posterior predictive simulations of the χ^2 statistic of composition homogeneity showed a poor fit ($p = 0.0$) of both the tree-homogeneous composition model (F_{CV1}) and the tree-heterogeneous composition (F_{NDCH2}) model, but the latter was a much improved fit by 2 orders of magnitude (see legends of **Figures S4** and **S5** for details). Site-heterogeneous composition model analyses using the Phylobayes ($GTR+\Gamma_4+F_{CAT}$) also placed the mosses as the sister-group to the other land plants, but with low branch support (PP = 0.84), whereas hornworts remained strongly supported as the sister-group to tracheophytes (PP = 1.0; **Figure S6**).

The ML bootstrap analyses of the codon-degenerate recoded nucleotide data set ($GTR+\Gamma_4+F_{est}$) did not resolve the position of either mosses or liverworts, and resolved hornworts as sister-group to tracheophytes with low branch support (BS = 76%; **Figure 2B**).

Amino Acid Data

ML bootstrap analyses of the concatenated amino acid data ($gcpREV+\Gamma_4+F_{mod}$) resolve bryophytes as monophyletic (BS = 77%) but fail to recover the monophyly of tracheophytes, showing ferns as the sister-group to the remaining embryophytes but with very low statistic support (BS = 56%; **Figure S7**). When the data were divided into 17 partitions, ML bootstrap support for the monophyly of the bryophytes increased to 81%, and the ferns were supported as the sister-group to all other land plants by 66% (**Figure S8**).

MCMC analyses of amino acid data under a tree-homogeneous composition model ($gcpREV+\Gamma_4+F_{est}$; **Figure S9**) and under the tree-heterogeneous composition model ($gcpREV+\Gamma_4+F_{NDCH2}$; **Figures 3A, B**; **Figures S10** and **11**) both show bryophytes as monophyletic with maximum support (PP = 1.0) in all replicates of the analyses. However, the four independent runs of non-stationary composition (F_{NDCH2}) analyses failed to converge on the same topology with respect to the relationships among the tracheophyte lineages. In two runs (runs 1 and 3, **Figures 3A** and **S10**, respectively), including the run with the best marginal-likelihood

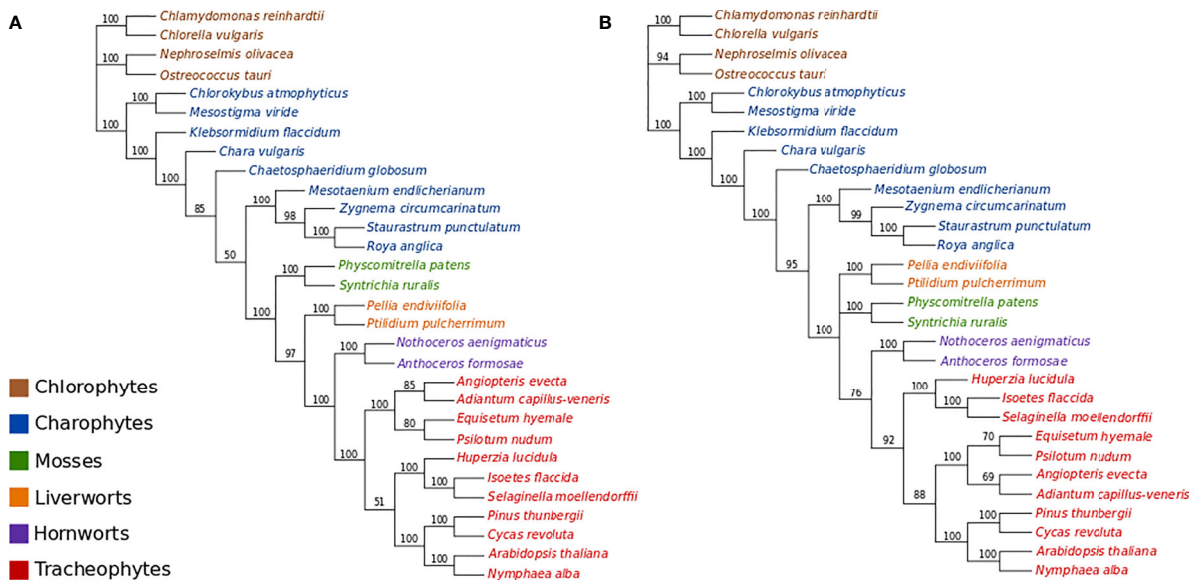


FIGURE 2 | (A) ML bootstrap analysis (400 replicates) of the nucleotide data ($GTR+\Gamma_4$). Optimal ML tree score: -ln L= 757308.3933. Length of optimal tree = 10.0 substitutions/site. **(B)** ML bootstrap analyses (400 replicates) of the codon-degenerate recoded nucleotide data ($GTR+\Gamma_4$). Optimal ML tree score: -ln L= 304862.2078. Optimal tree length = 2.7 substitutions/site.

score (run 1), tracheophytes were resolved as paraphyletic, with ferns placed as sister-group to the remaining embryophytes and lycopods as sister-group to the bryophyte clade. The two other runs (runs 2 and 4, **Figures 3B** and **S11**, respectively) recovered tracheophytes as monophyletic, with lycopods as the sister-group to the clade containing ferns and seed plants. All nodes on the trees obtained from every run received maximum support (PP = 1.0). Neither the tree-homogeneous or the tree-heterogeneous (F_{NDCH2}) composition model fit the data, according to posterior predictive simulations of χ^2, but the NDCH2 model was a much better approximation than the homogeneous model as the test statistic fell within the sample distribution of the runs, albeit outside the 95% confidence interval (see legends of **Figures S10** and **S11** for details).

The four independent MCMC analyses of the amino acid data with the site-heterogeneous composition model (GTR+G+F_{CAT}) resulted in trees showing the clade Setaphyta as the sister-group to the remaining land plants (PP = 0.96–0.98) and hornworts as the sister-group of tracheophytes (**Figures S12–S15**). Posterior predictive tests of the four runs showed that all but one run passed the site diversity test that estimates the fit of the model to describe the mean number of distinct amino acids per site. However, the null hypothesis was rejected (i.e. the model does not fit the data adequately) by posterior predictive simulations in other tests: a) the empirical convergence probability test which estimates the long-term probability of two sites converging on the same character state in two random taxa; b) across-site compositional heterogeneity test; c) across-taxa maximum heterogeneity test; and d) across-taxa mean squared heterogeneity test. Additional MCMC runs on amino acid data with constant sites removed, previously thought to influence tree topology in these analyses, did not show differences in topology or branch support (not shown). A summary of bryophyte relationships obtained from each data type, and each type of analysis is presented in **Figure 4**.

The conjugating algae group Zygnematophyceae was resolved as sister-group to land plants in all but one analysis where the sister-group to land plants was not resolved (**Figure S2**). When resolved, this relationship received high to maximum branch support (BS between 80% and 100%, PP > 0.95) except for analyses of nucleotide data with composition tree-homogeneous models (ML, **Figures S1** and **S3**; Bayesian MCMC, **Figure S4**) and with a site-heterogeneous composition (F_{CAT}) model (**Figure S6**).

DISCUSSION

Taxon Sampling and Model Fit

The chloroplast phylogeny of land plants and streptophytes has been inferred many times previously using different data and analytical approaches, but these studies have often resulted in conflicting phylogenetic patterns (e.g. Nishiyama et al., 2004; Chang and Graham, 2011; Ruhfel et al., 2014). However, almost all studies neglect to test the adequacy of the models used to reconstruct the phylogeny, while others fail even to report the model used (e.g. Gitzendanner et al., 2018). Here we highlight and distinguish between model-fit and model adequacy: while nearly all studies test model fit to choose a "best" substitution model from among a selection, model adequacy is concerned with how well the model fits the data in absolute terms — a best-

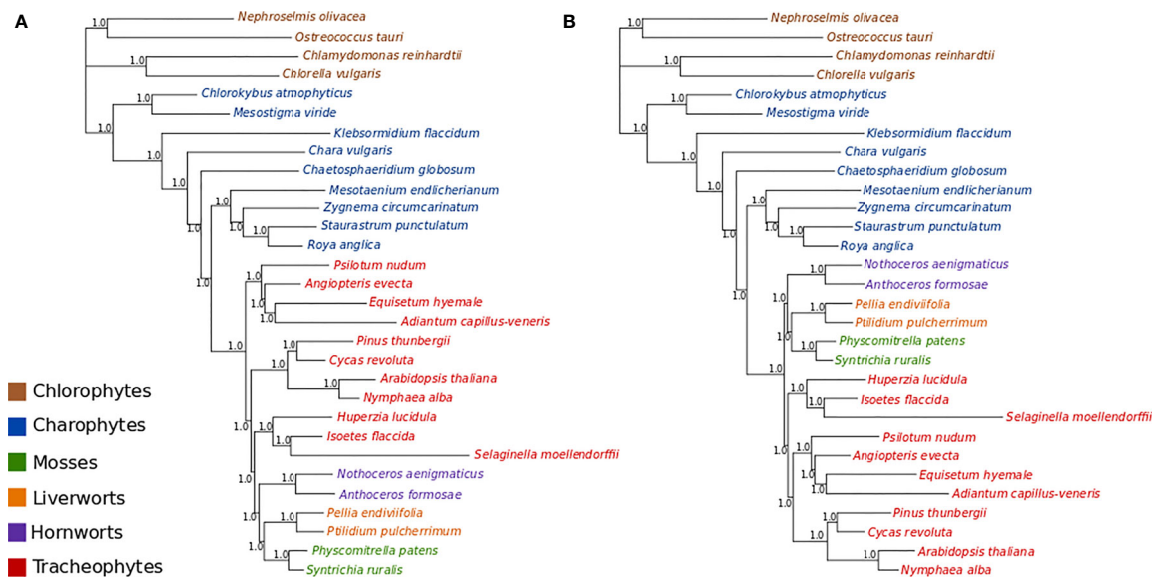


FIGURE 3 | Bayesian MCMC tree-heterogeneous composition analyses of the amino acid data (gcpREV+Γ_4+F_{NDCH2}) **(A)** Run1: Marginal likelihood: L_h= 302615.5702. Posterior predictive simulations of χ^2 statistic of composition homogeneity: original statistic = 3021.4465, sample distribution = 2040.7165 to 3069.3241, p-value= 0.0005. 4,000,000 generations, 40,000 samples, 10,000 discarded as burnin. Mean tree length = 5.6538 substitutions/site. **(B)** Run2: Marginal likelihood: L_h= 302634.1013. Posterior predictive simulations of χ^2 statistic of composition homogeneity: original statistic= 3021.4465, sample distribution = 2047.7021 to 3081.9710, p-value = 0.0002. 4,000,000 generations, 40,000 samples, 10,000 discarded as burnin. Mean tree length = 5.6872 substitutions/site.

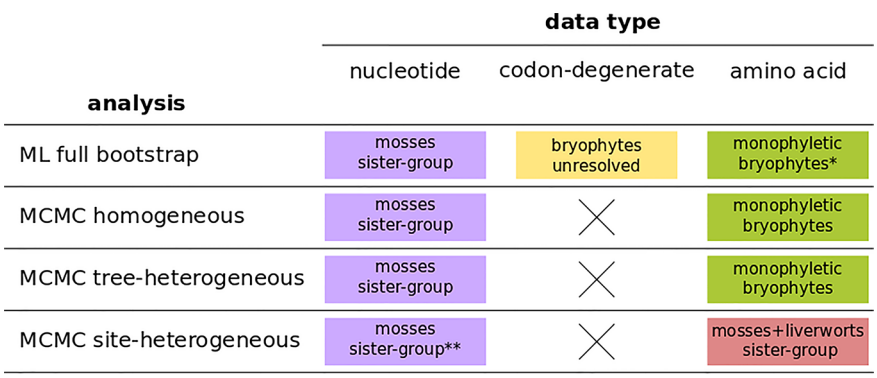


FIGURE 4 | A summary of bryophyte relationships obtained from nucleotide, codon-degenerate, and amino acid translation data using Maximum-likelihood bootstrap analyses (ML) and Bayesian (MCMC) homogeneous, tree-heterogeneous, and site-heterogeneous analyses. Tree nodes were considered supported if the bootstrap value was equal or higher than 80% or if the posterior probability was equal or higher than 95%. (*) part of the analyses without node support; (**) no node support.

fitting model may still be a very poor fit to the data. The three matched-pairs tests we conducted (**Figure 1**) show that all three data sets are neither composition nor rate homogeneous through time. Therefore tree-homogeneous models are likely to be a very poor fit to the data, and yet such models have been, and continue to be, widely used as the only means of reconstructing the phylogeny of land plants. In this paper we employed tree-heterogeneous composition (NDCH2) and site-heterogeneous composition (CAT) model analyses, but to date no single analyses have been conducted that account for both process, and no analyses that account for among-lineage rate variation have been conducted. We identify here that all three processes are likely important for the accurate reconstruction of the land plant phylogeny.

Two studies are notable for having used whole chloroplast genome data together with substitution models that account for composition heterogeneity across sites (Cox et al., 2014) and across taxa (Lemieux et al., 2016). The work of Cox et al. (2014) used a tree-heterogeneous composition model in both nucleotide and amino acid data, showing that amino-acid data support the monophyly of bryophytes, and that when synonymous substitutions are eliminated, support for the non-monophyly of bryophytes is lost in nucleotide data. By contrast, in the work by Lemieux et al. (2016), the analyses of amino acid data using a site-heterogeneous composition model instead showed maximum support for the placement of Setaphyta alone as sister-group to the remaining land plants with the hornworts the sole sister-group to the tracheophytes. In both data sets, sampling of bryophyte lineages was limited, with only one representative of hornworts and one of liverworts. In addition, the data set of Cox et al. (2014) had a very imbalanced proportion of bryophytes and tracheophytes (4:33), which may affect phylogenetic reconstruction, and the data set of Lemieux et al. (2016) lacked any representative of the ferns. The present data set aimed at correcting this sampling bias, and included two representatives of each bryophyte lineage, as well as a balanced representation of each tracheophyte lineage, including ferns. The two taxa for each of the bryophyte lineages were chosen (where possible) to span the likely ancestral node of the lineage with the intention of more accurately reconstructing ancestral states and reducing the length of the subtending branches. By doing this, the genetic distances between lineages were minimized and the likelihood of long-branch attraction reduced. There was a conscious decision to limit the numbers of taxa sampled while sampling as much data as was computationally tractable. Even so, the most complex Bayesian MCMC tree-heterogeneous composition (NDCH2) analyses took > 6 months single CPU computational time to complete per analytical run.

Recent maximum-likelihood analyses of protein-translated plastid transcriptome data spanning the entire green plant kingdom resulted in trees showing the monophyly of bryophytes (Gitzendanner et al., 2018; Leebens-Mack et al., 2019); a result similar to that presented here. However, these studies did not evaluate whether the time-homogeneous models they used in their studies were an adequate fit to their data. This is especially important as Cox et al. (2014) (as again in this study) have shown that land plant plastid data are time-heterogeneous, and therefore the results of these studies are difficult to interpret as they may by compromised by their use of poor-fitting time-homogeneous models.

Conflict Between Nucleotide and Amino Acid Chloroplast Data Is Reduced When Synonymous Substitutions Are Excluded

One common technique used to reduce the probability of systematic errors in phylogenetic reconstruction is to remove data that cannot be adequately modeled, thus increasing the fit of the model and the likely accuracy of the reconstructed trees (e.g. Goremykin et al., 2003). With time, a proportion of site characters uniting a lineage (synapomorphies) are inevitably erased by multiple substitutions and are said to be "saturated" when all phylogenetic signal is lost. Saturation in a protein-coding gene sequence can be reduced by eliminating substitutions which represent synonymous amino-acid replacements. These substitutions occur more rapidly than non-synonymous substitutions as they are not constrained by protein

structure and function and therefore are less likely to reflect accurate phylogenetic signal. By removing synonymous substitutions from the nucleotide data, the tree length was reduced from a very high estimated substitution rate of 9.9 substitutions per site to only 2.7 substitutions per site (**Figures 2A, B**). However, while using degenerate ambiguity recoding to eliminate synonymous substitutions can reduce the amount of non-historical signal present in the data, it does not eliminate it as composition biases can still be caused by different selective pressures for amino acids at protein sites and due to mutational biases. In our analyses, we show that excluding synonymous substitutions eliminates signal in the nucleotide data that supports mosses as sister-group to embryophytes and decreases support for the grouping of hornworts and tracheophytes. Consequently, we think it likely that support for the non-monophyly of the bryophytes in nucleotide sequences is due to non-historical signal (substitutional saturation) present in synonymous sites.

Composition Tree-Heterogeneous Analyses of Chloroplast Amino-Acid Data Support the Monophyly of Bryophytes

ML and Bayesian analyses of chloroplast protein data tend to support the monophyly of the bryophytes (**Figures S9–S15**), however this support is sometimes coincident with the non-monophyly of the tracheophytes. The non-monophyly of the tracheophytes in Bayesian homogeneous and tree-heterogeneous composition analyses, and indeed the implication that tracheophytes are ancestral to bryophytes, is a result that has not been reported before. The topologies where tracheophytes are paraphyletic with the ferns as the earliest-diverging lineage of all land plants, or the ferns are the earliest-diverging lineage of the tracheophytes alone, are almost certainly inaccurate. This is because both the ferns and seed plants share a unique 30-kb inversion in the large, single copy region of the chloroplast genome that is very likely a unique character uniting ferns and seed plants to the exclusion of other taxa as it is thought unlikely that such a structural rearrangement could be reversed (Raubeson and Jansen, 1992). The non-canonical early-branching of the fern lineage suggests that the ferns are being drawn toward the base of the land plants, possibly as an artifact caused by among-site compositions heterogeneity as Phylobayes CAT analyses strongly support the monophyly of the tracheophytes. Unfortunately, the better-fitting Bayesian tree-heterogeneous composition analyses were inconclusive with identical tree topologies having varying marginal likelihood scores: of 4 replicate runs the tracheophyte-paraphyletic topology scored both the best and 3rd best marginal likelihoods, while the tracheophyte-monophyletic topology scored both the 2nd and 4th best marginal likelihoods. This suggests that the composition values sampled at nodes are the critical factor, and not the topology, and that the mixing of the MCMC chains was not efficient enough to allow independent runs to converge to a single best solution.

By contrast, the analyses using composition site-heterogeneous models (Phylobayes CAT) support the Setaphyta as the earliest branching lineage with the hornworts as the sister-group to the tracheophytes. Here the monophyly of the tracheophytes is maximally supported, suggesting that perhaps the modeling of among-site composition heterogeneity is critical to resolving the tracheophytes with amino-acid data. However, posterior predictive tests of the CAT model showed that it failed to describe data heterogeneity across both sites and lineages, with a particularly strong rejection (high Z-scores) of the null hypothesis for the among-lineage composition heterogeneity test. Nevertheless, these analyses suggest the paraphyly of bryophytes, with hornworts the sister-group to tracheophytes, may be the result of among-lineage composition biases as the NDCH2 analyses show strong support for the monophyly of bryophytes. Indeed it may be that among-lineage composition heterogeneity is critical to resolving the monophyly of the bryophytes while at the same time among-site heterogeneity is critical to resolving the monophyly of tracheophytes in these data. Unfortunately, while models combining both these facets of the substitution process are available (e.g. NHPhylobayes, Blanquart and Lartillot, 2006 they are currently computationally intractable with a data set of this size.

The First Land Plants

While the analyses presented here for chloroplast data are inconclusive as to the relationships among the major lineages of plants, they do support bryophytes as a monophyletic group under tree-heterogeneous models. This observation is in accord with some recent analyses of nuclear data (Puttick et al., 2018; Sousa et al., 2019), but not the mitochondrial data (Liu et al., 2014; Sousa et al., 2020). The conclusion that bryophytes are monophyletic, and therefore that tracheophytes are not derived from a bryophyte ancestor, changes our perspective on trait evolution at the stem of the land plants. Indeed, a phylogeny wherein tracheophytes and bryophytes split from a common terrestrial ancestor implies that the alternation of generations in early land plants was not necessarily identical to that of extant bryophytes (Kenrick, 2017), which has an unbranched sporophyte that is fully dependent on the gametophyte. Instead, even if sporophytes were nutritionally dependent on gametophytes at early stages of development, it is possible that the first land plants had a branched sporophyte, and perhaps even near-isomorphic free-living alternate generations, as in the fossil plants *Horneophyton* and *Aglaophyton*, from the Rhynie chert flora. These are considered early polysporangiophytes (Kenrick and Crane, 1997) but, if tracheophytes are not directly derived from bryophytes, they could perhaps have retained traits from an ancestor that pre-dates the bryophyte-tracheophyte split. Thus, a scenario where both gametophytes and sporophytes possessed the necessary machinery for free-living, and became reduced in the tracheophyte and bryophyte lineages, respectively, is as possible as one where tracheophytes evolved from a simple, heterotrophic sporophyte. The evolution of stomata in land plants, given a monophyletic-bryophytes phylogeny, is not so clear, but they are not a shared-derived character (synapomorphy) uniting the hornworts, mosses, and tracheophytes (Mishler and Churchill, 1984; Ruszala et al., 2011). Only if it is assumed that the probability of loss of stomata was greater than the probability of gain (a not unreasonable

assumption, see Harris et al., 2020) then the evolution of stomata would be a synapomorphy uniting all the land plants, with losses in the liverworts and several early-branching moss lineages. Else, if stomata are not homologous among land plants, then they would have been gained independently in the hornworts, mosses, and tracheophytes.

As a corollary to bryophytes forming a monophyletic group, we suggest that a formal classification of the clade containing all three bryophyte lineages as Division (Phylum) Bryophyta Schimp., comprising the three Classes Anthocerotopsida (hornworts), Marchantiopsida (liverworts), and Bryopsida (mosses), will likely have a favorable impact on botanical and evolutionary teaching, as the morphological, reproductive, and ecological traits shared among these three lineages inevitably lead to an intuitive recognition of bryophytes as a natural group.

DATA AVAILABILITY STATEMENT

The accession numbers for the genomes analyzed in this article can be found in Table 1. Alignment files are available on Zenodo (DOI: 10.5281/zenodo.3886964).

AUTHOR CONTRIBUTIONS

CC and PF conceived the study. CC and FS performed analyses. CC, FS, PC, and PF wrote the paper.

FUNDING

This work was supported by FCT (Portuguese Foundation for Science and Technology) through project grant PTDC/BIA-EVF/1499/2014 to CC and national funds through project UIDB/04326/2020, and from the operational programs CRESC Algarve 2020 and COMPETE 2020 through projects EMBRC.PT ALG-01-0145-FEDER-022121 and BIODATA.PT ALG-01-0145-FEDER-022231.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01062/full#supplementary-material

REFERENCES

Ababneh, F., Jermiin, L. S., Ma, C., and Robinson, J. (2006). Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. *Bioinformatics* 22, 1225–1231. doi: 10.1093/bioinformatics/btl064

Abascal, F., Zardoya, R., and Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38, 7–13. doi: 10.1093/nar/gkq291

Birky, C. W. Jr. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.* 35, 125–148. doi: 10.1146/annurev.genet.35.102401.090231

Blanquart, S., and Lartillot, N. (2006). A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol. Biol. Evol.* 23, 2058–2071. doi: 10.1093/molbev/msl091

Bock, R. (2010). The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci.* 15, 11–22. doi: 10.1016/j.tplants.2009.10.001

Brown, J. M., Hedtke, S. M., Lemmon, A. R., and Lemmon, E. M. (2010). When trees grow too long: investigating the causes of highly inaccurate Bayesian branch-length estimates. *Syst. Biol.* 59, 145–161. doi: 10.1093/sysbio/syp081

Chang, Y., and Graham, S. W. (2011). Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. *Am. J. Bot.* 98, 839–849. doi: 10.3732/ajb.0900384

Cox, C. J., and Foster, P. G. (2013). A 20-state empirical amino-acid substitution model for green plant chloroplasts. *Mol. Phylogenet. Evol.* 68, 218–220. doi: 10.1016/j.ympev.2013.03.030

Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R., and Embley, T. M. (2008). The archeabacterial origin of eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20356–20361. doi: 10.1073/pnas.0810647105

Cox, C. J., Li, B., Foster, P. G., Embley, T. M., and Civáň, P. (2014). Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Syst. Biol.* 63, 272–279. doi: 10.1093/sysbio/syt109

Cox, C. J. (2018). Land plant molecular phylogenetics: a review with comments on evaluating incongruence among phylogenies. *Crit. Rev. Plant Sci.* 37, 113–127. doi: 10.1080/07352689.2018.1482443

Foster, P. G., Cox, C. J., and Embley, T. M. (2009). The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philos. Trans. R. Soc B* 364, 2197–2207. doi: 10.1098/rstb.2009.0034

Foster, P. G. (2004). Modeling compositional heterogeneity. *Syst. Biol.* 53, 485–495. doi: 10.1080/10635150490445779

Gao, L., Su, Y. J., and Wang, T. (2010). Plastid genome sequencing, comparative genomics, and phylogenomics: Current status and prospects. *J. Syst. Evol.* 48, 77–93. doi: 10.1111/j.1759-6831.2010.00071.x

Gitzendanner, M. A., Soltis, P. S., Wong, G. K. S., Ruhfel, B. R., and Soltis, D. E. (2018). Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *Am. J. Bot.* 105, 291–301. doi: 10.1002/ajb2.1048

Goremykin, V. V., Hirsch-Ernst, K.II, Wölfl, S., and Hellwig, F. H. (2003). Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol. Biol. Evol.* 20, 1499–1505. doi: 10.1093/molbev/msg159

Harris, B. J., Harrison, C. J., Hetherington, A. M., and Williams, T. A. (2020). Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. *Curr. Biol.* 30, 2001–2012. doi: 10.1016/j.cub.2020.03.048

Hedderson, T. A., Chapman, R. L., and Rootes, W. L. (1996). Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. *Plant Syst. Evol.* 200, 213–224. doi: 10.1007/BF00984936

Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231. doi: 10.1016/j.tig.2006.02.003

Jermiin, L. S., and Misof, B. (2020). Measuring historical and compositional signals in phylogenetic data. *bioRxiv*. doi: 10.1101/2020.01.03.894097

Jermiin, L. S., Jayaswal, V., Ababneh, F. M., and Robinson, J. (2017). "Identifying optimal models of evolution," in *Bioinformatics* (New York, NY: Humana Press), 379–420.

Jermiin, L. S., Lovell, D. R., Misof, B., Foster, P. G., and Robinson, J. (2020). Detecting and visualising the impact of heterogeneous evolutionary processes on phylogenetic estimates. *bioRxiv*. doi: 10.1101/828996

Karol, K. G., Arumuganathan, K., Boore, J. L., Duffy, A. M., Everett, K. D. E., Hall, J. D., et al. (2010). Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evol. Biol.* 10, 321. doi: 10.1186/1471-2148-10-321

Karol, K. G. (2001). The Closest Living Relatives of Land Plants. *Science* 294, 2351–2353. doi: 10.1126/science.1065156

Keane, T. M., Creevy, C. J., Pentony, M. M., Naughton, T. J., and McInerney, J. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* 6, 29. doi: 10.1186/1471-2148-6-29

Kenrick, P., and Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature* 389, 33. doi: 10.1038/37918

Kenrick, P. (2017). How land plant life cycles first evolved. *Science* 358, 1538–1539. doi: 10.1126/science.aan2923

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., and Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenetic datasets. *BMC Evol. Biol.* 14, 82. doi: 10.1186/1471-2148-14-82

Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095–1109. doi: 10.1093/molbev/msh112

Leebens-Mack, J., Barker, M., Carpenter, E., Deyholos, M., Gitzendanner, M., Graham, S., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

Lemieux, C., Otis, C., and Turmel, M. (2016). Comparative chloroplast genome analyses of streptophyte green algae uncover major structural alterations in the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae. *Front. Plant Sci.* 7, 697. doi: 10.3389/fpls.2016.00697

Lemmon, A. R., Brown, J. M., Stanger-Hall, K., and Lemmon, E. M. (2009). The effect of ambiguous data on phylogenetic estimates obtained by likelihood and Bayesian inference. *Syst. Biol.* 58, 130–145. doi: 10.1093/sysbio/syp017

Lewis, L. A., Mishler, B. D., and Vilgalys, R. (1997). Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene rbcL. *Mol. Phylogenet. Evol.* 7, 377–393. doi: 10.1006/mpev.1996.0395

Liu, Y., Cox, C. J., and Goffinet, B. (2014). Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Syst. Biol.* 63, 862–878. doi: 10.1093/sysbio/syu049

Magallón, S., Hilu, K. W., and Quandt, D. (2013). Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am. J. Bot.* 100, 556–573. doi: 10.3732/ajb.1200416

McCourt, R. M., Delwiche, C. F., and Karol, K. G. (2004). Charophyte algae and land plant origins. *Trends Ecol. Evol.* 19, 661–666. doi: 10.1016/j.tree.2004.09.013

Mishler, B. D., and Churchill, S. P. (1984). A cladistic approach to the phylogeny of the "bryophytes". *Brittonia* 36, 406–424. doi: 10.2307/2806602

Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., et al. (2018). The timescale of early land plant evolution. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2274–E2283. doi: 10.1073/pnas.1719588115

Nishiyama, T., and Kato, M. (1999). Molecular phylogenetic analysis among bryophytes and tracheophytes based on combined data of plastid coded genes and the 18S rRNA gene. *Mol. Biol. Evol.* 16, 1027–1036. doi: 10.1093/oxfordjournals.molbev.a026192

Nishiyama, T., Wolf, P. G., Kugita, M., Sinclair, R. B., Sugita, M., Sugiura, C., et al. (2004). Chloroplast phylogeny indicates that bryophytes are monophyletic. *Mol. Biol. Evol.* 21, 1813–1819. doi: 10.1093/molbev/msh203

Puttick, M. N., Morris, J. L., Williams, T. A., Cox, C. J., Edwards, D., Kenrick, P., et al. (2018). The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr. Biol.* 28, 733–745. doi: 10.1016/j.cub.2018.01.063

Raubeson, L. A., and Jansen, R. K. (1992). Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* 255, 1697–1699. doi: 10.1126/science.255.5052.1697

Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., and Burleigh, J. G. (2014). From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* 14, 23. doi: 10.1186/1471-2148-14-23

Ruszala, E. M., Beerling, D. J., Franks, P. J., Chater, C., Casson, S. A., Gray, J. E., et al. (2011). Land plants acquired active stomatal control early in their evolutionary history. *Curr. Biol.* 21, 1030–1035. doi: 10.1016/j.cub.2011.04.044

Sousa, F., Foster, P. G., Donoghue, P. C., Schneider, H., and Cox, C. J. (2019). Nuclear protein phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). *New Phytol.* 222, 565–575. doi: 10.1111/nph.15587

Sousa, F., Civáň, P., Brazão, J., Foster, P. G., and Cox, C. J. (2020). The mitochondrial phylogeny of land plants shows support for Setaphyta under composition-heterogeneous substitution models. *PeerJ* 8, e8995. doi: 10.7717/peerj.8995

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Talavera, G., and Castrasana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, C., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U. S. A.* 111, E4859–E4868. doi: 10.1073/pnas.1323926111

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., et al. (2014). Contribution of NAC transcription factors to plant adaptation to land. *Science* 343, 1505–1508. doi: 10.1126/science.1248417

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 05 August 2020
doi: 10.3389/fpls.2020.01185

Transcriptome Analysis of *ppdnmt2* and Identification of Superoxide Dismutase as a Novel Interactor of DNMT2 in the Moss *Physcomitrella patens*

Darshika Singh[1†], *Radha Yadav*[1†], *Shubham Kaushik*[2], *Nikita Wadhwa*[1], *Sanjay Kapoor*[3] *and Meenu Kapoor*[1*]

[1] University School of Biotechnology, Guru Gobind Singh Indraprastha University, New Delhi, India, [2] Vproteomics, Valerian Chem Private Limited Green Park Mains, New Delhi, India, [3] Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi, India

OPEN ACCESS

Edited by:
Stefan de Folter,
National Polytechnic Institute of Mexico (CINVESTAV), Mexico

Reviewed by:
Mario Alberto Arteaga-Vazquez,
University of Veracruz, Mexico
Pierre-François Perroud,
University of Marburg, Germany

***Correspondence:**
Meenu Kapoor
meenukapoor@me.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: 28 April 2020
Accepted: 21 July 2020
Published: 05 August 2020

Citation:
Singh D, Yadav R, Kaushik S, Wadhwa N, Kapoor S and Kapoor M (2020) Transcriptome Analysis of ppdnmt2 and Identification of Superoxide Dismutase as a Novel Interactor of DNMT2 in the Moss Physcomitrella patens. Front. Plant Sci. 11:1185. doi: 10.3389/fpls.2020.01185

DNMT2 is a DNA/tRNA cytosine methyltransferase that is highly conserved in structure and function in eukaryotes. In plants however, limited information is available on the function of this methyltransferase. We have previously reported that in the moss *Physcomitrella patens*, *DNMT2* plays a crucial role in stress recovery and $tRNA^{Asp}$ transcription/stability under salt stress. To further investigate the role of *PpDNMT2* at genome level, in this study we have performed RNA sequencing of *ppdnmt2*. Transcriptome analysis reveals a number of genes and pathways to function differentially and suggests a close link between *PpDNMT2* function and osmotic and ionic stress tolerance. We propose *PpDNMT2* to play a pivotal role in regulating salt tolerance by affecting molecular networks involved in stress perception and signal transduction that underlie maintenance of ion homeostasis in cells. We also examined interactome of PpDNMT2 using affinity purification (AP) coupled to mass spectrometry (AP-MS). Quantitative proteomic analysis reveals several chloroplast proteins involved in light reactions and carbon assimilation and proteins involved in stress response and some not implicated in stress to co-immunoprecipitate with PpDNMT2. Comparison between transcriptome and interactome datasets has revealed novel association between *PpDNMT2* activity and the antioxidant enzyme Superoxide dismutase (SOD), protein turnover mediated by the Ubiquitin-proteasome system and epigenetic gene regulation. PpDNMT2 possibly exists in complex with CuZn-SODs *in vivo* and the two proteins also directly interact in the yeast nucleus as observed by yeast two-hybrid assay. Taken together, the work presented in this study sheds light on diverse roles of *PpDNMT2* in maintaining molecular and physiological homeostasis in *P. patens*. This is a first report describing transcriptome and interactome of DNMT2 in any land plant.

Keywords: DNMT2, *Physcomitrella*, transcriptome, superoxide dismutase, stress

INTRODUCTION

Methylation of cytosine residues (m^5C) in transfer RNAs (tRNAs) is a conserved modification found in this class of non-coding RNAs across eukaryotes (Helm, 2006; Motorin et al., 2009). It has been described in the nuclear tRNAs in lower and higher eukaryotes including the single-celled green algae and the multicellular flowering plants (Wilkinson et al., 1995; Okano et al., 1998; Yoder and Bestor, 1998; Goll et al., 2006; Burgess et al., 2015). Methylation of tRNAs is known to stabilize RNA secondary structures and prevent its cleavage by ribonucleases under a variety of stress conditions and also affect developmental processes in both plants and animals (Rai et al., 2007; Schaefer et al., 2010; Tuorto et al., 2012; Burgess et al., 2015; David et al., 2017). There are two classes of RNA methyltransferases (RMTases) that catalyze m^5C in tRNAs: the Transfer RNA specific methyltransferase 4 (TRM4) or NOP2/Sun domain protein 2 (NSUN2) in yeast and animals that has two paralogs in Arabidopsis (TRM4a and TRM4b) (Pavlopoulou and Kossida, 2009; Chen et al., 2010; Tuorto et al., 2012; Burgess et al., 2015; Motorin et al., 2016) and the Transfer RNA aspartic acid methyltransferase 1 (TRDMT1) or DNA methyltransferase 2 (DNMT2) (Goll et al., 2006). In Arabidopsis, both the RMTase catalyze m^5C in the variable and anticodon loop in a number of tRNAs (Burgess et al., 2015; David et al., 2017).

DNMT2/TRDMT1 is a highly conserved cytosine methyltransferase in eukaryotes. It differs from other proteins in this family in lacking the N-terminal regulatory domains, being shorter in length and in methylating cytosines in both DNA and tRNAs (Wilkinson et al., 1995; Goll et al., 2006). It was first identified in the fission yeast, *Schizosaccharomyces pombe* as *pombe methyltransferase 1* (*pmt1*) and is known to be the only methyltransferase in the genomes of many model organisms including *Drosophila melanogaster*, *Dictyostelium discoidium*, and *Entamoeba histolytica* (Wilkinson et al., 1995; Okano et al., 1998; Yoder and Bestor, 1998). *DNMT2* expression is developmentally regulated in these organisms (Hung et al., 1999; Kuhlmann et al., 2005; Rai et al., 2007; Müller et al., 2013). Loss-of-function mutations in *DNMT2* does not have deleterious consequences on genomic methylation patterns, embryonic development, vegetative growth, chromosome stability, and mating type switching in mice, Drosophila, and yeast under standard conditions, but its role in affecting viability, silencing of mobile elements, innate immune response, telomere integrity, and histone H4K20 trimethylation is evident under stress conditions (Wilkinson et al., 1995; Lin et al., 2005; Katoh et al., 2006; Schaefer et al., 2010; Becker et al., 2012). DNMT2 also affects non-random segregation of sister chromatids in the asymmetrically dividing Drosophila male germline stem cells (Yadlapalli and Yamashita, 2013).

Though DNMT2 bears close similarity in sequence and structure to DNA methyltransferases, it however has weak DNA methyltransferase activity and a low preference for cytosines located as CpG in comparison to Cp(A/T) or CC(A/T)GG (Pinarbasi et al., 1996; Hermann et al., 2003; Kunert et al., 2003; Liu et al., 2003; Tang et al., 2003; Fisher et al., 2004; Mund et al., 2004; Kuhlmann et al., 2005). Active methylation of cytosines in tRNA molecules has been reported as a widely conserved function of DNMT2 in different organisms. DNMT2-mediated tRNA methylation is known to protect tRNAs against ribonucleases under heat and oxidative stress in Drosophila (Schaefer et al., 2010). This has also been correlated with small RNA homeostasis in flies (Durdevic et al., 2013).

Since DNMT2 has affinity for both DNA and RNA substrates and is involved in diverse biological processes, it is predicted to interact with a number of proteins. The putative interactors of DNMT2 based on high evolutionary rate covariation values have been identified as proteins belonging to either of the following categories: chromatin remodeling, transcription factor, gene expression regulation, DNA replication, stress response, and RNA editing (Vieira et al., 2018). The first *in vivo* protein partner reported for DNMT2 was the glycolytic protein Enolase that catalyzes conversion of 2-Phosphoglycerate to Phosphoenolpyruvate during glycolysis in Entamoeba (Tovy et al., 2010). Enolase binds to the *Entamoeba histolytica* 5-cytosine DNA methyltransferase (Ehmeth) and inhibits its tRNA methylation activity under glucose starvation conditions. In humans, DNMT2 interacts with a number of proteins involved in RNA processing, transcriptional control, and mRNA export (Thiagarajan et al., 2011). In plants such as Arabidopsis, interaction of AtDNMT2 with plant-specific histone deacetylases2 (HD2) proteins HD2A, 2B, and 2C suggests a close link between DNMT2 and epigenetic regulatory mechanisms (Song et al., 2010).

We have previously reported that the moss *Physcomitrella patens* encodes the longest eukaryotic DNMT2 with 477 residues that shares more than 50% identity with Arabidopsis DNMT2 and around 40% with corresponding proteins in yeast, humans, and mice, respectively (Malik et al., 2012; Arya et al., 2016). Although *PpDNMT2* differentially expresses in protonema and the leafy gametophores, its loss-of-function does not affect growth and development at these stages (Malik et al., 2012; Arya et al., 2016). Instead, *ppdnmt2* mutants display high sensitivity towards elevated salt and mannitol in the growth medium and are unable to recover from stress despite accumulation of normal levels of *PpDHNA* (encoding dehydrin) and the small heat shock protein encoding *PpHSP16.4* transcripts known to be essential for stress recovery in *P. patens* (Saavedra et al., 2006; Ruibal et al., 2012; Arya et al., 2016). To further investigate the functional capabilities of *PpDNMT2*, in this study we have analyzed the transcriptome of *ppdnmt2* by RNA sequencing and identified a number of differentially expressing genes and pathways that are affected by the loss of *PpDNMT2* function. Further, to gain insight into proteins existing in complex with PpDNMT2 *in vivo* we performed immunoprecipitation and identified proteins co-immunoprecipitated with PpDNMT2 by Liquid Chromatography-Mass Spectrometry (LC-MS). The transcriptome and interactome datasets reveal close link between *PpDNMT2* function and the antioxidant enzymes such as the Superoxide dismutases (SOD), osmotic and ionic homeostasis maintenance, regulation of protein turnover, and epigenetic regulation in *P. patens*.

MATERIALS AND METHODS

Plant Material and RNA Isolation

For RNA sequencing, *Physcomitrella patens* Gransden (Wild type and *ppdnmt2*#8) was propagated on solid BCDAT

medium (http://moss.nibb.ac.jp/protocol.html) overlaid with cellophane under continuous white light (40 μmol/m^2/s) at 22–24°C. The cellophanes were shifted on fresh BCDAT medium every 7th day, and tissue was harvested on the 24th day. Total RNA was isolated from three biological replicates of each sample type using Trizol (Sigma) and cleaned up using GeneJET RNA clean-up and concentration micro kit (Thermo Scientific, Inc.) before eluting in RNase-free water. RNA quality and quantity were checked by fractionating samples on 1.2% agarose gel and using Nanodrop (Nanodrop LITE Thermoscientific).

RNAseq Library Construction and Sequencing

RNA sequencing library was prepared using the Illumina-compatible NEBNext® Ultra™ Directional RNA Library Prep Kit (New England BioLabs, USA) at Genotypic Technology Pvt. Ltd., Bengaluru, India following manual instructions. Briefly, 1 μg of total RNA was taken for mRNA isolation, fragmentation, and priming. The fragmented and primed mRNAs were then subjected to first strand synthesis in the presence of Actinomycin D (Gibco, Life Technologies, USA) followed by second strand synthesis. The double stranded cDNA was purified using HighPrep magnetic beads (Magbio Genomics Inc, USA) followed by end repair, adenylation and ligation to Illumina multiplex barcode adapters as per NEBNext® Ultra™ Directional RNA library preparation kit protocol. Adapter ligated cDNAs were purified using HighPrep beads and subjected to 12 cycles of indexing-PCR (37°C for 15 min followed by denaturation at 98°C for 30s) cycling (98°C for 10 s, 65°C for 75 s) and at 65°C for 5 min to enrich adapter-ligated cDNA fragments. Final PCR products (sequencing library) were purified using the HighPrep beads followed by library quality control check. Illumina-compatible sequencing library was quantified using Qubit fluorometer (Thermo Fisher Scientific, USA), and its fragment size distribution (80 bp to 680 bp) was analyzed on Agilent 2200 Tapestation. Paired End (PE) sequencing (150 × 2) was then performed using the high-throughput Illumina HiSeq sequencing platform, HiSeq X Ten (Illumina, Inc.). Raw reads generated have been deposited in the NCBI BioSample database and can be accessed through Sequence Read Archive (SRA) accession # **SRR8555685, SRR8555689, SRR8555687, SRR8555688, SRR8555690, and SRR8574017**.

Processing of Reads, Alignment Against *P. patens* Reference Genome and Differential Gene Expression Analysis

Quality of raw reads was checked using Fast QC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and these were pre-processed to remove adaptor sequences and low quality bases using Cutadapt (Martin, 2011). The processed reads were then aligned to the reference genome of *Physcomitrella patens_318_V3.3.* (https://phytozome.jgi.doe.gov/pz/portal.html) using TopHat-2.0.13 (Trapnell et al., 2009). The resulting files in bam format were used for downstream analysis. Transcript assembly, read count and differential gene expression was performed using Cufflinks-2.2.1 (Trapnell et al., 2010). Group wise comparison was performed to identify differentially regulated genes between wild type and *ppdnmt2#8* using cuffdiff. Up and downregulated genes were identified using log2 fold cut-off (+/−1) with P-value <= 0.05. Heat map was generated in R using Bioconductor package gplots.

Gene Ontology and Pathway Analysis

All transcripts and proteins were annotated against Viridiplantae sequences downloaded from uniprot (http://www.uniprot.org/) and using The *Physcomitrella patens* Gene Model Lookup Database (PpGML DB; https://peatmoss.online.uni-marburg.de/ppatens_db/contact.php; Fernandez-Pozo et al., 2019), and GO terms were assigned accordingly. Fischer's one-tailed test was used to calculate p-values. GO terms with p-values below the threshold of <=0.05 were considered statistically significant. GO enrichment analysis along with statistical filtering using Bonferroni correction was performed using PANTHER database (http://pantherdb.org/). Pathway analysis was performed using the KEGG pathways database (https://www.genome.jp/kegg/pathway.html).

Antibody Preparation

Anti-PpDNMT2 antibody was synthesized using PpDNMT2-specific peptide 302-APPLLRKLIGDHYES-316 as antigen at BIONEEDS, Karnataka, India. For antisera development, rabbits were first immunized with the peptide antigen conjugated to a carrier protein followed by 2–4 rounds of boosters every 21 days. Titer was checked from 2nd booster onwards, 10 days after injection, and antisera was collected after the desired titer was achieved. The sera were then affinity purified using Protein A.

Stress Treatment, Protein Extraction and Western Blot

For AP-MS analysis, wild type protonema was propagated for nine days on solid BCDAT media under continuous white light at 22–25°C. The tissue was then transferred to liquid BCDAT for 5 days before supplementing the media with 400 mM NaCl and incubating further for 24 h. Tissue propagated under similar conditions and for the same period of time but without NaCl supplement was used as control. Three independent biological samples were used for control and salt-treated samples. Total protein was isolated by manually grinding 200 mg tissue in liquid nitrogen and adding 400 μl lysis buffer (150 mM NaCl, 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1 mM EGTA (pH 8.0), 0.2 mM sodium ortho-vanadate, 0.2 mM PMSF, 1% Triton X-100 (Sigma), 0.5% NP-40 (Himedia) and 1× ProteaseArrest™ (G Biosciences) to the homogenate. After incubating the protein extract on ice for 30 min it was sonicated four times for 8 s each with at least an 8 s pause in between. Cell debris was then pelleted by centrifugation at 14,000 rpm for 10 min at 4°C. The clear supernatant was collected, and protein concentration was estimated by Bradford method using CB protein assay™ reagent (G Biosciences). *PpDNMT2* expression was confirmed by performing western blot. For this, 50 μg total protein was denatured by adding 1× SDS gel loading buffer (375 mM 1M Tris-Cl pH 6.8, 9% SDS, 50% Glycerol, 0.03% Bromophenol blue) and heated at 99°C for 10 min. The samples were fractionated using 12% SDS polyacrylamide gel at 70 V. After

electrophoresis the gel was rinsed with sterile water and equilibrated in transfer buffer (25 mM Tris base, 192 mM Glycine, and 10% methanol) before blotting onto activated PVDF membrane (Merck Millipore) overnight at 4°C. After transfer, the membrane was washed once with TBS (20 mM Tris, 150 mM NaCl) and immersed in blocking solution (5% skimmed milk (Difco) in TBST: 1× TBS + 0.1% Tween 20) for 2 h at room temperature with constant shaking. Primary antibody (anti-PpDNMT2) at 1:500 dilution was then added and incubated for 60 min. After washing with TBST (15 min × 3), secondary antibody (Goat Anti-Rabbit IgG, HRP-conjugate, Merck Millipore) at 1:5,000 dilution was added, and membrane was further incubated for 60 min. It was then washed with TBST (15 min × 3) and developed using TMB (3,3′,5,5′-Tetramethylbenzidine, Sigma) substrate.

Immunoprecipitation and Silver Staining

Endogenous PpDNMT2 was immunoprecipitated using anti-PpDNMT2 antibody (Ab, Ligand) coupled to Dynabeads-Protein A conjugate. Protein extracts (500 µg each) prepared from control, and treated samples were incubated with 7 µg anti-PpDNMT2 Ab overnight at 4°C with constant shaking. 50 µl Dynabeads-Protein A conjugate (Invitrogen, Thermo Fisher Scientific) was then added and incubated further for 2 h at 4°C. The tube was then placed on the magnetic stand to collect the Dynabeads-Protein A-Ab complex at the tube wall while the supernatant was discarded. The beads were washed twice with 500 µl lysis buffer. Three independent immunoprecipitation experiments each for control and salt-treated samples were carried out. For analyzing the bound antigen on SDS-PAGE, 2× SDS-PAGE sample loading buffer (G Biosciences) was added to the bead bound complex and boiled at 70°C for 10 min. After electrophoresis using 12% SDS-polyacrylamide gel silver staining was performed. For this, the gel was first washed with sterile water and then fixed for 60 min using 30% ethanol, 10% acetic acid adding fresh fixing solution after 30 min. After rinsing the gel with 20% ethanol (10 min × 2) and then sterile water for the same time, it was soaked for 1 min in 0.02% Sodium thiosulphate pentahydrate. After again rinsing with water (1 min × 2), the gel was impregnated with 12 mM silver nitrate for 30 min in dark without shaking. The gel was then briefly washed with water and developed using freshly prepared 3% potassium carbonate containing 10% sodium thiosulphate and 37% formalin for 45 min. The reaction was terminated using stop solution (0.3 M Tris and 2% acetic acid).

LC-MS Analysis

After immunoprecipitation, the antigen along with the co-immunoprecipitated proteins was separated from Dynabeads-Protein A-Ab complex by adding 30 µl elution buffer (50 mM Glycine, pH 2.8). The pH of the eluate was adjusted using 1M Tris pH 7.5, and samples were sent to Central Instrumentation Facility at University of Delhi South Campus for LC-MS. Briefly, 25 µg of each protein sample was first reduced using 10 mM Dithiothreitol (DTT, Pierce) in 50 mM Ammonium BiCarbonate (ABC, pH~8) at 37°C for 30 min and then alkylated using 50 mM Iodoacetamide (IAA, Sigma) in 50 mM ABC (pH~8) in dark for another 30 min. After diluting the samples 10 times with water, the proteins were digested into smaller peptides using the serine protease Trypsin (Promega). This was added in the ratio of 1:50 (Trypsin : Lysate ratio) and incubated overnight at 37°C. The reaction was stopped by adding 10% Trifluoroacetic acid (TFA). Digests were then cleaned up using Pierce™ C18 Spin Columns according to manufacturer's protocol and dried using speed vac (Thermo Savant DNA 120). The pellet was finally resuspended in Buffer-A (0.1% formic acid in water). The clarified peptide digested samples (1 µg each) were then resolved on a 50-cm long EASY-Spray column (50 cm * 75 um) PepMap RSLC filled with 2 µm-C18 resin on nano1200 chromatography system (ThermoFischer Scientific) attached to QExactive mass spectrometer equipped with nano-electrospray ion source. The peptides were loaded with Buffer A and eluted with a 5–40% gradient of Buffer-B (100% acetonitrile, 0.1% formic acid) for 102 min, 40–90% gradient for 1 min, followed by 90% gradient for 15 min at a flow rate of 300 nl/min with a total run time of 123 min.

The QExactive spray voltage was set at 4 kV, S lens RF level at 60 and ITC heated capillary temperature at 300°C. The MS data were acquired in positive polarity using a data-dependent method choosing the 10 most intense peaks with charge state +2 to +5, exclude isotope option enabled and dynamic exclusion time of 12 s. The MS1 (mass range 350–2,000 m/z) and MS2 scans were acquired in Orbitrap Mass analyser with resolution of 70,000 and 17,500 at m/z 200–2,000, respectively with polydimethylcyclosiloxane (PCM) ions (m/z = 445.120025), and lock mass option was enabled for internal recalibration during the run.

The MS1 or Full scan target was 3×10^6 with a maximum fill time of 60 ms with mass range set to 350–2,000. Target value for MS2 or fragment scans was set at 1×10^5, and intensity threshold was set at 8.3×10^2. Isolation window of parent ion of interest was set at 1.5 m/z. Normalized collision energy for Higher-energy collisional dissociation (HCD) was set at 27. Peptide match option was set to preferred mode along with activation of isotope exclusion option.

Protein Identification and Quantification

The RAW files containing the MS/MS spectra were analyzed using Proteome Discoverer (v2.2) software. The search engines Sequest HT and MS Amanda 2.0 based algorithms SEQUEST (Yates et al., 1995) and AMANDA (Dorfer et al., 2014) respectively, were used for matching peptides with proteins downloaded from Uniprot Proteins database for *Physcomitrella patens* containing 61,529 entries (https://www.uniprot.org/uniprot/?query=physcomitrella+organism%3Apatns&sort=score). The protease Trypsin was used to generate peptides, and specificity for Trypsin/P was set (cleavage at the C terminus of "K/R: unless followed by "P") with maximum missed cleavage value of 2. Carbamidomethyl on cysteine was considered as fixed modification, while oxidation of methionine and N-terminal acetylation were considered as variable modifications. For Sequest HT and MS Amanda 2.0 search, the precursor and fragment mass tolerances were set at 10 ppm and 0.5 Da, respectively. Minimum peptide length for search was set at 6, while maximum peptide length was set at 150. Both peptide spectrum match and protein False Discovery Rate (FDR) were set to 0.01 and determined using percolator node. Relative protein

quantification was performed using Minora feature detector node of Proteome Discoverer 2.2 including a Minimum Trace Length of 5, Max. ΔRT of Isotope Pattern 0.2 min and considering only those with high PSM (peptide spectrum matches) confidence.

Differential Analysis

Differential analysis was performed using protein abundance values in each sample type. The abundance values in each biological replicates were Log2 transformed followed by Z-score standardization for normalization of data. Statistical significance was inferred by performing Student t-test and proteins with p-value <= 0.05 and fold change >= 2 were considered statistically significant. Data was visualized using in-house R scripts.

Construct Preparation

Constructs were prepared for yeast two-hybrid assays. PpDNMT2 coding region (1,434 bp) was amplified using the primer pair PpDNMT2TOPOY2HFP and PpDNMT2TOPOY2HRP and cloned in pENTR/D–TOPO vector (Invitrogen; **Table S1**). The fragment was then shuttled into the pGADT7 destination vector [pGADT7-GW (Clontech laboratories Inc. USA)] using the Gateway LR Clonase™ II enzyme mix (Invitrogen, USA) following manufacturer's instructions. Bait construct was prepared using the yeast expression vector pGBKT7 (Clontech laboratories Inc. USA). *Pp3c9_25690V3.1* coding fragment (621 bp) was amplified using cDNA prepared from total RNA isolated from protonema tissue subjected to 400 mM NaCl stress for 24 h. Gene-specific primer pair Pp3c9_25690Y2H_FP and Pp3c9_25690Y2H_RP were used for PCR. The amplicon was digested with and ligated to *Sal*I and *Pst*I digested pGBKT7 vector. The construct preparations were validated by restriction enzyme digestion and DNA sequencing.

Yeast Two-Hybrid Assay

Yeast transformations were performed using the EZ-Yeast transformation kit (MP Biomedicals, USA) as per manufacturer's recommendations. Briefly, 2 µg each of pGADT7 (AD) and pGBKT7 (BD) fusion constructs were co-transformed in yeast AH109 suspended in 125 µl EZ-Transformation solution along with 5 µl carrier DNA. The mixture was incubated at 42°C for 30 min. The cells were then plated on Synthetic Dropout/-Leucine/-Tryptophan, SD/-Leu/-Trp (SD-LW) plates and incubated at 30°C for 3–4 days. Transformants showing robust growth were then streaked on SD/-Leu/-Trp/-Histidine (SD-LWH), SD/-Leu/-Trp/-His/-Adenine (SD-LWHA) and SD-LWHA + X-α-Gal synthetic dropout solid media plates for selection of transformants displaying protein–protein interactions.

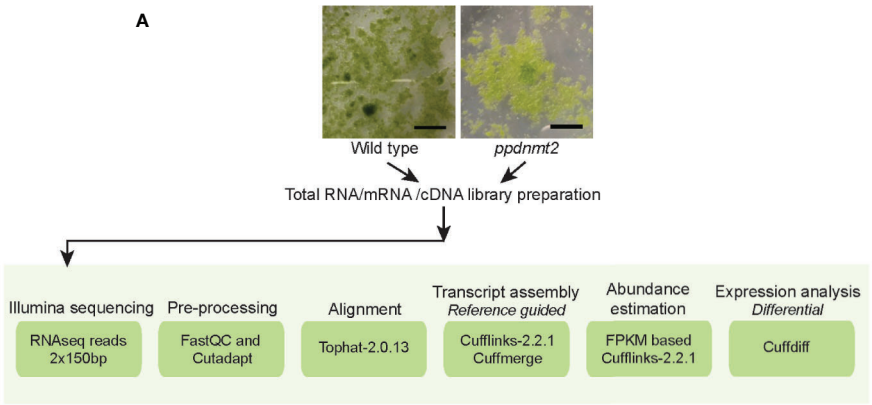
RESULTS

Identification of Differentially Expressed Genes in *PpDNMT2* Mutant

To gain insight into the molecular effects of loss of *PpDNMT2* function, differentially expressed genes in *ppdnmt2#8* were identified by comparing its transcriptome with wild type plants using reference-based transcriptome method. We had previously characterized *ppdnmt2#8* (here after referred to as *ppdnmt2)* and shown that these plants were highly sensitive to salt and mannitol stress (Arya et al., 2016). Using the experimental pipeline for RNAseq outlined in **Figure 1A**, a total of 181.60 million Illumina HiSeq reads were generated for both sample types by RNA sequencing out of which 165.09 million high quality adapter free processed reads were subjected to downstream analysis. On an average 90.91% of the reads in each biological replicate aligned to the reference genome (**Figure 1B**). Next, to correct biasness in gene length and sequencing depth, the data was normalized using FPKM metrics. Normalized read values obtained for all the biological replicates of control and *ppdnmt2* are summarized in **Table S2**. Applying a log2 fold cut-off (+/−1), 1,237 upregulated genes and 708 downregulated genes were identified in *ppdnmt2*, with seven genes expressing only in wild type and nine expressing only in *ppdnmt2* (**Figure 2A**). Out of these up and downregulated genes, 708 out of 1237 and 398 out of 708 genes were found to be P-significant with p-value <= 0.05 (**Table S3**).

Heat map of top 20 up and downregulated genes shows differential accumulation of transcripts in wild type and *ppdnmt2* (**Figure 2B**). Among the genes upregulated in *ppdnmt2* are those encoding the membrane localized oxidoreductase NADH : Fe(3^+)-EDTA reductase (*Pp3c27_2390V3.1*) that catalyzes reduction of ferric ions ($Fe3^+$) into the useful ferrous form; chloroplast localized SUPEROXIDE DISMUTASE [Fe], Fe-SOD (*Pp3c17_14510V3.1*) the antioxidant enzyme catalyzing dismutation of superoxide radicals to water and molecular oxygen (Mittler et al., 2004); copper chaperone for SOD (*Pp3c17_14500V3.1*) known to be involved in trafficking copper ions from cytosol to chloroplasts for photosynthesis (Yamasaki et al., 2008); Ribulose-bisphosphate carboxylase (*Pp3c3_14510V3.1*) that catalyzes formation of 3-phosphoglycerate from carbon dioxide and ribulose bisphosphate in the Calvin-Benson cycle; Arogenate dehydrogenase (*Pp3c12_25380V3.1)* that catalyzes oxidative decarboxylation of arogenate leading to biosynthesis of the two aromatic amino acids Tyrosine and Phenylalanine (Rippert and Matringe, 2002a; Rippert and Matringe, 2002b); the glycolytic enzyme Enolase (*Pp3c15_4000V3.1*) that catalyzes synthesis of phosphoenolpyruvate that serves as precursor for biosynthesis of aromatic amino acids in chloroplasts; membrane proteins involved in nitrogen uptake in the form of nitrate/nitrite (nitrate/nitrite porter, NNP family of Major Facilitator Superfamily, MFS; *Pp3c7_13340V3.1*) and the ammonium transporter proteins of AMT family (*Pp3c22_5530V3.1*); Dirigent protein-like (*Pp3c6_6610V3.1*) known to be involved in defense response by activating formation of lignin-like polymers in *P. patens* (Reboledo et al., 2015; **Figure 2B** right panel, **Table S3**).

On the other hand, the top 20 downregulated genes include those encoding the bHLH-MYC transcription factors (*Pp3c6_3180V3.1*) that are possibly involved in plant growth,



B Summary of reads and expressed genes

Features	WT-1	WT-2	WT-3	M8-1	M8-2	M8-3
Raw reads (millions)	30.92	36.22	27.44	31.34	26.74	28.94
Processed reads (millions)	28.03	32.88	24.64	28.46	24.58	26.51
Reads mapped to the reference genome (millions)	25.403589	29.845176	22.126720	25.38834	22.596394	24.280509
Percentage	90.63	90.77	89.8	90.79	91.93	91.59
Number of expressed genes	23763			23916		

FIGURE 1 | (A) Outline of RNA sequencing strategy employed for identification of differentially expressed genes in *PpDNMT2* knockout plants**. (B)** Table summarizing statistics of reads generated in sequencing and their assembly. Scale bar = 1 cm.

development, and stress response, R2R3-MYB transcriptional repressor (*Pp3c6_3186V3.1*) regulating phenylpropanoid pathways and lignin biosynthesis (Ma and Constabel, 2019); ribose-phosphate pyrophosphokinase (*Pp3c8_17920V3.1*), a known component of the multisubunit E3 ubiquitin ligase that forms anaphase promoting complex/cyclosome involved in phytohormone regulation (Yu et al., 2017); voltage-gated calcium channel proteins (*Pp3c7_19620V3.1*); Methylthioadenosine nucleosidase (*Pp3c26_5320V3.1*) that plays key roles in methionine recycling pathway essential in protein and S-adenosyl methionine synthesis and that serves as substrate for ethylene biosynthesis (Bürstenbinder et al., 2010); the Cytochrome P450 (*Pp3c9_7820V3.1*) possibly involved in phytohormone biosynthesis; the antioxidant enzyme peroxidase (*Pp3c12_16500V3.1*) and U5 SnRNP-specific protein (*Pp3c13_21340V3.1*) that functions as E3 ubiquitin-protein ligase RFWD2 that plays an important role in protein degradation *via* the ubiquitin proteasome system (**Figure 2B** right panel, **Table S3**).

To identify functions of other up and downregulated genes, Gene Ontology (GO) analysis was performed, and genes belonging to different categories of Biological Processes, Molecular Functions, and Cellular Compartments were identified (**Figure 3**; **Table S4**). Majority of the upregulated genes were observed to encode membrane proteins with ATP binding, oxidoreductase, metal ion binding activities involved in carbohydrate metabolism, intracellular protein transport, cell wall modification, DNA biosynthetic process, cell redox homeostasis, protein phosphorylation, response to oxidative stress, defense response, response to light stimulus, glycogen biosynthetic process. While, the downregulated genes were those involved in transcription regulation, carbohydrate metabolism, transmembrane transport, Gluconeogenesis, mRNA processing, signal transduction, methylation, cell redox homeostasis (**Figure 3**, **Table S4**).

To understand the biological significance of genes mis-expressed in *ppdnmt2*, GO enrichment analysis was performed. Loss of *PpDNMT2* was observed to significantly affect molecular and physiological processes with GO related to iron ion transmembrane transport, regulation of protein kinase activities, glutamate metabolic process, mitochondrial respiratory chain complex assembly besides others, with endopeptidase inhibitor/regulator activities, Iron ion transmembrane transporter activity, etc to be enriched in the upregulated GO data set (**Figure 4**; **Table S5**), while biological processes that recur with regularity/rhythmicity or processes related to generation and maintenance of such activities,

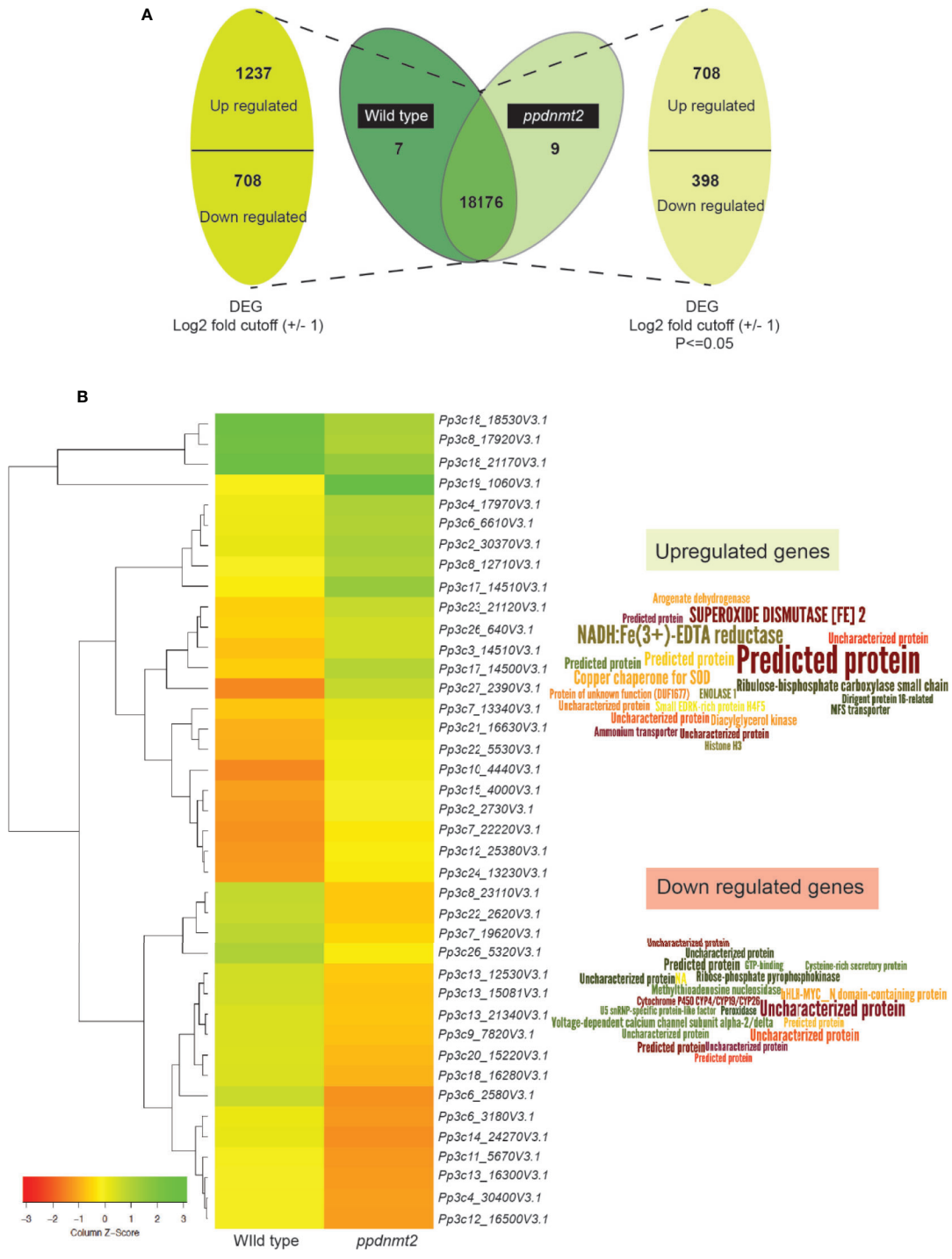


FIGURE 2 | (A) Representation of differentially expressed genes with numbers denoting upregulated/downregulated genes and those expressing in common and exclusively in wild type and *ppdnmt2#8*, respectively. **(B)** Left: Hierarchical clustering analysis of top 20 up and downregulated genes displayed in the form of a heat map where green color shows upregulated and red depicts downregulated genes. Right: Word cloud representation of top 20 differentially expressed genes in *ppdnmt2* where word size corresponds to Log_2 changes in transcript abundance (log_2 foldchange$_{mutant/wild\ type}$). Word clouds have been generated using Wordle.net.in using advanced settings.

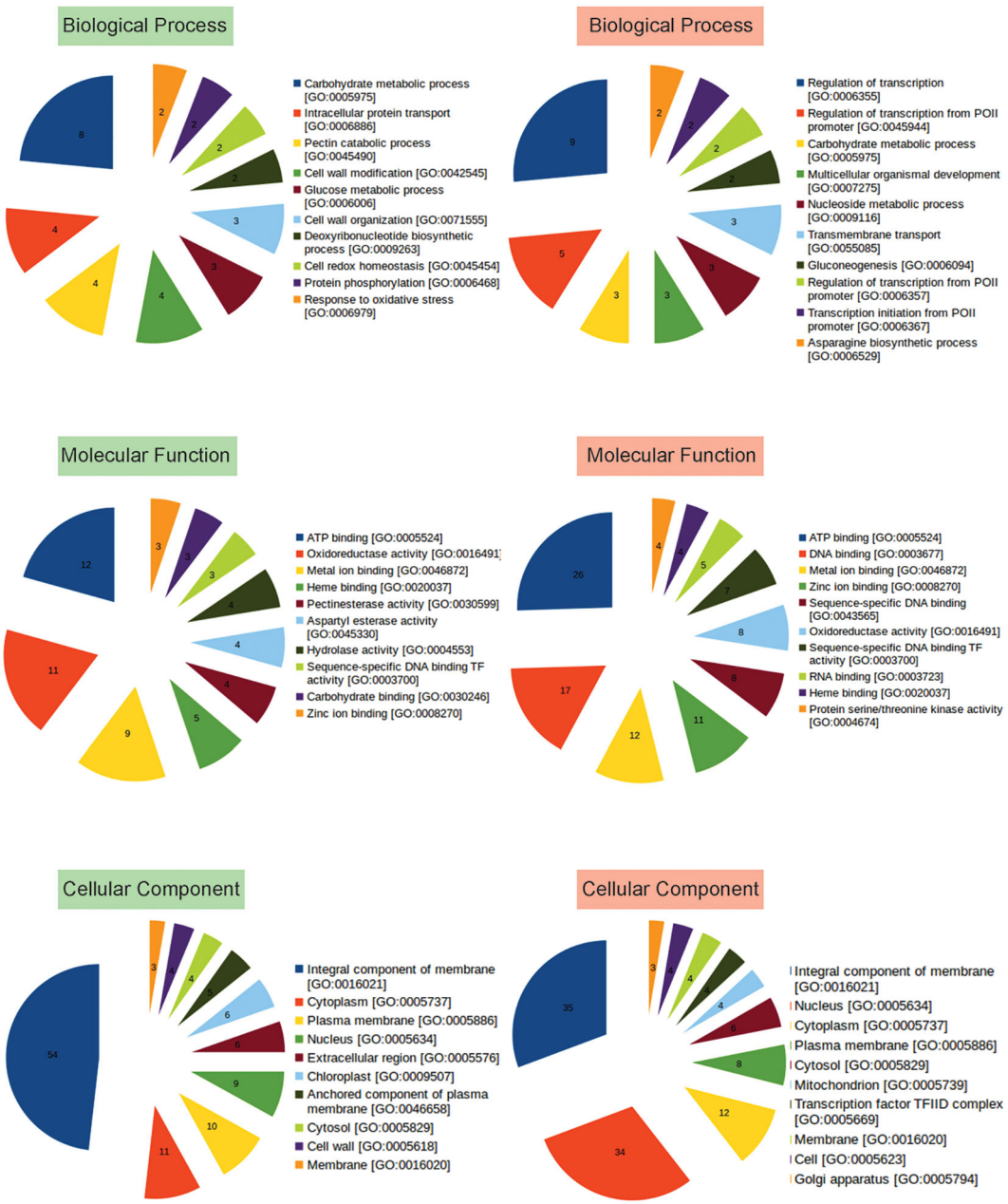


FIGURE 3 | Representation of functional classification of up and downregulated genes into Biological Process, Molecular Function, and Cellular Component categories by GO analysis. Number in each pie represents how many times the function (BP, CC, MF) is expressed (abundance) in the data. GO based categorization of upregulated genes are labeled in green, while those of downregulated genes are labeled in red.

response to jasmonic acid, regulation of signal transduction and cell communication with DNA transcription activity related terms were observed to be enriched among the downregulated GO data set (**Figure 4**; **Table S6**).

Analysis of biological pathways affected by loss of *PpDNMT2* function shows pathways involving transporters, exosome pathways related to RNA degradation/mRNA surveillance, membrane trafficking, phenylpropanoid biosynthesis, starch and sucrose metabolism, pathways involving glycotransferases and chromosome associated proteins to be upregulated. While chromosomes and associated protein pathways; transcription factors, Ubiquitin system, transcription machinery, transporters,

protein phosphatases, peptidases, plant pathogen interaction were downregulated (**Figure 4B**; **Table S7**).

Taken together, analysis of differentially expressed genes and associated pathways shed light on the biological processes affected by loss of *PpDNMT2* function that may correlate with the observed stress sensitivity in *ppdnmt2* mutants (Arya et al., 2016). The above results suggest that *PpDNMT2* may play a crucial role in regulating protein activities spanning diverse biological processes that directly or indirectly contribute towards maintenance of cellular homeostasis at molecular and physiological levels.

PpDNMT2 Regulates Stress Signal Transduction Pathways

Loss of *PpDNMT2* makes the plants sensitive to elevated salt in the growth medium (Arya et al., 2016). To understand the underlying molecular mechanism of *PpDNMT2* function in this process, we analyzed expression of genes encoding regulatory components of pathways known to be affected by salinity in *ppdnmt2*.

Salt stress disrupts ionic equilibrium in cells due to accumulation of excess toxic Na^+ in the cytoplasm and deficiency of essential ions such as K^+. This results in activation of the conserved Salt Overly Sensitive (SOS) pathway that functions to regulate sodium ion homeostasis (Zhu, 2002). Salt stress also activates osmotic stress signaling pathways and Abscisic acid (ABA)-dependent pathway that play key roles in osmotic adjustment and activation of transcription factors that regulate expression of stress-responsive genes necessary for stress tolerance/adaptation (**Figure 5**). Cell wall and plasma membrane play an important role in perception and transmission of stress signals and in plant defence response. Sodium ions induce biochemical changes in cell membranes that lead to alterations in their physical properties (López-Pérez et al., 2009; Moura et al., 2010; Ji et al., 2013; Van der Does et al., 2017). This triggers generation of secondary messengers such as cytosolic Ca^{2+}, Reactive Oxygen Species (ROS) and Inositol phosphates (**Figure 5**). Increased cytosolic Ca^{2+} are sensed by calcium sensors such as the EF (Extra Finger) hand-type calcium binding protein SOS3 that binds to and activates the Serine/Threonine kinase SOS2. SOS2 then phosphorylates the plasma membrane localized Na^+/H^+ antiporter SOS1 and the SOS3-LIKE CALCIUM BINDING PROTEIN 8 (SCaBP8) resulting in activation of the high affinity K^+ channel proteins (such as AKT1) that begin export of Na^+ from the cytoplasm and increase influx of K^+ (**Figure 5**). SOS2 also activates the vacuolar Na^+/H^+ exchanger (NHX) for compartmentalization of Na^+ into the vacuole (Qiu et al., 2004; Batelli et al., 2007). Increased levels of ROS during this process can damage proteins, lipids, DNA, and carbohydrates (Miller et al., 2010). To mediate ROS removal, cells activate both enzymatic and non-enzymatic ROS scavengers. Enzymatic scavengers include SOD, Catalase (CAT), Ascorbate peroxidase (APX), monodehydroascorbate reductase (MDHAR) among others, while the non-enzymatic scavengers include ascorbic acid, alkaloids, carotenoids, and other phenolic compounds (Zhu, 2002; Yang and Guo, 2018).

Transcriptome landscape of *ppdnmt2* reveals many genes involved at different levels of stress signal transduction pathways to be differentially expressed. At the level of cell wall, genes involved in cell wall biosynthesis and maintenance such as those encoding cellulose synthase, expansins, cross-linking glycans (Xyloglucan : Xylosyl transferases), glycoproteins (Arabinogalactans), and pectinesterases were observed to be upregulated (**Figure 6**; **Table S3**). While at the plasma membrane, genes involved in lipid metabolic process, lipid transport, fatty acid biosynthetic process, diacylglycerol were upregulated while those encoding lipid binding proteins were downregulated in *ppdnmt2* (**Figure 6**; **Table S3**). Among the genes encoding Ca^{2+} ion binding proteins, calmodulin encoding genes (*Pp3c25_1480V3.1* and *Pp3c14_8590V3.1*) were observed to be upregulated while SOS3 (*Pp3c13_5410V3.1*), the key regulator of the SOS pathway was downregulated in *ppdnmt2*. The Leucine-rich repeat receptor-like protein kinases (LRR-RLK) are membrane localized receptor proteins that sense and transduce signals downstream by autophosphorylation followed by phosphorylation of specific substrates. They are known to play crucial roles in stress response and plant development (Liu et al., 2017). Out of 119 genes encoding LRR-RLK in *P. patens*, eight were upregulated in *ppdnmt2* (*Pp3c6_23170V3.1*, *Pp3c1_18260V3.1*, *Pp3c7_22250V3.1*, *Pp3c1_25110V3.1*, *Pp3c25_12800V3.1*, *Pp3c5_19200V3.1*, *Pp3c4_19720V3.1*), while CPK13 homolog of calcium-dependent protein kinase (*Pp3c7_22440V3.1*) known to inhibit $K(^+)$ channel proteins KAT2 and KAT1 in Arabidopsis was downregulated (Ronzier et al., 2014; Liu et al., 2017). Expression of genes encoding Fe-SOD (*PpFeSD3*; *Pp3c17_14510V3.1*) was significantly upregulated along with those encoding peroxidases (*Pp3c1_19610V3.1*, *Pp3c16_10060V3.1*; *Pp3c12_19290V3.1*, *Pp3c26_2960V3.1*, *Pp3c19_20780V3.1*) and the non-enzymatic ROS scavenger ascorbic acid, monodehydroascorbate reductase (MDHAR, *Pp3c2_8410V3.1*) and Glutathione S-transferase (GST, *Pp3c22_5470V3.1*), while those encoding lactoylglutathione lyase glyoxalase I (*Pp3c24_12430V3.1*) were downregulated in *ppdnmt2* (**Figure 6**; **Table S3**).

Genes involved in maintaining ion homeostasis were also observed to be differentially expressed in *ppdnmt2*. Genes encoding copper ion transmembrane transporter (solute carrier family 31; *Pp3c27_4780V3.1*) and the copper chaperone for SOD (*Pp3c17_14500V3.1*) and ATOX1 chaperone (*Pp3c14_1600V3.1*) were observed to be significantly upregulated in *ppdnmt2*. Expression of three out of 22 Plastocyanin genes encoded in *P. patens* genome (*Pp3c3_25110V3.1*, *Pp3c6_3530V3.1*, *Pp3c16_22330V3.1*) that require copper for their activity were also observed to be upregulated in *ppdnmt2*. This suggests that *PpDNMT2* may play a crucial role in Cu^{2+} trafficking pathway from its point of entry in the plasma membrane to its delivery in the chloroplasts. Further, four genes encoding Aquaporins (*Pp3c10_21310V3.1*, *Pp3c11_23630V3.1*, *Pp3c3_3230V3.1*), the membrane bound channel proteins that facilitate transport of water and small solutes across plasma membrane; genes encoding potassium channel (AKT1-related, *Pp3c2_28940V3.1*) proteins were also observed to be upregulated in *ppdnmt2*.

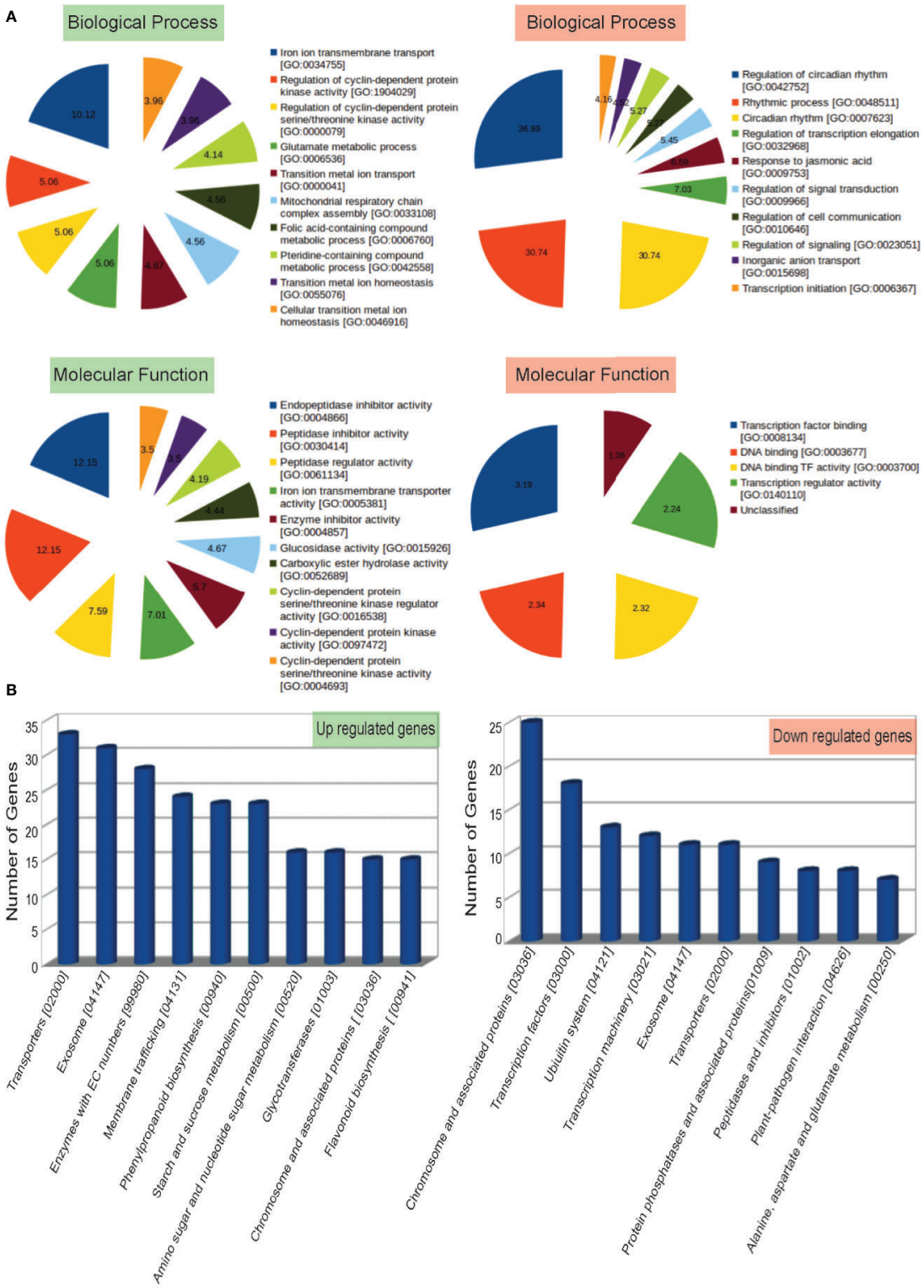


FIGURE 4 | (A) Top 10 Gene Ontology terms of Biological Process (BP) and Molecular Functions (MF) enriched in up (labeled in green) and downregulated (labeled in red) genes. Enriched BP and MF with p value < 0.05 and with an enrichment factor > 1 are shown. Numbers represent fold change enrichment value of a particular function in each dataset. **(B)** Graphical representation of up and downregulated pathways (X-axis) with number of genes identified in each pathway category plotted on the Y-axis.

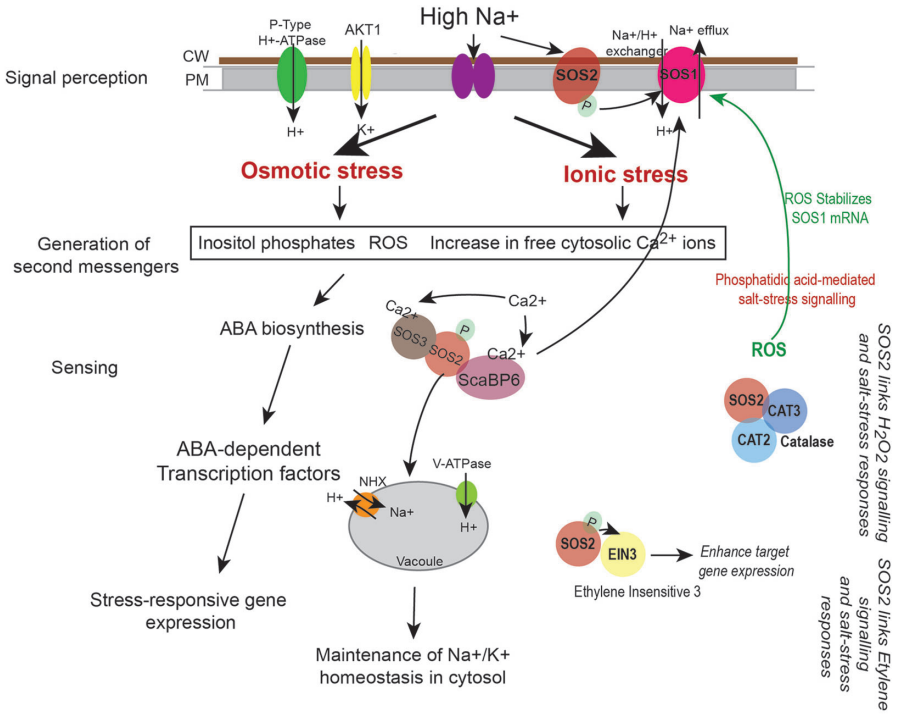


FIGURE 5 | Regulation of sodium (Na^+) and potassium (K^+) ion homeostasis by SOS pathway. High Na^+ ions induce osmotic and ionic stress in cells. This leads to increase in levels of cytosolic Ca^{2+}, generation of reactive oxygen species (ROS), and increase in Inositol phosphates. Perturbations in Ca^{2+} levels is sensed by calcium binding proteins such as SOS3 and ScaBP6 that activate the protein kinase SOS2 thus initiating a phosphorylation cascade. This leads to activation of the plasma membrane localized Na^+/H^+ exchanger SOS1 (that effluxes Na^+ from the cytoplasm) and the NHX exchanger on the vacuoles (that accumulate Na^+ in the vacuole) thus leading to maintenance of ion homeostasis in cells. Salt stress response is also linked to hydrogen peroxide (H_2O_2) signaling as SOS2 directly interacts with CATALASE (CAT2 and CAT3), while its interaction with Ethylene Insensitive 3 (EIN3) links SOS pathway to ethylene signaling processes (Verslues et al., 2007; Quan et al., 2017). ROS and stress responsive changes in gene expression may also lead to biosynthesis of phytohormones such as ABA that can amplify the stress signal by activating ABA-dependent transcription factors that lead to expression of stress-responsive genes.

Salt stress is also known to increase endocytosis and active vesicle movement that leads to Na^+ accumulation in vacuoles (Hamaji et al., 2009). Genes encoding proteins regulating membrane vesicle trafficking such as Clathrin (*Pp3c13_20171V3.1*), ADP-ribosylation factor 1 (ARF1, *Pp3c10_25070V3.1*, *Pp3c1_20430V3.1*), Secretory Carrier-Associated Membrane Proteins (SCAMPs, *Pp3c5_14240V3.1*), Soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNARE, *Pp3c18_17330V3.1*) were observed to be upregulated in *ppdnmt2* (**Figure 6**; **Table S3**).

Cytoskeleton dynamics regulated by microtubules, microfilaments, and their associated proteins also play crucial roles in stress tolerance in plants. Depolarization of microtubules followed by their repolarization is critical for stress tolerance (Wang et al., 2007). Loss of *PpDNMT2* function results in upregulation of Microtubule-localized SPIRAL1 protein encoding gene (*Pp3c22_7610V3.1*; **Figure 6**, **Table S3**). These proteins are known to be required for anisotropic root growth in rapidly elongating cells in Arabidopsis (Shoji et al., 2004; Nakajima et al., 2006).

A number of genes encoding transcription factors that affect stress responsive gene expression and other related functions were also observed to be differentially expressed in *ppdnmt2*. Genes encoding the APETALA2/ETHYLENE-RESPONSIVE FACTOR (AP2/ERF) were both up and downregulated (upregulated: *Pp3c21_13130V3.1*, *Pp3c6_26080V3.1*, *Pp3c8_7340V3.1*, *Pp3c4_23930V3.1*, *Pp3c5_2900V3.1*; down: *Pp3c22_1800V3.1*, *Pp3c2_11580V3.1*, *Pp3c25_4250V3.1*, *Pp3c19_14950V3.1*, *Pp3c4_24710V3.1*). AP2/ERF forms a large group of plant-specific transcription factors that activate genes in response to ABA, cold, drought, and salinity. They also regulate chloroplast division under salt stress in *P. patens* (Mizoi et al., 2012; Do et al., 2020). Plant-specific Plant AT-rich sequence-and zinc-binding transcription factor (PLATZ) transcription factors were also differentially expressed (upregulated: *Pp3c13_4760V3.1*; Downregulated: *Pp3c15_23190V3.1*, *Pp3c10_12560V3.1*). These proteins function in RNA polymerase III transcription machinery to regulate biogenesis of tRNA and 5S rRNAs in maize (*FL3*) and rice (*GL6*) that affects endosperm storage filling (Li et al., 2017) and rice grain length and number (Wang and Schippers, 2019). Two out of 22 small Heat Shock proteins (sHSPs) encoded in *P. patens* genome were also differentially expressed (upregulated: *Pp3c8_6770V3.1* and

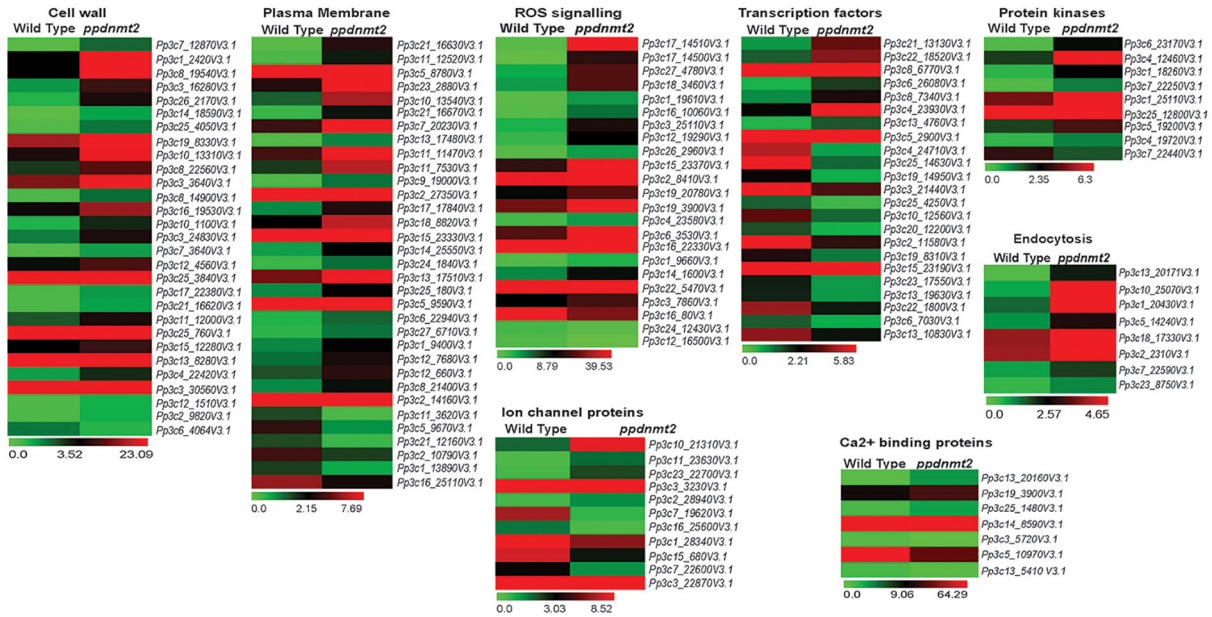


FIGURE 6 | Expression patterns of genes encoding proteins involved in stress sensing at cell wall and plasma membrane, ion homeostasis regulation, ROS signaling, cytoskeletal dynamics, and those encoding transcription factors. Average FPKM values of genes in wild type and *ppdnmt2* are presented by cluster display (heat map) with color scale representing range of expression values shown at the bottom of each heat map.

Pp3c25_14630V3.1: downregulated). sHSPs are known to prevent protein aggregation and facilitate protein refolding by chaperons during stress and are required for stress recovery in *P. patens* (Ruibal et al., 2012). *Pp3c22_18520V3.1* encoding the sigma factor (SigE) of the plastid RNA polymerase, a multisubunit bacteria-type enzyme in chloroplasts was observed to be upregulated in *ppdnmt2*. The plastid sigma factors are known to integrate clock and light regulatory mechanisms that underly daily expression patterns of plastid genes in *P. patens* (Ichikawa et al., 2004). Three genes (*Pp3c13_10830V3.1*, *Pp3c20_12200V3.1* and *Pp3c3_21440*) encoding WRKY transcription factors were observed to be downregulated in *ppdnmt2*. These transcription factors form important components of stress signaling cascades and signaling in response to internal developmental cues in plants (Bakshi and Oelmüller, 2014). *PpDNMT2* may also affect plant-specific phytohormone signaling responses. Genes involved in cytokinin metabolism (degradation, activation, signaling; *Pp3c6_7030*, *Pp3c23_17550*, and *Pp3c13_19630*); DELLA protein encoding genes (*Pp3c19*_8310) known to function as negative regulators of gibberellin signaling and in integration of multiple hormone signaling pathways were observed to be downregulated (Davière and Achard, 2016) (**Figure 6**; **Table S3**). Differential expression of genes involved in phytohormone signal transduction may explain the abnormal branching pattern observed in *ppdnmt2* (Arya et al., 2016).

The above results therefore suggest that *PpDNMT2* may play a pivotal role in maintaining ion homeostasis by regulating genes involved in stress signaling pathways and in plant development by affecting genes involved in phytohormone signaling in *P. patens*.

Identification of Proteins Co-Immunoprecipitating With PpDNMT2

To gain insight into the interactome of PpDNMT2, immunoprecipitation coupled with Mass spectrometry (IP-MS) was performed. First, to check specificity of the PpDNMT2 antibody, a western blotting was performed using equal amounts of total protein extracted from wild type and *ppdnmt2* protonemata tissue and probed with anti-PpDNMT2-HRP conjugate. A single protein band of expected size (~51 kDa) corresponding to PpDNMT2 was detected in wild type samples and not in *ppdnmt2* indicating the antibody prepared was specific for PpDNMT2 (**Figure 7A**). Next, to select the conditions under which *PpDNMT2* is optimally expressed, we took note of the fact that among all the cytosine methyltransferases expressing in protonema, *PpDNMT2* expression has been observed to be the weakest (Malik et al., 2012). We have also previously shown that *PpDNMT2* is differentially expressed under salt and mannitol stress with transcript levels decreasing under stress conditions (Arya et al., 2016). To analyze the expression at protein level under stress, we performed western analysis using total protein extracted from control protonema (non-stress) and protonema exposed to 24 h salt stress (referred to as treated from here on) and probing with anti-PpDNMT2-HRP. Low levels of PpDNMT2 were detected in both control and treated samples. To assess the levels of endogenous PpDNMT2 more clearly we enriched the target

protein by affinity purification in each sample type and then performed western blotting. Strong induction in PpDNMT2 expression was observed in samples harvested after 24 h stress treatment in comparison to control (**Figure 7B**). Hence, we selected 0 and 24 h stress as the time points for protein extraction and affinity purification. The interacting protein partners in the immune complex were then identified by LC-MS. IP procedure with beads devoid of antibody was not included in this study.

Following the experimental pipeline outlined in **Figure 7C**, a total of 101 proteins were identified in control out of which 98 were present in all three biological replicates. Similarly, a total 188 proteins were identified in treated samples out of which 121 were common in all the three replicates (**Figure 7D**). Strong correlation among the biological replicates of control and treated samples with Pearson correlation coefficient >= 0.8 among C1–C3 (control) and >=0.7 between S1–S3 (treated) was observed (**Figure S1**). Heat map generated using abundance values of P-significant proteins show differential accumulation of proteins under control and stress conditions. Out of 98 and 121 proteins identified in the control and treated samples respectively, 69 were common to both control and treated while 32 were present only in the control and 119 only in the stress samples. After statistical analysis applying p value <= 0.05 and fold change >= 2 cut-off, 169 proteins were shortlisted of which 64 proteins were observed to be present in >=2 fold abundance while 105 proteins below this cut-off in treated sample (**Figure 7D**, **Table S8**).

To gain insight into biological processes, molecular function and cellular localization of proteins co-immunoprecipitated with PpDNMT2, GO analysis was performed. GO terms with photosynthesis, protein-chromophore linkage (Biological process) and chlorophyll binding, ATP binding (Molecular function) affiliation were observed to be enriched. Though annotations for a number of proteins were not available in public databases, on the basis of available information, it was observed that majority of the proteins were integral components of membranes/chloroplast thylakoid membranes involved in photosynthesis, response to light stimulus, carbohydrate metabolism, photorespiration, glycolytic process with chlorophyll, or ATP binding activities (**Figure 8A**). High abundance of proteins related to chloroplast and photosynthesis is most likely a reflection of abundant chloroplasts in chloronema tissue used for proteome analysis.

PpDNMT2 Regulates Ubiquitin-26S Proteasome Degradation Pathway and Epigenetic Regulatory Pathway Genes

Comparison between the gene set affected by loss of *PpDNMT2* and the co-IP protein data reveals novel roles for *PpDNMT2* in *P. patens*. Among the proteins identified in PpDNMT2 immunocomplex under both control and salt stress was RPT4, one of the six AAA-ATPases of the 19S regulatory particle of the Ubiquitin-26S proteasome system that regulates recognition and unfolding of substrates during protein degradation (**Table S8**). Our transcriptome dataset also revealed genes of the ubiquitin-26S degradation pathway to be differentially expressed. Genes encoding the ubiquitin-conjugating enzyme E2C (*Pp3c6_12520V3.1*), E3 ubiquitin ligase MIEL1 (*Pp3c14_5600V3.1*), proteins with ubiquitin transferase (*Pp3c24_4740V3.1*, *Pp3c19_19700V3.1*, *Pp3c1_8490V3.1*) and ligase (*Pp3c2_34310V3.1*) activities, a subunit of the 19S regulatory particle involved in substrate recognition and unfolding (Rpn12) and the 26S proteasome regulatory subunit N12 (*Pp3c14_17630V3.1*) were upregulated, while the E3 ubiquitin-protein ligase MARCH6 (*Pp3c15_7960V3.1*) was downregulated in *ppdnmt2* (Wang and Schippers, 2019; **Figure 8B** left panel; **Table S3**). To experimentally check if loss of *PpDNMT2* function does affect protein degradation/senescence in response to salt stress, wild type and *ppnmt2* were propagated in BCDAT media supplemented with high NaCl under similar growth conditions. Early senescence in *ppdnmt2* (within 6 days) observed by visual discoloration of *ppdnmt2* protonema in comparison to wild type plants was observed indicating low stress tolerance of *ppdnmt2* plants (**Figure 8B**, right panel). This suggests a role for *PpDNMT2* in sensing/maintaining cellular proteostasis by affecting expression of Ubiquitin-26S proteasome pathway genes.

In the co-IP protein set, four histone 4 proteins (*Pp3c27_7460V3.1*, *Pp3c14_26430V3.1*, *Pp3c22_22190V3.1* and *Pp3c21_19690V3.1*) known to be core components of nucleosomes were identified with high confidence in treated samples (**Table S8**). Our transcriptome data set also revealed that genes encoding proteins involved in histone modification are differentially expressed. Three genes encoding histone methyltransferases were downregulated in *ppdnmt2*. These include *Pp3c24_20920V3.1* encoding the SET domain containing histone lysine N methyltransferase and *Pp3c19_9880V3.1 and Pp3c22_17280V3.1* encoding the homolog of Arabidopsis *SUVR5* that mediates H3K9me2 methylation and is required for transcriptional silencing (Caro et al., 2012). The homolog of Arabidopsis transcriptional repressor *SIN3* (a component of Histone deacetylase complex) encoded by *Pp3c24_13770V3.1* was also downregulated in *ppdnmt2*, while *Pp3c3_10480V3.1* encoding the RNA-dependent DNA polymerase was upregulated. It was also observed that *Pp3c20_2710V3.1* encoding gag protein was activated in *ppdnmt2* (**Figure 8C**; **Table S3**). Transcription of GAG structural proteins encoded by the long terminal repeat (LTR) copia-type retrotransposon is the first/obligatory step required for mobility and transposition of repetitive elements. Hence, transcription of *gag* in *ppdnmt2* and not in wild type suggests active transposition of this class of retrotransposons in cells and a role for *PpDNMT2* in silencing LTR copia type retrotransposons. Our previous observation that PpDNMT2 localizes both in the nucleus and the cytoplasm further supports its role in epigenetic silencing of TE observed in this study (Malik et al., 2012). Taken together, our proteomic and transcriptome data suggests a key role for *PpDNMT2* in chromatin and epigenetic gene regulation in *P. patens*.

PpDNMT2 Interacts With Superoxide Dismutase

To shortlist putative interactors of PpDNMT2 for protein interaction validation we noted a close link between PpDNMT2 and SOD from our transcriptome and interactome studies. Our transcriptome data shows strong upregulation of *PpFSD3* (*Pp3c17_14510V3.1*) encoding Fe-SOD in *ppdnmt2* (5.9 folds). Target P2.0 (Emanuelsson et al., 2000) and ChloroP1.1

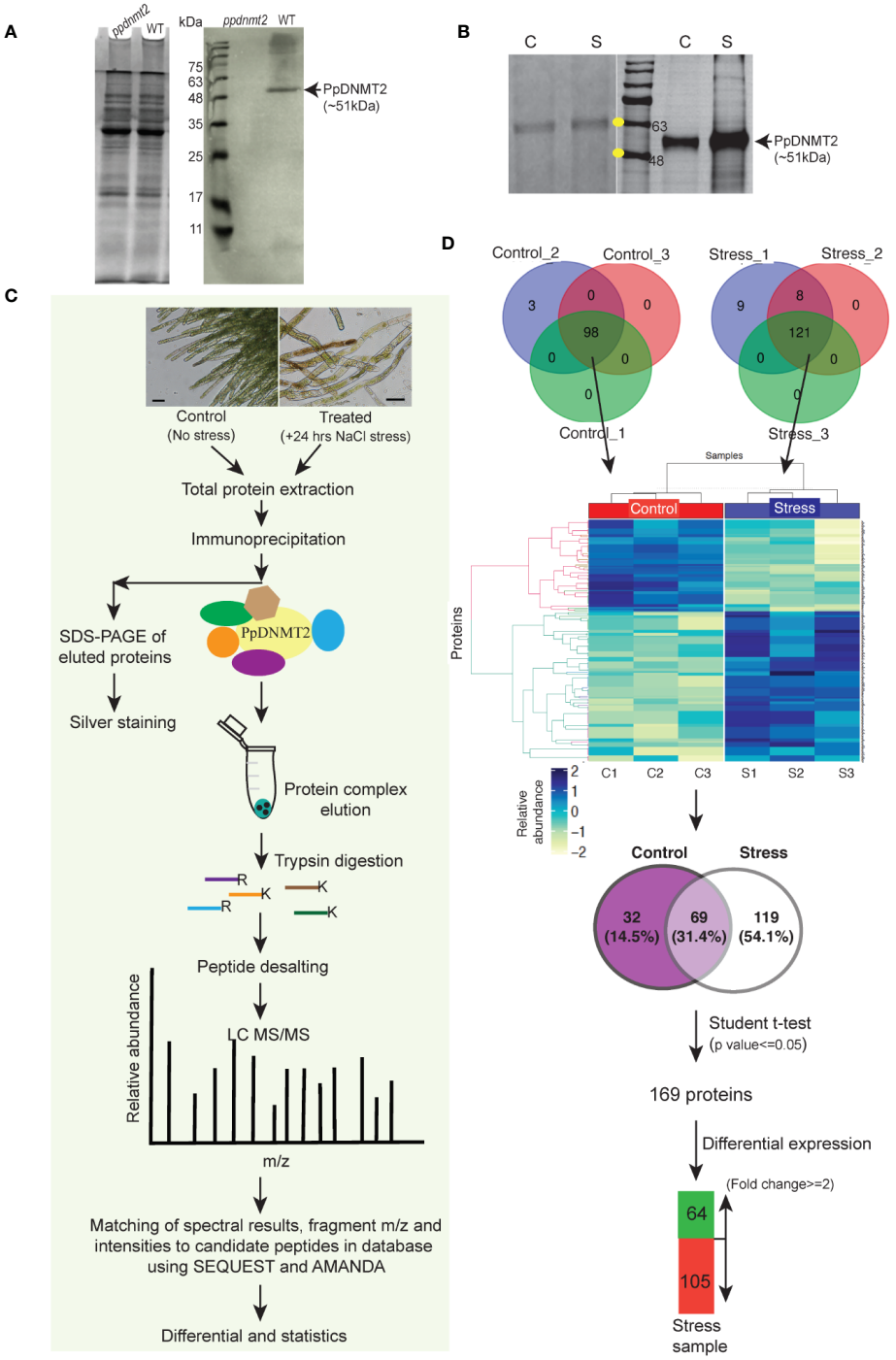


FIGURE 7 | (A) (Left) Coomassie stained image of polyacrylamide gel showing fractionation of equal amounts of total protein (50 µg each) extracted from *ppdnmt2* and wild type protonema tissue. (Right) Western blot probed with PpDNMT2-specific antibody showing presence of PpDNMT2 protein band of expected size in wild type plants and its absence in *ppdnmt2*. **(B)** (Left) Silver stained image of polyacrylamide gel showing proteins isolated from control (C, 0 h) and stress-treated, (S) protonema tissue. (Right) Immunoblot of proteins probed with PpDNMT2-specific antibody. Arrow points to the band position of ~51 kDa PpDNMT2, and the yellow dots on the marker bands in the protein ladder indicate 48 and 63 kDa proteins, respectively. **(C)** Outline of steps followed to immunoprecipitate proteins in complex with PpDNMT2 followed by their identification by LC-MS/MS. Scale bar = 50 µm. **(D)** Summary of analyzed data. Venn diagram on top shows number of proteins identified in each of the three biological replicates of control and stress samples. Distribution of these proteins in both samples before and after application of Student t-test. Heat map showing differential abundance of 169 proteins in the Control (C1–C3) and Stress (S1–S3) samples. For generating heat map abundance, values were log2 transformed and z-normalized. Hierarchical clustering was then performed using Euclidian distance and average linkage using in-House R script (Package: ComplexHeatmap). Pale yellow and blue colors, respectively, represent low and high protein abundance.

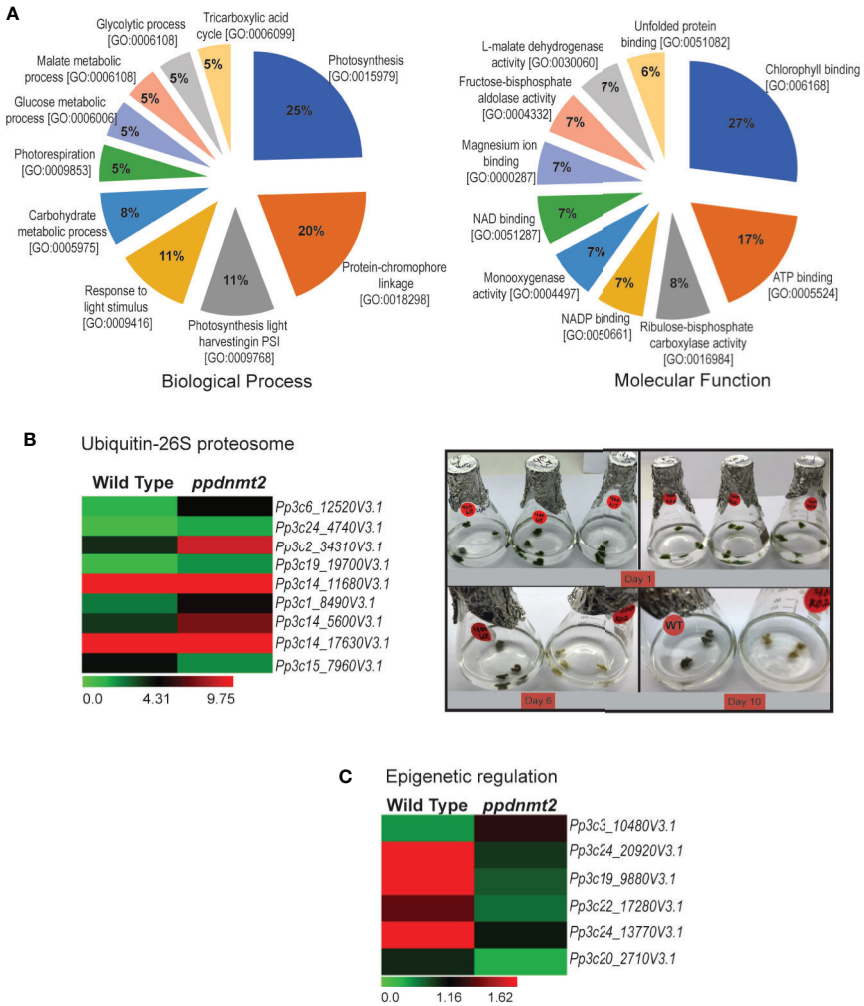


FIGURE 8 | (A) Pie charts depicting top 10 Biological Process and Molecular Function Gene Ontology categories to which P-significant (p value <= 0.05) proteins co-immunoprecipitated with PpDNMT2 can be classified. **(B)** Left: Differential expression pattern of genes involved in Ubiquitin-26S proteasome in wild type and *ppdnmt2* presented by cluster display with color scale representing FPKM expression values; Right: Wild type (WT) and *ppdnmt2* (KO#2) protonemata propagated in BCDAT media supplemented with 400 mM NaCl under similar growth conditions for 10 days to analyze stress tolerance of plants lacking *PpDNMT2* function. **(C)** Differential expression of genes in wild type and *ppdnmt2* involved in epigenetic regulation displayed in the form of a heat map with color scale representing range of FPKM values of the selected genes.

(Emanuelsson et al., 1999) servers predict PpFSD3 to be localized to the chloroplast. Interestingly, among the proteins existing in PpDNMT2 immunocomplex, four CuZn-SODs were identified in the treated samples (**Table S8**). Two of these CuZn-SODs, PpCSD1 (accession#A9SX31; *Pp3c9_24840*) and PpCSD2 (accession# A9SX65; *Pp3c9_25690*) have been previously described (Higashi et al., 2013). Expression of genes encoding PpCSD1 and PpCSD2 is known to be repressed under copper deficient conditions. Downregulation of CuZn-SODs is known to be functionally compensated by increase in Fe-SOD levels as an adaptive strategy in higher plant chloroplasts to mitigate harmful effects of ROS under copper limiting conditions in chloroplasts (Yamasaki et al., 2008). These observations suggested a close link between PpDNMT2 and the antioxidant enzyme activities. To strengthen this hypothesis we performed directed yeast two hybrid assay to study interaction between PpDNMT2 and PpCSD2 (*Pp3c9_25690V3.1*) in yeast nucleus.

Complementary DNA (cDNA) encoding PpCSD2 was fused with GAL4 DNA binding domain (BD) and used as bait, while PpDNMT2 fused with GAL4 Activation domain (AD) was used as prey. Co-expression of the fusion proteins in yeast and growth of colonies on selective media lacking leucine, tryptophan, histidine (SD/LWH), and Adenine (SD/LWHA) indicated interaction between the bait and prey proteins *in vivo* (**Figure 9**). Further validation was performed by streaking the positive colonies on quadruple dropout media supplemented with the chromogenic

substrate X-α-Gal (SD/LWHA X-α-Gal). Development of blue color in colonies indicated activation of the endogenous reporter *MEL1* by the reconstituted GAL4 as a result of interaction between PpDNMT2 and PpCSD2. SV40 large T-antigen (RecT) co-expressed with human Lamin C did not show interaction as expected and was used as negative control, while growth of colonies harboring RecT co-expressed with p53 confirmed the known interaction between RecT and p53 and was used as positive control. AD-PpDNMT2 co-expressed with empty BD-vector and BD-prey proteins co-expressed with empty AD-vectors did not result in any yeast growth and were used as controls for monitoring specificity of interaction. Strength of interaction between PpDNMT2 and PpCSD2 was also assessed by adding optimized concentration (15 mM) of 3-AT (3-amino-1,2,4-triazole), a competitive inhibitor of the His3 protein in SD-LWH selective media. Optimum growth of yeast co-expressing PpDNMT2 and PpCSD2 fusion proteins on 3-AT supplemented selective media indicated strong interaction between the two proteins (**Figure 9**).

The above results therefore show that PpDNMT2 can exist in complex with CuZn-SOD *in vivo*, and together these proteins may play a pivotal role in navigating cellular response during stress recovery.

DISCUSSION

The biological significance and molecular mechanism of cytosine DNA/tRNA methyltransferase *DNMT2* have remained elusive in land plants. In this study, an attempt has been made to understand the molecular mechanism of *PpDNMT2* function that is crucial for stress recovery in *P. patens* (Arya et al., 2016). We have analyzed the transcriptome of *PpDNMT2* knockout mutant (*ppdnmt2#8*) and identified proteins present in the PpDNMT2 immunocomplex by mass spectrometry after exposing wild type plants to salt stress conditions.

Among the cell organelles affected by salinity, chloroplasts are the most sensitive. These plastids are not only sites for photosynthetic activity, but they also participate in many metabolic processes including biosynthesis of lipids and fatty acids, aromatic amino acids and reduction of nitrites and sulphates (Baginsky and Gruissem, 2004). Due to the high rate of oxidizing metabolic activities and increased rate of electron flow during photosynthesis, chloroplasts are highly prone to ROS production. Further, salt stress also induces production of ROS. Hence, chloroplasts experience significant ROS-mediated damages under stress. Accordingly, the expression of many nuclear genes targeted to chloroplasts is known to be finely tuned under salt stress in flowering plants (Zörb et al., 2009; Fan et al., 2011; Wang et al., 2014; do Amaral et al., 2016). Also, in *P. patens* photosynthesis in leafy gametophores is known to increase under salt stress to reduce negative effects of salinity (Wang et al., 2008). The co-IP data set generated from stress-treated chloroplast-rich protonema tissue show enrichment of proteins involved in thylakoid membrane organization, PSII activity, CO_2 assimilation, ROS scavenging, maintenance of ionic and osmotic homeostasis and protein turnover in PpDNMT2 immunocomplex. This suggests that in protonema cells too photosynthetic activity and the antioxidative enzyme activity are elevated to enhance stress tolerance and protection from oxidative damage though the possibility of non-specific contamination from chloroplast proteins due to their abundance in the chloroplast-rich tissue used for protein extraction cannot be ruled out. Our transcriptome data also shows upregulation of genes encoding ROS scavenging proteins (*Pp3c9_25690v3.1*, ~5.9 folds) in *ppdnmt2* in comparison to wild type plants propagated under similar conditions of continuous light. Continuous light conditions in some plants are known to affect photosynthesis rate and photochemical efficiency of PSII leading to increase in SOD activity (Haque et al., 2015). The tissue used for transcriptome analysis and proteomic studies was propagated on solid and in liquid media, respectively but these contained common moss media supplements. Hence, upregulation of genes encoding SOD in *ppdnmt2* and identification of SODs in PpDNMT2-immunocomplex under salt stress suggest an important role for *PpDNMT2* in stress mitigation pathways involved in minimizing ROS-mediated damages in *P. patens*.

P. patens genome encodes eight SOD isozymes that include four CuZn-SODs (PpCSD1–4), three Fe-SODs (PpFSD1–3) and one Mn-SOD (PpMSD) that are localized in the chloroplasts (PpCSD1, PpCSD2, PpFSD3), cytosol (PpCSD3, PpCSD4), mitochondria (PpMSD), and the apoplast (PpFSD1) (Higashi et al., 2013). The expression of genes encoding these metalloenzymes is known to be co-ordinately regulated depending on the bioavailability of metal ions. The CuZn-SOD and Fe-SOD encoding genes are known to express differentially under copper deficient conditions with expression of *PpCSD1* and *PpCSD2* repressed while that of *PpFSD3* strongly induced under copper limiting conditions in *P. patens* (Higashi et al., 2013). Copper is essential for activity of chloroplast electron carrier protein Plastocyanin that catalyzes transfer of electrons from the $Cytb_6f$ complex to PS1 in the Z-scheme of photosynthesis (Gorman and Levine, 1965). Chloroplasts are known to sustain photosynthesis under low copper conditions by downregulating CuZn-SOD and increasing levels of copper chaperone proteins to enable transfer of limited copper to plastocyanin in chloroplasts (Abdel-Ghany et al., 2005; Nagae et al., 2008). Further, to minimize ROS-mediated damage under these conditions, a switch from CuZn-SOD to Fe-SOD occurs that leads to increase in levels of Fe-SOD activity. Our proteome and transcriptome data suggest a similar adaptive mechanism to be functional in *P. patens* as well. We observe low levels of CuZn-SOD (*Pp3c9_24840V3.1*, *Pp3c9_25690V3.1*) in PpDNMT2 immunocomplex under salt stress. Further, *ppdnmt2* transcriptome shows increased levels of *Pp3c17_14500V3.1*, *Pp3c27_4780V3.1*, and *Pp3c17_14510V3.1* encoding the copper chaperone for SOD, the copper ion transmembrane transporter, and the chloroplast localized Fe-SOD. These observations suggest that *PpDNMT2* may play a crucial role in modulating SOD activities in response to copper availability in chloroplasts. We also show PpDNMT2 to interact with PpCSD2 in yeast nucleus. GFP-PpDNMT2 is known to be distributed both in the nucleus and cytoplasm in moss protonema cells (Malik et al.,

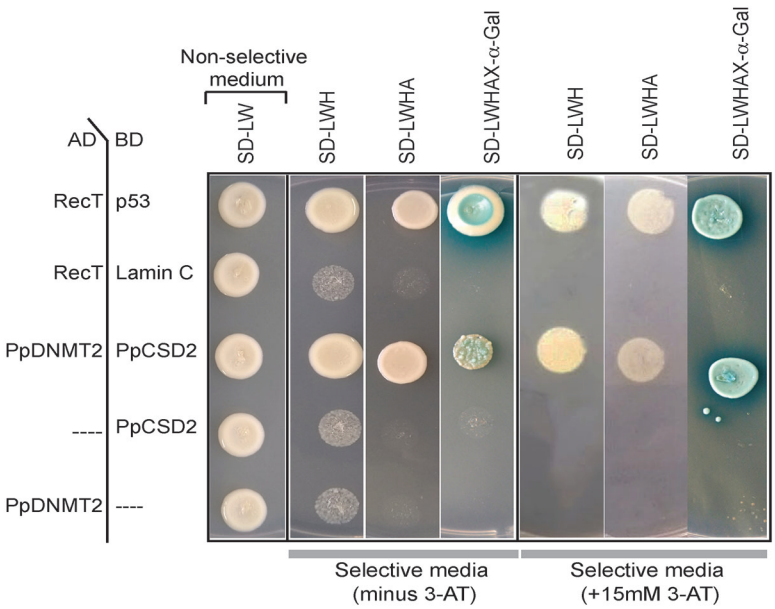


FIGURE 9 | Yeast two-hybrid assay showing direct interaction between PpDNMT2 and PpCSD2. PpDNMT2 and the SV40 large T antigen (RecT) are expressed as GAL4-AD fusion protein while p53, human Lamin C, and the bait protein (PpCSD2) are expressed as GAL4 DNA binding domain (BD) fusion proteins. Non-selective and Selective synthetic dropout media lacking leucine (L), Tryptophan (W), Histidine (H) and Adenine (A) are mentioned above.

2012). While PpCSD2 is predicted to be localized to the chloroplasts, no chloroplast transit peptide sequence is present in PpDNMT2. Since PpCSD2 was identified in PpDNMT2 immunocomplex under stress conditions, it is plausible that PpDNMT2 may get localized to the chloroplasts under stress, and the two proteins may interact in the plastids. In animal systems, GFP-Dnmt2 is known to shuttle between the nucleus and the cytoplasm in response to cellular stress. Dnmt2 is normally present in the nucleus but under stress conditions it re-localizes to cytoplasmic granules and RNA processing bodies (Thiagarajan et al., 2011).

Transcriptome profile also shows *ppdnmt2* cells to accumulate increased levels of apoptosis associated protein encoding genes and transcripts encoding Ubiquitin-26S proteasome components, while genes affecting telomere integrity are downregulated along with genes involved in signal transduction, cell–cell communication and intracellular protein transport. Though the levels of changes in gene expression is low or below threshold, the transcript signatures in *ppdnmt2* reflect a molecular environment that provides a possible explanation for early senescence and increased sensitivity of *ppdnmt2* under salt and osmotic stress. The molecular mechanism of how PpDNMT2-Fe-SOD complex allows plants to tolerate stress is not known yet. It is plausible that Fe-SOD may modulate tRNA methylation activity of PpDNMT2 to stabilize tRNAs and prevent their fragmentation under stress as has been shown in Drosophila (Schaefer et al., 2010). In the lower eukaryote Entamoeba the glycolytic protein Enolase is known to bind to DNMT2 (Ehmet) and inhibit its $tRNA^{Asp}$ methylation activity under glucose starvation (Tovy et al., 2010).

Among the RNA methyltransferases in plants *TRM4B* loss-of-function mutants in Arabidopsis are known to be sensitive to oxidative stress. They also lack m5C sites on mRNA and non-coding RNAs and have reduced $tRNA^{Asp(GTC)}$ stability (David et al., 2017). *PpDNMT2* loss-of-function mutants are also sensitive to salinity, and *PpDNMT2* may modulate activity of antioxidant proteins under stress as inferred from this study. We had also previously reported that transcription/stability of $tRNA^{Asp(GTC)}$ is highly reduced under stress in *ppdnmt2*. This suggests that the stability of $tRNA^{Asp(GTC)}$ mediated by RNA methyltransferase activities is critical for stress tolerance in both Arabidopsis and *P. patens* thus indicating this to be an evolutionarily conserved mechanism in land plants.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. NCBI SRA BioProject ID: PRJNA514203.

AUTHOR CONTRIBUTIONS

DS, RY, and NW conducted wet lab work, analyzed transcriptome data, and wrote the first draft of the manuscript. SKau analyzed the interactome data. MK and SKap planned the experiments. MK acquired financial support and finalized the manuscript with RY, DS, NW, SKau, and SKap. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by funds received from the Science and Engineering Research Board (SERB; EMR/2016/000513) and the Department of Biotechnology Government of India (DBT; BT/PR20801/BPA/118/199/2016) to MK.

ACKNOWLEDGMENTS

The authors are grateful to Prof. Prakash C. Sharma for his help in facilitating the project sanctioned from the Department of Biotechnology (DBT) Government of India and to Dr. Ranjith Kumar C.T. for help in proteomics experiments. Financial support in the form of Indraprastha Research Fellowship to DS and NW and Research Associateship to RY from DBT is gratefully acknowledged. The authors also thank Kathakali Banerjee for help in transcriptome work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01185/full#supplementary-material

REFERENCES

Abdel-Ghany, S. E., Müller-Moulé, P., Niyogi, K. K., Pilon, M., and Shikanai, T. (2005). Two P-type ATPases are required for copper delivery in *Arabidopsis thaliana* chloroplasts. *Plant Cell* 17, 1233–1251. doi: 10.1105/tpc.104.030452

Arya, D., Kapoor, S., and Kapoor, M. (2016). *Physcomitrella patens DNA methyltransferase 2* is required for recovery from salt and osmotic stress. *FEBS J.* 283, 556–570. doi: 10.1111/febs.13611

Baginsky, S., and Gruissem, W. (2004). Chloroplast proteomics: potentials and challenges. *J. Exp. Bot.* 55, 1213–1220. doi: 10.1093/jxb/erh104

Bakshi, M., and Oelmüller, R. (2014). Wrky transcription factors jack of many trades in plants. *Plant Signaling Behav.* 9, e27700. doi: 10.4161/psb.27700

Batelli, G., Verslues, P. E., Agius, F., Qiu, Q., Fujii, H., Pan, S., et al. (2007). SOS2 promotes salt tolerance in part by interacting with the vacuolar H^+-ATPase and upregulating its transport activity. *Mol. Cell. Biol.* 27, 7781–7790. doi: 10.1128/MCB.00430-07

Becker, M., Müller, S., Nellen, W., Jurkowski, T. P., Jeltsch, A., and Ehrenhofer-Murray, A. E. (2012). Pmt1, a Dnmt2 homolog in *Schizosaccharomyces pombe*, mediates tRNA methylation in response to nutrient signaling. *Nucleic Acids Res.* 40, 11648–11658. doi: 10.1093/nar/gks956

Burgess, A. L., David, R., and Searle, I. R. (2015). Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. *BMC Plant Biol.* 15, 1–17. doi: 10.1186/s12870-015-0580-8

Bürstenbinder, K., Waduwara, I., Schoor, S., Moffatt, B. A., Wirtz, M., Minocha, S. C., et al. (2010). Inhibition of 5'-methylthioadenosine metabolism in the Yang cycle alters polyamine levels, and impairs seedling growth and reproduction in Arabidopsis. *Plant J.* 62, 977–988. doi: 10.1111/j.1365-313X.2010.04211.x

Caro, E., Stroud, H., Greenberg, M. V. C., Bernatavichute, Y., Feng, S., Groth, M., et al. (2012). The SET-Domain Protein SUVR5 Mediates H3K9me2 Deposition and Silencing at Stimulus Response Genes in a DNA Methylation-Independent Manner. *PloS Genet.* 8, e1002995. doi: 10.1371/journal.pgen.1002995

Chen, P., Jäger, G., and Zheng, B. (2010). Transfer RNA modifications and genes for modifying enzymes in *Arabidopsis thaliana*. *BMC Plant Biol.* 10:201. doi: 10.1186/1471-2229-10-201

David, R., Burgess, A., Parker, B., Li, J., Pulsford, K., Sibbritt, T., et al. (2017). Transcriptome-wide mapping of RNA 5-methylcytosine in Arabidopsis mRNAs and noncoding RNAs. *Plant Cell* 29, 445–460. doi: 10.1105/tpc.16.00751

Davière, J. M., and Achard, P. (2016). A Pivotal Role of DELLAs in Regulating Multiple Hormone Signals. *Mol. Plant* 9, 10–20. doi: 10.1016/j.molp.2015.09.011

do Amaral, M. N., Arge, L. W. P., Benitez, L. C., Danielowski, R., da Silveira Silveira, S. F., da Rosa Farias, D., et al. (2016). Differential expression of photosynthesis-related genes and quantification of gas exchange in rice plants under abiotic stress. *Acta Physiol. Plantarum* 38, 153. doi: 10.1007/s11738-016-2176-9

Do, T. H., Pongthai, P., Ariyarathne, M., Teh, O. K., and Fujita, T. (2020). AP2/ERF transcription factors regulate salt-induced chloroplast division in the moss *Physcomitrella patens*. *J. Plant Res.* 133, 537–548. doi: 10.1007/s10265-020-01195-y

Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., et al (2014). MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.* 13, 3679–3684. doi: 10.1021/pr500202e

Durdevic, Z., Mobin, M. B., Hanna, K., Lyko, F., and Schaefer, M. (2013). The RNA methyltransferase dnmt2 is required for efficient dicer-2-dependent siRNA pathway activity in Drosophila. *Cell Rep.* 4, 931–937. doi: 10.1016/j.celrep.2013.07.046

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903

Fan, P., Feng, J., Jiang, P., Chen, X., Bao, H., Nie, L., et al. (2011). Coordination of carbon fixation and nitrogen metabolism in *Salicornia europaea* under salinity: Comparative proteomic analysis on chloroplast proteins. *Proteomics* 11, 4346–4367. doi: 10.1002/pmic.201100054

Fernandez-Pozo, N., Haas, F. B., Meyberg, R., Ullrich, K. K., Hiss, M., Perroud, P. F., et al. (2019). PEATmoss (*Physcomitrella* Expression Atlas Tool): a unified gene expression atlas for the model plant *Physcomitrella patens*. *Plant J.* 102, 165–177. doi: 10.1111/tpj.14607

Fisher, O., Siman-Tov, R., and Ankri, S. (2004). Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (Ehmeth) in the protozoan parasite *Entamoeba histolytica*. *Nucleic Acids Res.* 32, 287–297. doi: 10.1093/nar/gkh161

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C.-L., Zhang, X., et al. (2006). Methylation of $tRNA^{Asp}$ by the DNA Methyltransferase Homolog Dnmt2. *Science* 311, 395–398. doi: 10.1126/science.1120976

Gorman, D. S., and Levine, R. P. (1965). Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardi*. *Proc. Natl. Acad. Sci.* 54, 1665–1669. doi: 10.1073/pnas.54.6.1665

Hamaji, K., Nagira, M., Yoshida, K., Ohnishi, M., Oda, Y., Uemura, T., et al. (2009). Dynamic aspects of ion accumulation by vesicle traffic under salt stress in Arabidopsis. *Plant Cell Physiol.* 50, 2023–2033. doi: 10.1093/pcp/pcp143

Haque, M. S., Kjaer, K. H., Rosenqvist, E., and Ottosen, C. O. (2015). Continuous light increases growth, daily carbon gain, antioxidants, and alters carbohydrate metabolism in a cultivated and a wild tomato species. *Front. Plant Sci.* 6:522:522. doi: 10.3389/fpls.2015.00522

Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* 34, 721–733. doi: 10.1093/nar/gkj471

Hermann, A., Schmitt, S., and Jeltsch, A. (2003). The human Dnmt2 has residual DNA-(Cytosine-C5) methyltransferase activity. *J. Biol. Chem.* 278, 31717–31721. doi: 10.1074/jbc.M305448200

Higashi, Y., Takechi, K., Takano, H., and Takio, S. (2013). Involvement of MicroRNA in copper deficiency-induced repression of chloroplastic CuZn-Superoxide Dismutase genes in the moss *Physcomitrella patens*. *Plant Cell Physiol.* 54, 1345–1355. doi: 10.1093/pcp/pct084

Hung, M. S., Karthikeyan, N., Huang, B., Koo, H. C., Kiger, J., and Shen, C. K. J. (1999). Drosophila proteins related to vertebrate DNA (5-cytosine) methyltransferases. *Proc. Natl. Acad. Sci.* 96, 11940–11945. doi: 10.1073/pnas.96.21.11940

Ichikawa, K., Sugita, M., Imaizumi, T., Wada, M., and Aoki, S. (2004). Differential expression on a daily basis of plastid sigma factor genes from the moss *Physcomitrella patens*. Regulatory interactions among *PpSig5*, the circadian

clock, and blue light signaling mediated by cryptochromes. *Plant Physiol.* 136, 4285–4298. doi: 10.1104/pp.104.053033

Ji, H., Pardo, J. M., Batelli, G., van Oosten, M. J., Bressan, R. A., and Li, X. (2013). The salt overly sensitive (SOS) pathway: Established and emerging roles. *Mol. Plant* 6, 275–286. doi: 10.1093/mp/sst017

Katoh, M., Curk, T., Xu, Q., Zupan, B., Kuspa, A., and Shaulsky, G. (2006). Developmentally regulated DNA methylation in *Dictyostelium discoideum*. *Eukaryotic Cell* 5, 18–25. doi: 10.1128/EC.5.1.18-25.2006

Kuhlmann, M., Borisova, B. E., Kaller, M., Larsson, P., Stach, D., Na, J., et al. (2005). Silencing of retrotransposons in Dictyostelium by DNA methylation and RNAi. *Nucleic Acids Res.* 33, 6405–6417. doi: 10.1093/nar/gki952

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development* 130, 5083–5090. doi: 10.1242/dev.00716

Li, Q., Wang, J., Ye, J., Zheng, X., Xiang, X., Li, C., et al. (2017). The maize imprinted gene floury3 encodes a PLATZ protein required for tRNA and 5s rRNA transcription through interaction with RNA polymerase III. *Plant Cell* 29, 2661–2675. doi: 10.1105/tpc.17.00576

Lin, M. J., Tang, L. Y., Reddy, M. N., and Shen, C. K. J. (2005). DNA methyltransferase gene dDnmt2 and longevity of Drosophila. *J. Biol. Chem.* 280, 861–864. doi: 10.1074/jbc.C400477200

Liu, K., Wang, Y. F., Cantemir, C., and Muller, M. T. (2003). Endogenous Assays of DNA Methyltransferases: Evidence for Differential Activities of DNMT1, DNMT2, and DNMT3 in Mammalian Cells In Vivo. *Mol. Cell. Biol.* 23, 2709–2719. doi: 10.1128/mcb.23.8.2709-2719.2003

Liu, P. L., Du, L., Huang, Y., Gao, S. M., and Yu, M. (2017). Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evol. Biol.* 17, 1–16. doi: 10.1186/s12862-017-0891-5

López-Pérez, L., Martínez-Ballesta, M., del, C., Maurel, C., and Carvajal, M. (2009). Changes in plasma membrane lipids, aquaporins and proton pump of broccoli roots, as an adaptation mechanism to salinity. *Phytochemistry* 70, 492–500. doi: 10.1016/j.phytochem.2009.01.014

Ma, D., and Constabel, C. P. (2019). MYB Repressors as Regulators of Phenylpropanoid Metabolism in Plants. *Trends Plant Sci.* 24, 275–289. doi: 10.1016/j.tplants.2018.12.003

Malik, G., Dangwal, M., Kapoor, S., and Kapoor, M. (2012). Role of DNA methylation in growth and differentiation in *Physcomitrella patens* and characterization of cytosine DNA methyltransferases. *FEBS J.* 279, 4081–4094. doi: 10.1111/febs.12002

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Miller, G., Suzuki, N., Ciftci-Yilmaz, S., and Mittler, R. (2010). Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant Cell Environ.* 33, 453–467. doi: 10.1111/j.1365-3040.2009.02041.x

Mittler, R., Vanderauwera, S., Gollery, M., and van Breusegem, F. (2004). Reactive oxygen gene network of plants. *Trends Plant Sci.* 9, 490–498. doi: 10.1016/j.tplants.2004.08.009

Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1819, 86–96. doi: 10.1016/j.bbagrm.2011.08.004

Motorin, Y., Lyko, F., and Helm, M. (2009). 5-methylcytosine in RNA: Detection, enzymatic formation and biological functions. *Nucleic Acids Res.* 38, 1415–1430. doi: 10.1093/nar/gkp1117

Motorin, Y., Seidu-Larry, S., and Helm, M. (2016). DNA and RNA pyrimidine nucleobase alkylation at the carbon-5 position. *Adv. Exp. Med. Biol.* 945, 19–33. doi: 10.1007/978-3-319-43624-1_2

Moura, J. C. M. S., Bonine, C. A. V., de Oliveira Fernandes Viana, J., Dornelas, M. C., and Mazzafera, P. (2010). Abiotic and biotic stresses and changes in the lignin content and composition in plants. *J. Integr. Plant Biol.* 52, 360–376. doi: 10.1111/j.1744-7909.2010.00892.x

Müller, S., Windhof, I. M., Maximov, V., Jurkowski, T., Jeltsch, A., Förstner, K. U., et al. (2013). Target recognition, RNA methylation activity and transcriptional regulation of the *Dictyostelium discoideum* Dnmt2-homologue (DnmA). *Nucleic Acids Res.* 41, 8615–8627. doi: 10.1093/nar/gkt634

Mund, C., Musch, T., Str, M., Assmann, B., and Lyko, F. (2004). Comparative analysis of DNA methylation patterns in transgenic *Drosophila* overexpressing mouse DNA methyltransferases. *Biochem. J.* 378, 763–768. doi: 10.1042/bj20031567

Nagae, M., Nakata, M., and Takahashi, Y. (2008). Identification of negative cis-acting elements in response to copper in the chloroplastic iron superoxide dismutase gene of the moss *Barbula unguiculata*. *Plant Physiol.* 146, 1687–1696. doi: 10.1104/pp.107.114868

Nakajima, K., Kawamura, T., and Hashimoto, T. (2006). Role of the *SPIRAL1* gene family in anisotropic growth of *Arabidopsis thaliana*. *Plant Cell Physiol.* 47, 513–522. doi: 10.1093/pcp/pcj020

Okano, M., Xie, S., and Li, E. (1998). Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* 26, 2536–2540. doi: 10.1093/nar/26.11.2536

Pavlopoulou, A., and Kossida, S. (2009). Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. *Genomics* 93, 350–357. doi: 10.1016/j.ygeno.2008.12.004

Pinarbasi, E., Elliott, J., and Hornby, D. P. (1996). Activation of a Yeast Pseudo DNA Methyltransferase by Deletion of a Single Amino Acid. *J. Mol. Biol.* 257, 804–813. doi: 10.1006/jmbi.1996.0203

Qiu, Q. S., Guo, Y., Quintero, F. J., Pardo, J. M., Schumaker, K. S., and Zhu, J. K. (2004). Regulation of Vacuolar Na^+/H^+ Exchange in *Arabidopsis thaliana* by the Salt-Overly-Sensitive (SOS) Pathway. *J. Biol. Chem.* 279, 207–215. doi: 10.1074/jbc.M307982200

Quan, R., Wang, J., Yang, D., Zhang, H., Zhang, Z., and Huang, R. (2017). EIN3 and SOS2 synergistically modulate plant salt tolerance. *Scientific reports* 7, 44637. doi: 10.1038/srep44637

Rai, K., Chidester, S., Zavala, C. V., Manos, E. J., James, S. R., Karpf, A. R., et al. (2007). Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes Dev.* 21, 261–266. doi: 10.1101/gad.1472907

Reboledo, G., del Campo, R., Alvarez, A., Montesano, M., Mara, H., and de León, I. P. (2015). *Physcomitrella patens* activates defense responses against the pathogen *Colletotrichum gloeosporioides*. *Int. J. Mol. Sci.* 16, 22280–22298. doi: 10.3390/ijms160922280

Rippert, P., and Matringe, M. (2002a). Molecular and biochemical characterization of an *Arabidopsis thaliana* arogenate dehydrogenase with two highly similar and active protein domains. *Plant Mol. Biol.* 48, 361–368. doi: 10.1023/A:1014018926676

Rippert, P., and Matringe, M. (2002b). Purification and kinetic analysis of the two recombinant arogenate dehydrogenase isoforms of *Arabidopsis thaliana*. *Eur. J. Biochem.* 269, 4753–4761. doi: 10.1046/j.1432-1033.2002.03172.x

Ronzier, E., Corratgé-Faillie, C., Sanchez, F., Prado, K., Brière, C., Leonhardt, N., et al. (2014). CPK13, A noncanonical Ca^{2+}-Dependent protein kinase, Specifically inhibits KAT2 and KAT1 shaker K^+ channels and reduces stomatal opening. *Plant Physiol.* 166, 314–326. doi: 10.1104/pp.114.240226

Ruibal, C., Salamó, I. P., Carballo, V., Castro, A., Bentancor, M., Borsani, O., et al. (2012). Differential contribution of individual dehydrin genes from *Physcomitrella patens* to salt and osmotic stress tolerance. *Plant Sci.* 190, 89–102. doi: 10.1016/j.plantsci.2012.03.009

Saavedra, L., Svensson, J., Carballo, V., Izmendi, D., Welin, B., and Vidal, S. (2006). A dehydrin gene in *Physcomitrella patens* is required for salt and osmotic stress tolerance. *Plant J.* 45, 237–249. doi: 10.1111/j.1365-313X.2005.02603.x

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., et al. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* 24, 1590–1595. doi: 10.1101/gad.586710

Shoji, T., Narita, N. N., Hayashi, K., Asada, J., Hamada, T., Sonobe, S., et al. (2004). Plant-specific microtubule-associated protein SPIRAL2 is required for anisotropic growth in Arabidopsis. *Plant Physiol.* 136, 3933–3944. doi: 10.1104/pp.104.051748

Song, Y., Wu, K., Dhaubhadel, S., An, L., and Tian, L. (2010). *Arabidopsis* DNA methyltransferase AtDNMT2 associates with histone deacetylase AtHD2s activity. *Biochem. Biophys. Res. Commun.* 396, 187–192. doi: 10.1016/j.bbrc.2010.03.119

Tang, L. Y., Reddy, M. N., Rasheva, V., Lee, T. L., Lin, M. J., Hung, M. S., et al. (2003). The Eukaryotic DNMT2 Genes Encode a New Class of Cytosine-5 DNA Methyltransferases. *J. Biol. Chem.* 278, 33613–33616. doi: 10.1074/jbc.C300255200

Thiagarajan, D., Dev, R. R., and Khosla, S. (2011). The DNA methyltranferase Dnmt2 participates in RNA processing during cellular stress. *Epigenetics* 6, 103–113. doi: 10.4161/epi.6.1.13418

Tovy, A., Tov, R. S., Gaentzsch, R., Helm, M., and Ankri, S. (2010). A new nuclear function of the *Entamoeba histolytica* glycolytic enzyme enolase: The metabolic regulation of cytosine-5 methyltransferase 2 (Dnmt2) activity. *PloS Pathog.* 6, e1000775. doi: 10.1371/journal.ppat.1000775

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., et al. (2012). RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* 19, 900–905. doi: 10.1038/nsmb.2357

Van der Does, D., Boutrot, F., Engelsdorf, T., Rhodes, J., McKenna, J. F., Vernhettes, S., et al. (2017). The Arabidopsis leucine-rich repeat receptor kinase MIK2/LRR-KISS connects cell wall integrity sensing, root growth and response to abiotic and biotic stresses. *PloS Genet.* 13. doi: 10.1371/journal.pgen.1006832

Verslues, P. E., Batelli, G., Grillo, S., Agius, F., Kim, Y. S., Zhu, J., et al. (2007). Interaction of SOS2 with nucleoside diphosphate kinase 2 and catalases reveals a point of connection between salt stress and H2O2 signaling in Arabidopsis thaliana. *Molecular and Cellular Biology* 27, 7771–7780. doi: 10.1128/mcb.00429-07

Vieira, G. C., D'Ávila, M. F., Zanini, R., Deprá, M., and da Valente, V. L. S. (2018). Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions. *Genet. Mol. Biol.* 41, 215–234. doi: 10.1590/1678-4685-gmb-2017-0056

Wang, C., Li, J., and Yuan, M. (2007). Salt tolerance requires cortical microtubule reorganization in Arabidopsis. *Plant Cell Physiol.* 48, 1534–1547. doi: 10.1093/pcp/pcm123

Wang, X., Yang, P., Gao, Q., Liu, X., Kuang, T., Shen, S., et al (2008). Proteomic analysis of the response to high-salinity stress in *Physcomitrella patens*. *Planta* 228, 167–177. doi: 10.1007/s00425-008-0727-z

Wang, R., Liu, S., Zhou, F., Ding, C., and Hua, C. (2014). Exogenous ascorbic acid and glutathione alleviate oxidative stress induced by salt stress in the chloroplasts of *Oryza Sativa* L. *Z. fur Naturforschung - Section C J. Biosci.* 69 (C), 226–236. doi: 10.5560/ZNC.2013-0117

Wang, A., Hou, Q., Si, L., Huang, X., Luo, J., Lu, D., et al. (2019). The PLATZ transcription factor GL6 affects grain length and number in rice. *Plant Physiol.* 180, 2077–2090. doi: 10.1104/pp.18.01574

Wang, H., and Schippers, J. H. M. (2019). The role and regulation of autophagy and the proteasome during aging and senescence in plants. *Genes* 10, 267. doi: 10.3390/genes10040267

Wilkinson, C. R. M., Bartlett, R., Nurse, P., and Bird, A. P. (1995). The fission yeast gene *pmt1*$^{+}$ encodes a DNA methyltransferase homologue. *Nucleic Acids Res.* 23, 203–210. doi: 10.1093/nar/23.2.203

Yadlapalli, S., and Yamashita, Y. M. (2013). Chromosome-specific nonrandom sister chromatid segregation during stem-cell division. *Nature* 498, 251–254. doi: 10.1038/nature12106

Yamasaki, H., Pilon, M., and Shikanai, T. (2008). How do plants respond to copper deficiency? *Plant Signaling Behav.* 3, 231–232. doi: 10.4161/psb.3.4.5094

Yang, Y., and Guo, Y. (2018). Unraveling salt stress signaling in plants. *J. Integr. Plant Biol.* 60, 796–804. doi: 10.1111/jipb.12689

Yates, J. R., Eng, J. K., and McCormack, A. L. (1995). Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Analytical Chem.* 67, 3202–3210. doi: 10.1021/ac00114a016

Yoder, J. A., and Bestor, T. H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum. Mol. Genet.* 7, 279–284. doi: 10.1093/hmg/7.2.279

Yu, H., Zhang, Y., Zhang, D., Lu, Y., He, H., Zheng, F., et al. (2017). Identification of a ribose-phosphate pyrophosphokinase that can interact in vivo with the anaphase promoting complex/cyclosome. *Int. J. Mol. Sci.* 18, 617. doi: 10.3390/ijms18040617

Zhu, J. K. (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53, 247–273. doi: 10.1146/annurev.arplant.53.091401.143329

Zörb, C., Herbst, R., Forreiter, C., and Schubert, S. (2009). Short-term effects of salt exposure on the maize chloroplast protein pattern. *Proteomics* 9, 4209–4220. doi: 10.1002/pmic.200800791

Conflict of Interest: SKau was employed by the company Vproteomics, Valerian Chem Private Limited.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 07 August 2020
doi: 10.3389/fpls.2020.01178

It Is Hot in the Sun: Antarctic Mosses Have High Temperature Optima for Photosynthesis Despite Cold Climate

Alicia V. Perera-Castro[1,2], Melinda J. Waterman[2], Johanna D. Turnbull[2], Michael B. Ashcroft[2], Ella McKinley[3], Jennifer R. Watling[3,4], Jessica Bramley-Alves[2], Angelica Casanova-Katny[5], Gustavo Zuniga[6], Jaume Flexas[1] and Sharon A. Robinson[2,7]*

[1] Department of Biology, Universitat de les Illes Balears, INAGEA, Palma de Mallorca, Spain, [2] Centre for Sustainable Ecosystem Solutions, School of Earth, Atmosphere and Life Sciences, University of Wollongong, Wollongong, NSW, Australia, [3] School of Biological Sciences, The University of Adelaide, Adelaide, SA, Australia, [4] Manchester Metropolitan University, Manchester, United Kingdom, [5] Laboratorio de Ecofisiología Vegetal y Cambio Climático y Núcleo de Estudios Ambientales (NEA), Facultad de Recursos Naturales, Universidad Católica de Temuco, Temuco, Chile, [6] Facultad de Química y Biología, Universidad de Santiago de Chile, Santiago, Chile, [7] Global Challenges Program, University of Wollongong, Wollongong, NSW, Australia

OPEN ACCESS

Edited by:
Jeffrey Graham Duckett, Natural History Museum, United Kingdom

Reviewed by:
Peter Convey, British Antarctic Survey (BAS), United Kingdom
Jeffrey William Bates, Imperial College London, United Kingdom

***Correspondence:**
Alicia V. Perera-Castro
pereraalicia11@gmail.com

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *20 January 2020*
Accepted: *21 July 2020*
Published: *07 August 2020*

Citation:
Perera-Castro AV, Waterman MJ, Turnbull JD, Ashcroft MB, McKinley E, Watling JR, Bramley-Alves J, Casanova-Katny A, Zuniga G, Flexas J and Robinson SA (2020) It Is Hot in the Sun: Antarctic Mosses Have High Temperature Optima for Photosynthesis Despite Cold Climate. Front. Plant Sci. 11:1178. doi: 10.3389/fpls.2020.01178

The terrestrial flora of Antarctica's frozen continent is restricted to sparse ice-free areas and dominated by lichens and bryophytes. These plants frequently battle sub-zero temperatures, extreme winds and reduced water availability; all influencing their ability to survive and grow. Antarctic mosses, however, can have canopy temperatures well above air temperature. At midday, canopy temperatures can exceed 15°C, depending on moss turf water content. In this study, the optimum temperature of photosynthesis was determined for six Antarctic moss species: *Bryum pseudotriquetrum*, *Ceratodon purpureus*, *Chorisodontium aciphyllum*, *Polytrichastrum alpinum*, *Sanionia uncinata*, and *Schistidium antarctici* collected from King George Island (maritime Antarctica) and/or the Windmill Islands, East Antarctica. Both chlorophyll fluorescence and gas exchange showed maximum values of electron transport rate occurred at canopy temperatures higher than 20°C. The optimum temperature for both net assimilation of CO_2 and photoprotective heat dissipation of three East Antarctic species was 20–30°C and at temperatures below 10°C, mesophyll conductance did not significantly differ from 0. Maximum mitochondrial respiration rates occurred at temperatures higher than 35°C and were lower by around 80% at 5°C. Despite the extreme cold conditions that Antarctic mosses face over winter, the photosynthetic apparatus appears optimised to warm temperatures. Our estimation of the total carbon balance suggests that survival in this cold environment may rely on a capacity to maximize photosynthesis for brief periods during summer and minimize respiratory carbon losses in cold conditions.

Keywords: Antarctica, bryophytes, carbon balance, electron transport rate, mesophyll conductance, net CO_2 assimilation, non-photochemical quenching, respiration

INTRODUCTION

Antarctica is considered the coldest continent on Earth, since the surface air temperature can reach annual means of -23°C (-45°C in interior regions higher than 1500 m a. s. l.) (Fortuin and Oerlemans, 1990). However, outside the Antarctic circle the meteorological conditions have been reported to be relatively milder. For instance, in the South Shetlands Islands of Maritime Antarctica the daytime mean air temperatures vary between -5°C and 13°C in the summer and only reach -30°C in winter (Convey and Smith, 2005; Pearce, 2008). In these southern latitudes, terrestrial vegetation – mainly lichens and bryophytes – is restricted to ice-free areas (Peat et al., 2007; Ochyra et al., 2008). Soil surface temperatures have been recorded to be much warmer than the ~2 m air temperatures reported by meteorological stations, with maximum differences of 10.7°C (Schenker and Block, 1986), 25°C (Smith, 1996) or even 27°C (Matsuda, 1968) in summer. In winter, Antarctic mosses will normally be in a dormant state protected by a thick, insulating layer of snow.

Likewise, the microclimate of mosses has been described to be radically different from air temperature recorded in Antarctica (Longton, 1974; Walton, 1982; Edwards and Smith, 1988; Smith, 1988; Hovenden et al., 1994; Melick and Seppelt, 1994; Davey and Rothery, 1997; Green et al., 2005; Pannewitz et al., 2005; Block et al., 2009; Bramley-Alves et al., 2014; Zúñiga, 2016; Convey et al., 2018). Block et al. (2009) reported daily cycles of temperature ranging from 0°C to 44.4°C during the day and then to -2.2°C at night in the moss *Andreaea regularis* on a rock surface at Signy Island (60°S). Such a variation between the plant surface and air temperatures has been attributed to radiation, the angle of its incidence and wind speed in polar (Wilson, 1957) and alpine (Körner, 2003) environments. The fact that daily temperature variation is so high raises questions about the actual period during which Antarctic mosses have optimum conditions for physiological processes, such us carbon fixation.

Several researchers have reported a high optimum temperature in Antarctic mosses for both CO_2 uptake (Longton, 1988b; Kappen et al., 1989; Davey and Rothery, 1997; Green et al., 2000; Pannewitz et al., 2005; Block et al., 2009) and O_2 evolution (Rastorfer, 1970; Ino, 1990; Wilson, 1990; Smith, 1999; Newsham, 2010), suggesting that the studied species were not truly psychrophilic. Longton et al. (1988a; 1988b) observed high optimum temperatures for Antarctic mosses but with a broad curve resulting in a positive intercept at 0° C. In contrast, some authors have considered bryophytes to generally have lower temperature optima for photosynthesis at about 5–15°C (He et al., 2016 and references therein). Although, Ino (1990) pointed out that only mosses inhabiting locations that were frequently submersed in cold water had low optimum temperatures for photosynthesis. So far, very few researchers have addressed the effect of the high specific heat capacity of water on the maximum temperatures experienced by mosses during a daily cycle (Block et al., 2009). Water availability has been suggested to be more relevant than temperature to biology (Kennedy, 1993) and to be the main factor ruling species distribution in Antarctica (Davey and Rothery, 1997; Convey et al., 2014). However, the interactive effect that water and temperature have in providing a favourable environment for positive carbon balance in Antarctica remains unknown.

The effect of temperature on photosynthesis has been thoroughly studied (Sage and Kubien, 2007; Flexas et al., 2014; von Caemmerer and Evans, 2015). Several biochemical and biophysical processes involved in photosynthesis are affected by temperature: (1) thylakoid membrane fluidity (Hirano et al., 1981), (2) kinetics of electron transport and the Calvin-Benson cycle (Holaday et al., 1992; Sage, 2002; Walker et al., 2013), (3) stomatal aperture, and (4) CO_2 diffusivity in membranes and aqueous/wall phase of mesophyll (as a sum, termed mesophyll conductance, g_m). g_m is one of the more relevant limiting factors of photosynthesis, especially in bryophytes, which present the lowest values of g_m of the plant kingdom (Flexas et al., 2012; Carriquí et al., 2019; Gago et al., 2019). The short-term response of g_m to temperature is considered species-specific (Bunce, 2008; von Caemmerer and Evans, 2015), although most of the studied species have shown a decrease of g_m at low temperatures (Bernacchi et al., 2002; Warren and Dreyer, 2006; Scafaro et al., 2011; Ubierna et al., 2017; but see also Qiu et al., 2017). The response of g_m to temperature, however, is unknown for bryophytes.

As a consequence of reduced photosynthetic rates at low temperatures is that the capacity to use light energy in carbon assimilation will decrease and saturation will occur at lower irradiances (Huner et al., 1993; Ensminger et al., 2006). Many photoprotection mechanisms have been described in both vascular plants and bryophytes to avoid photodamage of photosynthetic apparatus in this situation (Robinson and Waterman, 2014). One of the more short-term dynamic mechanisms of photoprotection is the regulated heat dissipation of excess energy (estimated by a chlorophyll fluorescence parameter termed non-photochemical quenching, NPQ), which is dependent on the build-up of a gradient in pH across the thylakoid membrane and has been reported to broadly increase during acclimation to low temperatures (Hendrickson et al., 2004; Míguez et al., 2015; Yang et al., 2018). To our knowledge, despite its relevance to understanding the limitations of photosynthesis at low, variable temperatures in Antarctic mosses, neither the temperature response of photoprotective heat dissipation nor g_m has so far been reported for bryophytes (with the exception of unsteady-state NPQ for one Mediterranean moss species in Deltoro et al., 1999).

Thus, the objectives of this study were (1) to model the daily carbon balance of Antarctic mosses during summer based on canopy surface temperature and its effect on photosynthesis, (2) to test the interspecific differences and the possible buffering effect of water content on moss canopy temperature, and (3) to determine the temperature responses of net CO_2 assimilation, electron transport rate, g_m and photoprotection mechanisms assessed by NPQ.

MATERIALS AND METHODS

Study Site and Plant Material

Two Antarctic locations were included in this study: Casey station (66°16′57″S, 110°31′36″E) on Bailey Peninsula (Windmill Islands

region, East Antarctica) and Fildes Peninsula (62°12′05″S, 58°57′44″W) on King George Island (South Shetland Islands) (**Figure 1**). According to the Australian Bureau of Meteorology, mean maximum and minimum air temperature at Casey Station during the hottest month of the Antarctic summer (January) is 2.3 and -2.5° C, respectively, whereas in winter (July) mean maximum temperatures can drop to -10.8°C (data from 1989 to 2019). Fildes Peninsula in Maritime Antarctica present similar air temperatures in summer with mean max/min temperatures of 2.8/0.1°C (data from 1969–2012, for Bellingshausen Station, consistent with data reported for Frei Montalva Station by Carrasco and González, 2007).

Six species of bryophytes were studied during different Antarctic campaigns (**Table 1**). *Bryum pseudotriquetrum*, *Ceratodon purpureus* and *Schistidium antarctici* were found on Bailey Peninsula near Casey Station (see Robinson et al., 2018 and King et al., 2020 for detailed maps and site descriptions). *B. pseudotriquetrum*, *S. antarctici*, *Chorisodontium aciphyllum*, *Polytrichastrum alpinum* and *Sanionia uncinata* were located at Fildes Peninsula and Ardley Island near Escudero Station. Mosses were identified to species by ACK and MJW (King George Island) and SR, JBA, and JDT (Windmill Islands). Specimen vouchers of each species were deposited in either the Janet Cosh Herbarium (University of Wollongong, Australia) or the CONC Herbarium (Universidad de Concepción, Chile).

Microclimatic Conditions and Moisture Effect

Daily Moss Surface Temperature Near Casey Station, East Antarctica

The surface temperature of East Antarctic *B. pseudotriquetrum*, *C. purpureus*, and *S. antarctici* was recorded during the Antarctic summer of 2003 (from 16/01/2003 to 28/01/2003) at six locations around Casey Station (ASPA 135). At each location a polyurethane sealed iBCod temperature sensor (Thermodata Pty. Ltd., Brisbane, Australia) was placed on the moss surface for 13 days of continuous recording. The frequency at which mosses experienced a determined interval of temperature (intervals of 2°C from -4 to 28°C) was calculated as: Time (%) = 100 · t/T, where t is the number of records of each interval of temperature and T is the total of records for each species.

For the same species and locations, additional spot measurements of photosynthetic photon flux density (PPFD), wind gust speed, moss surface temperature, and air temperature were collected at midday (local time, UTC +13) for a wider period (22 days distributed from 9/11/2002 to 01/02/2003). Moss surface temperature was measured with an infrared thermometer (Scotchtrack T Heat tracer IR1600L; 3M, Austin TX, USA). A high correlation (r = 0.914) was found between iBCod and infrared thermometer recordings for four random days when samples were measured with both sensors at midday (**Supplemental Figure 1**). PPFD was measured with a LS-C mini quantum sensor attached to the leaf clip holder of a Walz MINI-PAM Photosynthesis Yield Analyser (WALZ, Effeltrich, Germany) placed at moss surface level during measurement. Wind gust speed and air temperature were obtained from the Australian Bureau of Meteorology at Casey station.

Daily Moss Surface Temperature on Fildes Peninsula and Ardley Island, Maritime Antarctica

Daily air and moss surface temperature of maritime Antarctic *B. pseudotriquetrum*, *C. aciphyllum*, *P. alpinum*, and *S. uncinata* were recorded during the Antarctic summer of 2019 (from 08/01/2019 to 30/01/2019) at seven locations around Ardley Island (ASPA150). In this case, one HOBO 4-channel thermocouple datalogger (UX120-014M, Onset Computer Corporation, Bourne, MA, USA) was placed at each location, so that the moss surface temperature of three specimens and air temperature could be recorded simultaneously. The frequency with which mosses experienced a particular temperature interval

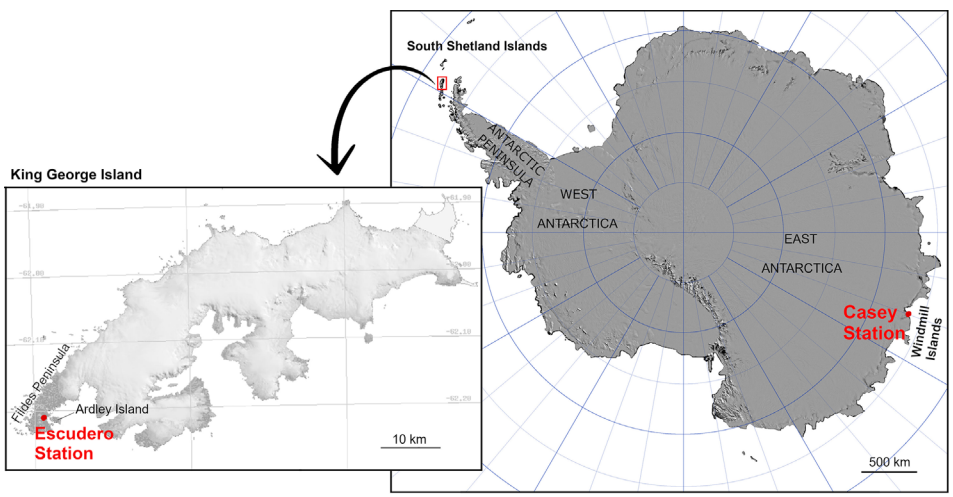


FIGURE 1 | Location of the South Shetland Islands and Windmill Islands (Casey Station), Antarctica with inset of King George Island showing location of Escudero Station (original source: the Scientific Committee on Antarctic Research).

TABLE 1 | Description of studied species with their habitat in Antarctica and phytogeography according to Ochyra et al., 2008.

Species	Habitat in Antarctica	World distribution	Loc	Moss surface temperature recording	ETR measurements	Gas exchange and NPQ measurements[§]
Bryum pseudotriquetrum (Hedw.) P.Gaert., B.Mey. & Scherb.	Ubiquitous.	Bipolar, transitional at higher tropical and subtropical elevations	CS	2003 (ASPA 135)	2011/12 (Casey Red Shed and Science sites)	2014 (Casey Red Shed site)
			KGI	2019 (Ardley Island, ASPA150)	2015 (Fildes Peninsula)	
Ceratodon purpureus (Hedw.) Brid.	Mostly in dry and exposed sites	Cosmopolitan**	CS	2003 (ASPA 135)	2011/12 (Casey Red Shed and Science sites)	2014 (Casey Red Shed site)
Schistidium antarctici (Cardot) L.I.Savicz	Wide range of habitats and expositions*	Endemic, pan-continental Antarctic	CS	2003 (ASPA 135)	2011/12 (Casey Red Shed and Science sites)	2014 (Casey Red Shed site)
			KGI		2015 (Fildes Peninsula)	
Sanionia uncinata (Hedw.) Loeske	Wide range of substrates, usually in drier habitats	Bipolar, transitional at higher tropical and subtropical elevations	KGI	2019 (Ardley Island, ASPA150)	2015 (Ardley Island, ASPA150)	
Chorisodontium aciphyllum (Hook.f. & Wilson)	In well drained substrates in the form of moss turf	South-temperate, widespread throughout the Southern Hemisphere	KGI	2019 (Ardley Island, ASPA150)	2015 (Ardley Island, ASPA150)	
Polytrichastrum alpinum (Hedw.) G.L.Sm.	In well drained gravelly/stony ground	Bipolar, transitional at higher tropical and subtropical elevations	KGI	2019 (Ardley Island, ASPA150)	2015 (Ardley Island, ASPA150)	

Year when samples were collected and measurements made is indicated below each kind of measurement (with sampling sites in brackets). CS, Casey Station; KGI, King George Island.

*See also Robinson et al. (2000), [§]Lab measurements performed in 2018.

**See also Biermsa et al. (2020).

was calculated as in *Daily Moss Surface Temperature Near Casey Station, East Antarctica*.

Environmental Moisture Effect on Moss Surface Temperature

In order to test the effect of water on the seasonal shift in moss surface temperature, 90 iBcod and iButton sensors (Maxim Integrated, San Jose, USA) were deployed across water gradients on top of cushions of *S. antarctici* at the Red Shed site at Casey Station. Three environments were described: wet, dry and intermediate. Wet environments were located adjacent to a meltwater stream (**Figure 2**); dry environments were located 2 m from the water edge, with intermediate environment located in the middle. Moss turf water content (TWC) of each environment was estimated after Lucieer et al. (2014) and King (2017) by submerging sponges within the moss turf for 24 h. Significantly different values of TWC were obtained for each environment (**Supplemental Figure 2**), partially published at Bramley-Alves et al. (2015). Moss surface temperatures were logged continuously from 29/11/2011 to 26/01/2012 and from 13/01/2013 to 31/01/2013. Frequencies of time at which mosses experienced a particular temperature interval was also calculated as in *Daily Moss Surface Temperature Near Casey Station, East Antarctica*.

Temperature Responses of Chlorophyll Fluorescence and Gas Exchange

Electron Transport Rate at Casey and Escudero Station

In order to determine the optimum temperatures for electron transport rate (ETR) of the moss species, measurements of chlorophyll fluorescence were performed under lab conditions with a Walz MINI-PAM Photosynthesis Yield Analyser fitted to a Walz external halogen lamp (FL 400). Fresh samples of moss tissue were collected from sunny microhabitats on Fildes Peninsula, King George Island (January 2015) and Casey Station, Antarctica (December 2011) and measured within 2 days. Moss samples were maintained under natural sunlight and temperature levels prior to measurement. Moss cushions were divided into moss plugs of 1–2 cm^2 diameter and subjected randomly to temperatures from 5 to 40°C in groups of 5 replicates per temperature curve (from 4 to 28°C for Casey measurements) in the laboratory. Replicate plugs were maintained in aluminum cups in a water bath set to the target temperature and moss surface temperature were monitored. The aluminum cups allowed heat transfer within the water bath but prevented submergence and ensured the photosynthetic surface of the moss remained exposed to air. At each temperature, the specimens were pre-illuminated at PPFD of ~100 μmol photons m^{-2} s^{-1} while moss temperature equilibrated with the water temperature. They were then maintained at temperature for 5 min before a rapid light response curve was performed. Thermocouples were used to measure temperature of the photosynthetic tissue before and after each light response curve. ETR was calculated according to Krall and Edwards (1992): $ETR = \phi_{PSII} \cdot PPFD \cdot \alpha\beta$, where ϕ_{PSII} is the yield of

FIGURE 2 | A typical wet site abutting a melt stream at the Red Shed site at Casey (left). Close up (right) shows iBCod sensor setup on a moss cushion.

PSII and αβ is the product of absorbance and the partitioning of absorbed quanta between PSI and PSII. ϕ_{PSII} was calculated according to Genty et al. (1989): $\phi_{PSII} = (F_m'-F_s)/F_m'$, where F_s and F_m' are the steady-state and maximal fluorescence at light adapted conditions, respectively.

Since αβ was unknown for this set of measurements, a provisional value of 0.42 was used (Maxwell and Johnson, 2000) and was assumed to remain constant with temperature. The maximum light-saturated ETR (ETR_{max}) was obtained by fitting each light curve to a rational model (Smith, 1936) or to the waiting-in-line model (Ritchie, 2008) by using the Microsoft Excel Solver tool (adapted to ETR light curves from Lobo et al., 2013). The lowest square sum errors of a non-photoinhibited light curve were obtained with rational model (eqn 1), while the waiting-in-line model (eqn 2) was used for photoinhibited curves.

$$ETR = \frac{AQE \cdot ETRmax \cdot PPFD}{\sqrt{AQE^2 + (ETRmax \cdot PPFD)^2}} \quad \text{(eqn 1)}$$

$$ETR = \frac{ETRmax \cdot AQE \cdot PPFD \cdot e^{1-AQE \cdot PPFD/(ETRmax \cdot e)}}{ETRmax \cdot e} \quad \text{(eqn 2)}$$

where AQE is the Apparent Quantum Efficiency, also fitted by the model.

Mesophyll Conductance and NPQ Measurement

In order to determine optimum temperature for CO_2 assimilation (A_N) at saturating light (A_{sat}) and g_m, 6-7 samples of approximately 3 cm^2 of three study species – *S. antarctici*, *C. purpureus* and *B. pseudotriquetrum* – were collected near the Red Shed at Casey Station in 2014. Each sample was air-dried and stored at -20°C until analysis. In 2018, samples were thawed and rehydrated with distilled water for 10-16 h in dark conditions at 4°C prior to measurement under laboratory conditions. Prior to further measurements moss health was assessed by chlorophyll fluorescence, with high maximum quantum yield of PSII (F_v/F_m) indicating full recovery.

All gas exchange measurements were performed with a LiCOR 6800 system (LiCOR Biosciences, Lincoln, NE, USA). Between 40 and 47 dark-adapted sub-samples of each species were introduced into a custom-made cuvette consisting of a gasket affixed to a piece of thin polyester stocking fabric (**Supplementary Figure 3**). The size of these gaskets was equal to the chamber size to ensure proper closure of the chamber and achieve a minimum CO_2 leakage (**Supplementary Figure 4**). CO_2 concentration was standardized at 400 μmol CO_2·mol^{-1} air, relative humidity at 60%–75% and the flow rate within the chamber was 700 μmol·s^{-1}. The temperature of the chamber was varied between 5-35°C in steps of 5°C (n = 6–7 for each temperature of *S. antarctici*, and *B. pseudotriquetrum*; n = 4–5 for *C. purpureus*). After 5 min inside the chamber in dark conditions, dark respiration (R_D) was measured and a saturating pulse was applied in order to measure basal and maximum chlorophyll fluorescence (F_0 and F_m, respectively) and to calculate $F_v/F_m = (F_m-F_0)/F_m$. Then, the sample was exposed to saturating red light, with a maximum emission at 625 nm (800 μmol·m^{-2}·s^{-1}, which was determined with partial light curves performed *a priori*). A_{sat} at minimum saturating light was recorded at steady-state conditions, when diffusion limitations due to interstitial water were null and biochemistry was fully light-adapted (**Supplementary Figure 5**). At this stage, a second saturating pulse was applied to determine ϕ_{PSII} (as explained above) and non-photochemical quenching (NPQ) at each temperature, the latter calculated according to Bilger and Björkman (1990): $NPQ = (F_m - F_m')/F_m'$. ETR was calculated as explained in *Daily Moss Surface Temperature on Fildes Peninsula and Ardley Island, Maritime Antarctica* without assuming a constant αβ. For the samples where A_{sat} was also measured, αβ was determined as 4/slope of the relationship between ϕ_{PSII} and ϕ_{CO2} ($(A_N$ + light respiration)/PPFD) obtained by varying PPFD under non-photorespiratory conditions in an atmosphere containing <1% O_2 (Valentini et al., 1995) (n = 3–5 light curves per temperature and species). Since the ratios ϕ_{PSII} and ϕ_{CO2} increased significantly at higher PPFD, the αβ of the highest PPFD (the same used for measuring A_{sat}) was chosen for

calculating ETR. Light respiration (R_L) was calculated from the initial light-limited portion of the low-O_2-light curves as the negative intercept of the relationship between A_N and (ϕ_{PSII} · PPFD)/4 according to Yin et al. (2011). In order to avoid desiccation, if required, the sample was fully rehydrated before the low-O_2-light curves were measured; by immersing in distilled water for 1–2 min with excess water removed gently with a paper tissue before placing the sample back in the chamber.

g_m was estimated according to Harley et al. (1992) with the modifications of Carriquí et al. (2019):

$$g_m = \frac{A_{sat}}{C_a - \frac{\Gamma^*(ETR+8(A_{sat}+R_L))}{ETR-4(A_{sat}+R_L)}} \quad \text{(eqn 3)}$$

where Γ^* is the chloroplastic hypothetical CO_2 compensation point in the absence of respiration and C_a is the atmospheric CO_2 concentration. Since stomata are absent in gametophytes of bryophytes, stomatal CO_2 concentration (C_i) is substituted by C_a in Harley original formula. Γ^* was calculated from the Rubisco specificity factor ($S_{C/O}$) as:

$$\Gamma* = 0.5 \quad O/S_{C/O} \quad \text{(eqn 4)}$$

$S_{C/O}$ was averaged from the bryophytes species reported by Font et al. (Font and Galmés, 2016). The temperature coefficient Q_{10} was calculated for intervals of linear g_m-temperature as follow (Van't Hoff, 1884):

$$Q_{10} = \left(\frac{g_{m2}}{g_{m1}}\right)^{10/(T2-T1)} \quad \text{(eqn 5)}$$

The relative mesophyll (l_m) and biochemical (l_b) limitations to photosynthesis were calculated according to Grassi and Magnani (2005) with the modifications of Carriquí et al. (2019):

$$l_m = \frac{\partial A / \partial C_c}{g_m + \partial A / \partial C_c} \quad \text{(eqn 6)}$$

$$l_b = \frac{g_m}{g_m + \partial A / \partial C_c} \quad \text{(eqn 7)}$$

As a proxy to $\delta A/\delta C_c$ the quotient A_N/C_c at 400 μmol CO_2 · mol^{-1} air was calculated.

Estimation of Carbon Gain

Estimations of the carbon gain of East Antarctic *B. pseudotriquetrum*, *C. purpureus*, and *S. antarctici* were made by combining the surface temperature recorded at Casey during the summer of 2003 (see *Daily Moss Surface Temperature Near Casey Station, East Antarctica*) and the temperature response curves of A_{sat} measured in lab conditions for the same species collected from the same location during the summer of 2014 (see *Mesophyll Conductance and NPQ Measurement*). Surface temperature values of *S. antarctici* at different moist environments at Casey during the summer of 2011/12 and 2013 were also analyzed. Each interval of temperature experienced by mosses was assigned a corresponding mean value of A_{sat} (calculated from the polynomial curve fitting of the corresponding temperature response of A_{sat}) and its contribution to the total net CO_2 assimilation over the study period ($A_{sat,T}$) was estimated as: $A_{sat,T} = A_{sat} \cdot f$, where f is the frequency of time at which this interval of temperature was recorded. The balance between net carbon fixation (sum of positive $A_{sat,T}$) and carbon lost (sum of negative $A_{sat,T}$) during the studied period was calculated per species as:

$$\%\ C\ fixation = \frac{\sum A^+_{sat,T}}{\sum A^+_{sat,T} + \left|\sum A^-_{sat,T}\right|} \cdot 100 \quad \text{(eqn 8)}$$

$$\%\ C\ lost = \frac{\left|\sum A^-_{sat,T}\right|}{\sum A^+_{sat,T} + \left|\sum A^-_{sat,T}\right|} \cdot 100 \quad \text{(eqn 9)}$$

These calculations were done by assuming that: (1) all high temperatures are experienced under high light conditions, (2) A_{sat} at the lowest temperatures (<4°C) is negative (as a conservative worst case scenario) and similar to a ratio of R_D experienced at 5°C (three scenarios were modeled based on 50%, 33.3%, 25% and 5% of R_D at 5°C), and (3) the water content of the measured specimens allowed optimum gas exchange.

Statistical Analysis

All analyses were performed using the R statistical software (R Core Team, 2015). The packets used were: *plyr* (Wickham, 2011), *ggplot2* (Wickham, 2016), *nmle* (Pinheiro et al., 2019), and *agricolae* packages (de Mendiburu, 2009).

Microclimate Data

Differences between species in microclimate data (see Daily Moss Surface Temperature Near Casey Station, East Antarctica and Daily Moss Surface Temperature on Fildes Peninsula and Ardley Island, Maritime Antarctica) were tested after logarithmic transformation of daily mean, maximum, and minimum temperatures by analysing a mixed ANOVA where localization of the sensors and date were considered as random variables and species as fixed factor. One-way ANOVA was performed to test the effect of the water content on the moss surface temperature (daily mean, maximum and minimum). The relationships between PPFD or wind gust speed and the difference between moss surface temperature and air temperature were tested by Pearson correlation test.

Chlorophyll Fluorescence and Gas Exchange Data

ETR_{max}/temperature curves were fitted to a 3-degree polynomial equation. Optimum temperatures for ETR_{max} were obtained for each sample of the studied species by determining where the 1st derivative of fitted polynomials was zero. Then, one-way ANOVA was used to test differences between species in optimum temperature for ETR_{max}. The effect of temperature on gas exchange derived parameters in each species was tested by two-way ANOVA. The relationships between the parameters l_b, l_m, αβ, and F_v/F_m with temperature were tested by Pearson correlation test.

RESULTS

Microclimatic Conditions and Moisture Effect

Daily Moss Surface Temperature Near Casey Station, East Antarctica

Microclimate data of *B. pseudotriquetrum*, *C. purpureus*, and *S. antarctici* recorded during the summer of 2002/3 near Casey Station (ASPA 135) are shown in **Figure 3**. No significant effect of species on daily mean, maximum, and minimum temperatures was observed (P = 0.564, 0.995, and 0.795, respectively). The highest mean temperatures of the mosses surface were obtained between 10:00 and 13:00 (local time, UTC +13) and reached values around 11°C in the three species. Although maximum mean temperatures were only on average 2.3°C higher than mean surface temperatures, absolute maximum temperatures of 19, 18, and 17°C were recorded at midday for *B. pseudotriquetrum*, *C. purpureus*, and *S. antarctici*, respectively. At night, mean surface temperatures remained positive in the three studied species and only absolute minimum temperatures declined to -1°C. At midday, absolute daily minimum moss surface temperature was never below +4°C. During 13 days of measurements over the peak Antarctic summer (January), the mosses experienced temperatures below +4°C for 56.6% of the time. Moss temperatures exceeded 14°C for an average of just 2.5% of the time.

The difference between moss surface temperature and air temperature at midday was predominantly driven by solar radiation (**Figure 4A**, P<0.05, R^2 = 0.546) and in a weaker but significant way by maximum speed of wind gust (**Figure 4B**,

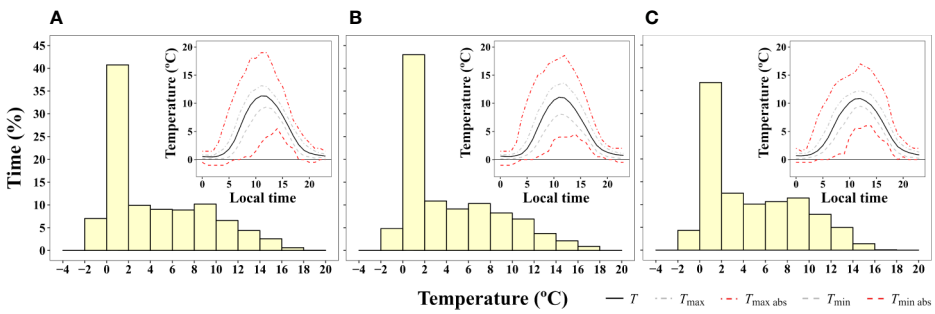


FIGURE 3 | Frequency of time (% of hours) during which a certain two degree interval of surface temperature was experienced by **(A)** *B. pseudotriquetrum*, **(B)** *C. purpureus*, and **(C)** *S. antarctici* averaged over 13 days during the 2003 summer season (from 16/01/2003 to 28/01/2003). Moss temperature was recorded with iBcod sensors placed in moss beds at six locations within 100 m of Casey Station. Inset graph shows diurnal course of surface temperature.

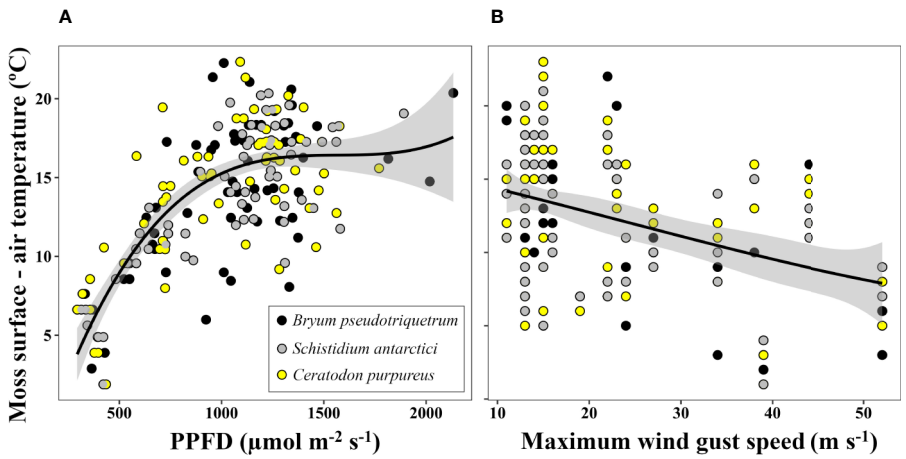


FIGURE 4 | Relationship between the increase of moss surface temperature relative to air temperature and **(A)** solar photosynthetic photon flux density (PPFD) (P<0.05, R^2 = 0.546 for logarithmic transformed data) and **(B)** daily maximum wind gust speed (P<0.05, R^2 = 0.165). Lines represent quadratic polynomial fittings of temperature with their respective 95% confidence intervals (shaded areas). Moss surface temperature and PPFD were measured at 12:00 (UTC +13) with an infrared thermometer and a LS-C mini quantum sensor, respectively. Air temperature at this time and daily maximum wind gust speed were obtained from the Australian Bureau of Meteorology at Casey station. Data represent eight days of measurement within a period of 22 days (from 09/11/2002 to 01/02/2003) at three locations (one daily measurement per species and location).

$P<0.05$, $R^2 = 0.165$). On clear days, when irradiation exceeded 1,000 $\mu mol \cdot m^{-2} \cdot s^{-1}$, air temperature and moss surface temperature reached a mean maximum difference of 16.2°C for all three species, with the highest absolute maximum difference recorded for *C. purpureus* (22.3°C).

Daily Moss Surface Temperature on Fildes Peninsula and Ardley Island, Maritime Antarctica

Microclimate data for the maritime Antarctic species studied during summer 2019 on Fildes Peninsula and Ardley Island are shown in **Figure 5**. No significant differences between species surface temperatures were found ($P = 0.785$, 0.461, and 0.972 for mean, maximum and minimum moss surface temperature, respectively). As in the Windmill Islands, the highest mean moss surface temperatures were obtained at midday and reached values around 8.6°C. Absolute maximum temperatures of 29.7 and 34.2°C were measured in *C. aciphyllum* and *S. uncinata*, respectively, meanwhile *B. pseudotriquetrum* and *P. alpinum* showed absolute maximums around 20.4°C. In the coldest hours of night, mean surface temperatures dropped to +0.1°C and absolute minimum temperatures declined to -3°C on average. Absolute minimum temperatures remained close to 0°C at midday. As in Windmill Islands, the studied mosses experienced temperatures below +4°C most of the time (66.7%) and moss surface temperature only exceeded 14°C for 3% of the time.

Environmental Moisture Effect on Moss Surface Temperature

Mean maximum and minimum surface temperature of *S. antarctici* were significantly affected by hydration status (**Figure 6**). Thus, dry canopy temperatures reached significantly higher maximum (15.0 ± 0.9 and 21.9 ± 0.9°C for data of 2011/12 and 2013, respectively) and lower minimum (-3.0 ± 0.4 and -3.1 ± 0.4°C) daily mean temperatures. While in intermediate and wet environments, extreme temperatures were more buffered. Mean temperatures were also significantly higher in dry sites during the summer of 2013 but not in the previous summer. The percentage of time when moss surface temperatures exceeded 14°C was 10.2%–19.5% for dry moss, 10.1%–4.8% for intermediate sites and only 2.4%–1.6% of time for wet sites (data not shown).

Temperature Responses of Chlorophyll Fluorescence and Gas Exchange

ETR light curves at each studied temperature for species from East and maritime Antarctic locations are shown in **Supplementary Figures 6** and **7**. The ETR_{max} from rational and waiting-in-line light curve models (**Supplementary Figure 8**) gave maximum values at temperatures between 19–26.3°C (**Table 2**). The only endemic species studied, *S. antarctici*, had the lowest optimum temperature for ETR (19.00 ± 0.9°C), followed by *C. purpureus* (21.3 ± 1.9°C) a cosmopolitan species. *Polytrichastrum alpinum*, which is associated with polar and alpine habitats, showed the highest optimum temperature (26.3 ± 0.7°C). No significant

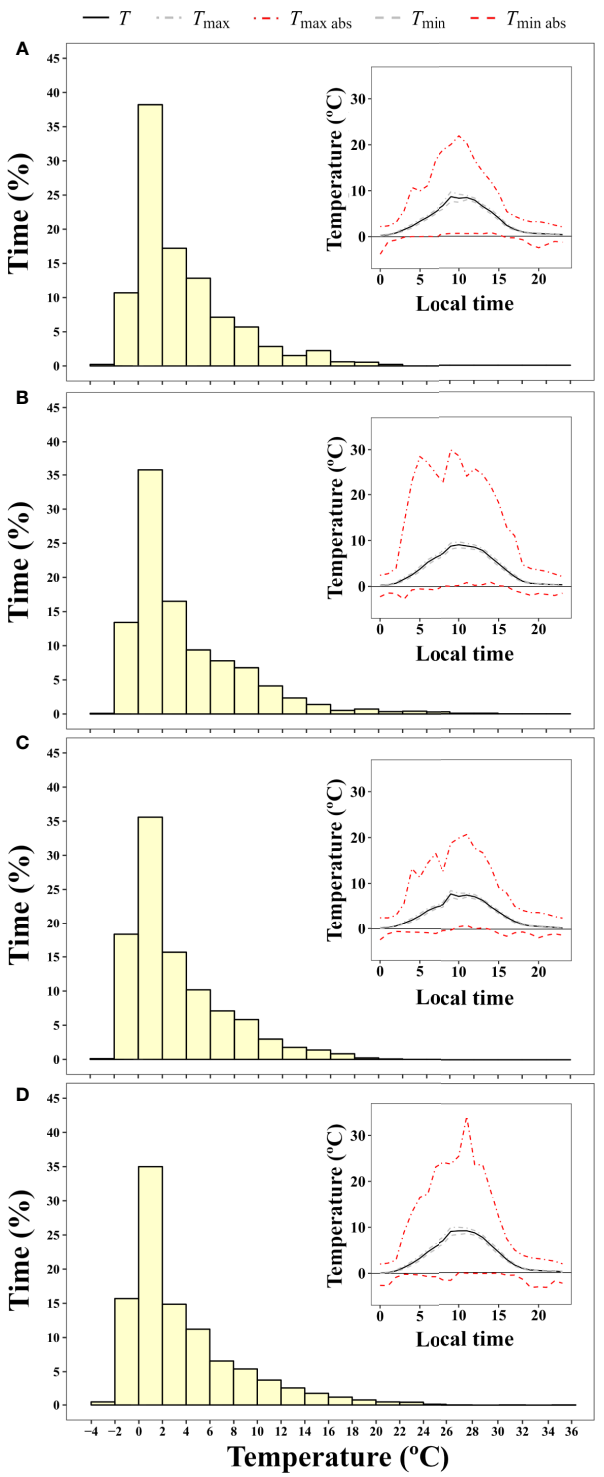


FIGURE 5 | Frequency of time (% of hours) during which a certain two degree interval of surface temperature was experienced by **(A)** *B. pseudotriquetrum*, **(B)** *C. aciphyllum*, **(C)** *P. alpinum*, and **(D)** *S. uncinata* averaged for 20 days over the 2019 summer season (from 07/01/2019 to 29/01/2019). Thermocouple dataloggers were placed in moss beds at three to nine locations per species around Ardley Island. Inset graphs show diurnal course of surface temperature.

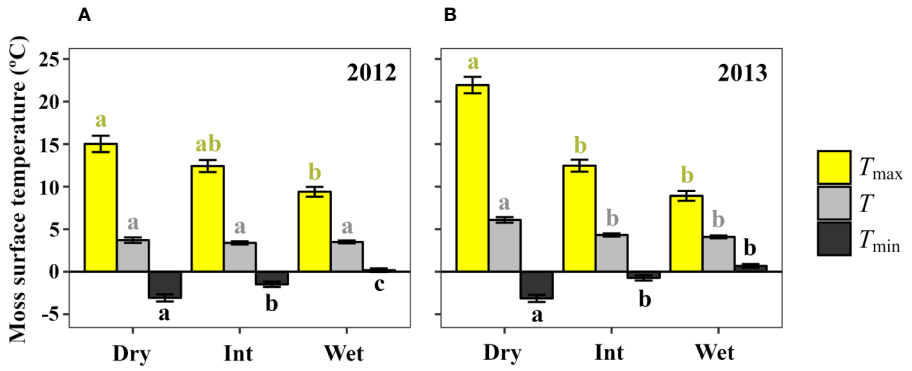


FIGURE 6 | Daily moss surface temperature mean (T), maximum (T_{max}) and minimum (T_{min}) in dry, intermediate and wet environments (3 sensors each) across a *S. antarctici* moss bed near Casey station (East Antarctica) during two summer seasons: data from **(A)** 29/12/2011 to 26/01/2012 (87 days), and **(B)** 13/01/2013 to 31/01/2013 (54 days). Data are means ± se. Different letters denote significant differences between environments by Duncan *post-hoc* ($P < 0.05$).

TABLE 2 | Optimum temperatures (T_{opt}) of ETR_{max} for the studied species.

Loc	Year	Species	T_{opt}
Casey station	2011/12	*B. pseudotriquetrum*	25.3 ± 1.6^{ab}
		C. purpureus	21.3 ± 1.9^{bc}
		S. antarctici	19.0 ± 0.9^{c}
Fildes Peninsula	2015	*B. pseudotriquetrum*	24.1 ± 0.7^{ab}
		S. antarctici	22.8 ± 1.2^{abc}
Ardley Island		*P. alpinum*	26.3 ± 0.7^{ab}
		C. aciphyllum	23.8 ± 0.4^{ab}
		S. uncinata	24.1 ± 0.7^{ab}

Data from Casey station: Dec/Jan 2011/12 (n = 5). Data from Fildes Peninsula and Ardley Island (King George Island): January 2015 (n = 6). Data are means ± se. Letters: significant differences between species by Duncan post-hoc (P < 0.05).

difference was found between optimum temperatures of *B. pseudotriquetrum* measured at the two study sites.

Figure 7 shows the change in gas exchange and associated fluorescence parameters with temperature. The highest A_{sat} were recorded at 25–30°C by *B. pseudotriquetrum* and *C. purpureus*, meanwhile *S. antarctici* showed its optimum A_{sat} at 20–25°C (**Figure 7A**). Interestingly, *B. pseudotriquetrum* could not maintain a positive carbon balance at 5 or 10°C. R_D was also strongly inhibited at low temperatures, showing reductions of around 80% at 5°C in the three studied species. Conversely maximum values were found at the highest tested temperature, 35°C (**Figure 7B**).

The optimum temperatures for electron transport were also above 25°C in the three species studied (**Figure 7C**). The absorbance of PSII used for calculating ETR did not vary significantly across temperature ($R^2 < 0.1$ in all species). Maximum NPQ occurred at 20°C or higher temperatures (**Figure 7D**), whereas F_v/F_m varied between 0.574 and 0.794 independent of the temperatures and the species ($R^2 < 0.09$ in all cases).

The estimated mesophyll conductance also showed maximum values at high temperatures – *B. pseudotriquetrum*, *C. purpureus* and *S. antarctici* showed optimum g_m values at >30, 30, and 15–35°C, respectively (**Figure 8**). g_m values at 5 and 10°C were not significantly different from zero. At low temperatures the diffusional limitation due to the mesophyll (l_m) increased significantly in opposition to biochemical limitation (l_b, $P < 0.001$ in *B. pseudotriquetrum*, $P = 0.017$ and 0.008 in *S. antarctici* and *C. purpureus*, respectively). Q_{10} of g_m in *B. pseudotriquetrum*, *C. purpureus*, and *S. antarctici* were calculated as 1.38, 6.6, and 1.31, respectively (using mean values of g_m between 15 and 25°C).

Estimation of Carbon Gain

Based on the temperature data from the field (**Figure 3**), Windmill Island moss surface temperatures only exceeded 14°C for 2.5% of the time during midsummer. However, modeling of carbon gain using temperature and gas exchange data shown in **Figure 7A** indicate that 36.7%, 12%, and 8.4% of total positive net CO_2 was fixed during this short period in *B. pseudotriquetrum*, *C. purpureus*, and *S. antarctici*, respectively (data not shown). The total balance of positive and negative net CO_2 exchanged (lost *vs* fixed) is shown in **Table 3**. A positive carbon balance –*i.e.* when carbon fixation exceeds 50% of the CO_2 exchanged– was only obtained when inhibition of respiration is high for temperatures below 4°C (33.3% and 25% of the R_D at 5°C in *C. purpureus* and *S. antarctici* and 5% in *B. pseudotriquetrum*).

When carbon balance was modeled for *S. antarctici* across a hydrological gradient during the summer of 2011/12 and 2013 in East Antarctica (**Figure 6**), positive carbon balance was obtained for dry and intermediate environments under most R_D scenarios (except when 50% of R_D at 5°C was modeled for dry environment in 2011/12 and intermediate moist environment in 2013) (**Table 4**). On the contrary, the wet environments only presented positive carbon balance when the inhibition of R_D was the highest modeled (5% of R_D at 5°C in 2011/12 and below 25% in 2013).

DISCUSSION

In line with previous research, our study verified that Antarctic bryophytes are not psychrophilic plants, since all the measured species presented optimum temperatures for ETR_{max} and A_{sat}

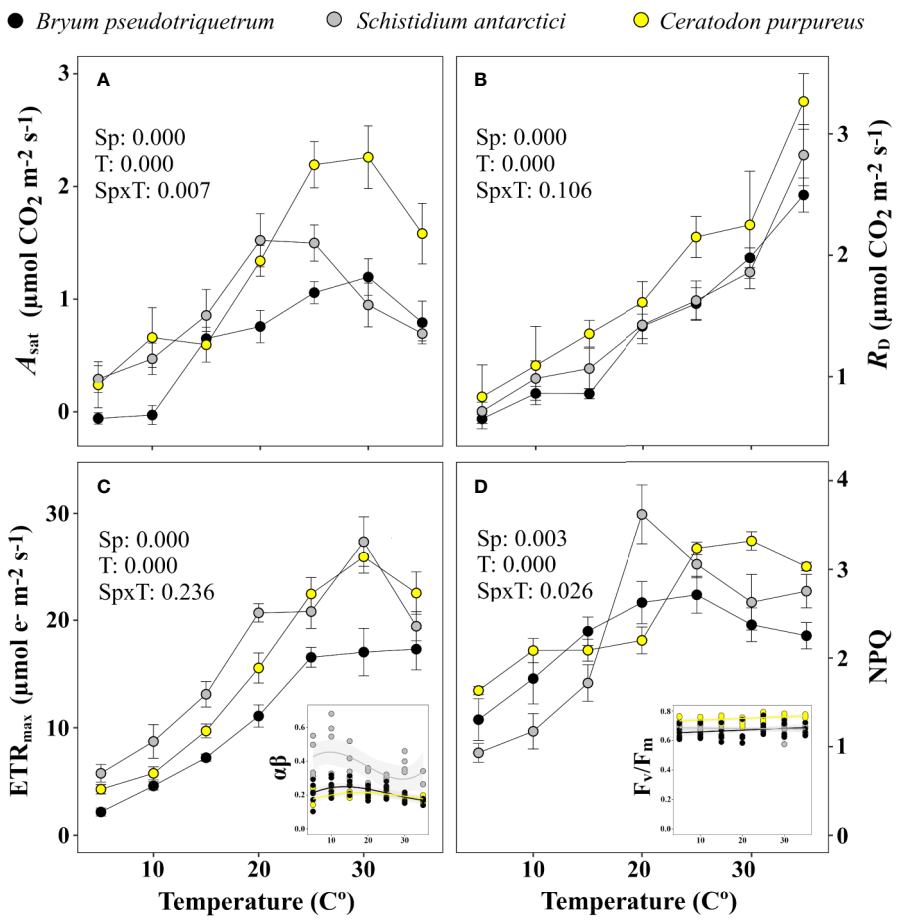


FIGURE 7 | Temperature response curve of three East Antarctic moss species for **(A)** saturating net CO_2 assimilation (A_{sat}), **(B)** dark respiration (R_D), **(C)** maximum electron transport rate (ETR_{max}), and **(D)** non-photochemical quenching (NPQ). Inset graph of **(C)** shows variation in the product of absorbance and partitioning of photons (αβ) used for calculation of ETR across temperature (R^2 = 0.068, 0.063, and 0.095 for *C. purpureus*, *S. antarctici*, and *B. pseudotriquetrum*, respectively). Inset graph of **(D)** shows stability of F_v/F_m across temperature (R^2 = 0.086, 0.028 and 0.046 for *C. purpureus*, *S. antarctici* and *B. pseudotriquetrum*, respectively). *P* values results from two-way ANOVA are shown in the upper-left corner of each graph. Mean ± se (*n* = 4-6, except for the light curve derived parameter αβ, where *n* = 3-5).

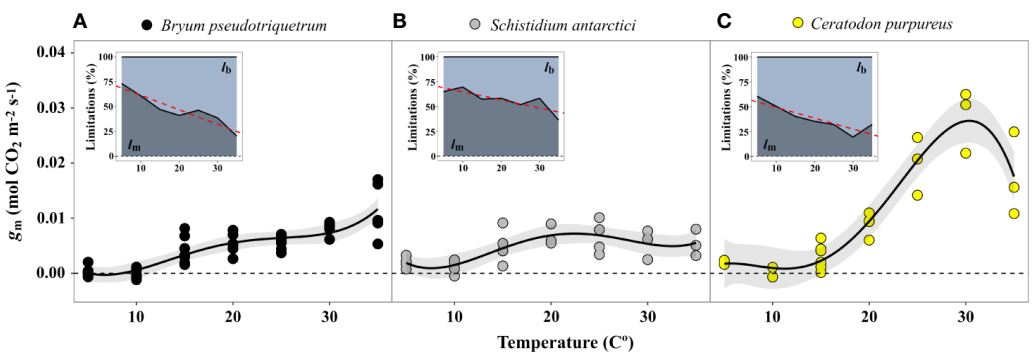


FIGURE 8 | Temperature response of mesophyll conductance (g_m) of **(A)** *B. pseudotriquetrum*, **(B)** *S. antarctici*, and **(C)** *C. purpureus*. Lines represent quadratic polynomial fittings of g_m with their respective 95% confidence intervals (shaded areas, *n* = 3–5). Inset graphs show the temperature relationship of the percentage of biochemical (l_b) *vs* mesophyll limitations (l_m) to photosynthesis. Linear regression of l_m-temperature relation is represented by a red dashed line ($P < 0.05$).

TABLE 3 | Estimated carbon balance for three East Antarctic species calculated from surface temperature data (see *Daily Moss Surface Temperature Near Casey Station, East Antarctica*) and temperature response of A_{sat} (see *Mesophyll Conductance and NPQ Measurement*).

Species	$\%R_D$	% C lost	% C fixation
B. pseudotriquetrum	50%	82.2	17.8
	33.3%	76.0	24.0
	25%	70.9	29.1
	5%	41.5	**58.4**
C. purpureus	50%	55.9	44.1
	33.3%	45.8	**54.2**
	25%	38.8	**61.2**
	5%	11.2	**88.8**
S. antarctici	50%	52.4	47.6
	33.3%	42.3	**57.7**
	25%	35.5	**64.5**
	5%	9.9	**90.1**

For temperatures below 4°C a conservative negative net assimilation of 50%, 33.3%, 25%, or 5% of the respiration in the dark (R_D) at 5°C was assumed. Bold shows positive carbon balance.

TABLE 4 | Estimated carbon balance for East Antarctic *S. antarctici* calculated from surface temperature data (see *Environmental Moisture Effect on Moss Surface Temperature*) and temperature response of A_{sat} (see *Mesophyll Conductance and NPQ Measurement*).

Environment	Year	$\%R_D$	% C lost	% C fixation
Dry	2011/12	50%	50.3	49.7
		33.3%	40.2	**59.8**
		25%	33.6	**66.4**
		5%	9.2	**90.8**
	2013	50%	35.1	**64.9**
		33.3%	26.5	**73.5**
		25%	21.3	**78.7**
		5%	5.1	**94.9**
Int	2011/12	50%	44.9	**55.1**
		33.3%	35.2	**64.8**
		25%	28.9	**71.1**
		5%	7.5	**92.5**
	2013	50%	53.2	46.8
		33.3%	43.1	**56.9**
		25%	36.2	**63.8**
		5%	10.2	**89.8**
Wet	2011/12	50%	75.4	24.6
		33.3%	67.1	32.9
		25%	60.5	39.5
		5%	23.4	**76.6**
	2013	50%	61.0	39.0
		33.3%	51.1	48.9
		25%	43.9	**56.1**
		5%	13.5	**86.5**

For temperatures below 4°C a conservative negative net assimilation of 50%, 33.3%, 25%, or 5% of the respiration in the dark (R_D) at 5°C was assumed. Bold shows positive carbon balance.

between 19 and 26.3°C, as has been reported for mesic and tropical bryophytes (Dilks and Proctor, 1975; Furness and Grime, 1982; Glime, 2011; Wagner et al., 2013; He et al., 2016). Despite the relationship between mean temperatures experienced during the growing season and temperature optima of net photosynthesis reported from polar, alpine, temperate, desert and tropical ecosystems (Wagner et al., 2013), high specific plasticity can be observed for Antarctic species (**Table 5**). The lowest optimum temperature has been found in the endemic species *S. antarctici* between 0 and 10°C (Kappen et al., 1989; Davey and Rothery, 1997; Block et al., 2009), although in the present study ETR_{max} of this species was obtained between temperatures of 19–30°C, also previously reported for O_2 evolution (30°C, Wilson, 1990). High optima for photosynthesis have also been found in non-endemic species with polar distributions, such as *Hennediella heimii* (previously *Bryum antarcticum*) (19°C, Rastorfer, 1970), or species such as *P. alpinum* which show a moderate bipolar distribution with transitional populations in alpine environments, (T_{opt} = 26.3°C here), but note (T_{opt} = 10°C) in a previous study by Davey and Rothery (1997). This suggests that possessing low optimum temperatures for photosynthesis is not a strict requirement for surviving in Antarctic environments (at least until 63–64°S), even in species that restrict their distributions to these habitats (e.g., *S. antarctici*).

The fact that all Antarctic mosses measured showed such high temperature optima for photosynthesis even when summer mean maximum temperatures are much cooler (2.3°C in January at Casey Station) suggests that moss surface temperatures must regularly exceed air temperatures. This was evidenced in our microclimate analysis where the temperature experienced by mosses was measured with iBCods, infrared thermometers and/or thermocouples. Ecological researchers have pointed out the importance of a deep description of microclimate in understanding and modeling present and future species distribution and ecosystem functioning, specially in small-stature species (Convey et al., 2018; Lembrechts et al., 2019; Lembrechts and Lenoir, 2020). In our study, around midday moss surfaces were elevated above mean air temperatures by 16.2 (22.3)°C (**Figure 4**) enhanced by high irradiation and low wind speed, as has been described for Arctic and alpine ecosystems (Wilson, 1957; Körner, 2003). However, most of the time Antarctic mosses experienced suboptimal conditions for photosynthesis (during the night or cloudy/windy days) such that their surface temperatures exceed 14°C only 2.5%–3% of the time. This is a lower percentage of time than reported by Smith (1988), where moss temperatures exceeded 20°C 24% of the time. Even so, the percentage of time at which the temperature of mosses allows a positive net CO_2 assimilation must be enough to compensate for loss of carbon by respiration in order to achieve the very low growth rates (average 1.33 mm per year) reported for these Antarctic mosses (Clarke et al., 2012). In future, under climate change, Antarctic mosses are also expected to experience an increase in air temperatures and this would be expected to lead to an increase in the percentage of time they spend at optimal temperatures (Robinson et al., 2020). Our current estimation of carbon balance suggests that carbon balance can only be positive if a large reduction of carbon loss by respiration is assumed for the lowest temperatures (**Figure 3**). In environments with high nocturnal temperatures, such as tropical regions, bryophytes can lose more than 50% of the CO_2 fixed during the daytime each night (Zotz et al., 1997). So, moss survival in Antarctica may be more related to an ability to

TABLE 5 | Temperature optimum (T_{opt}) for photosynthesis for a range of Antarctic bryophytes measured under field and laboratory conditions and as CO_2 assimilation, O_2 evolution, or ETR.

Species	T_{opt} (°C)	Method	Reference
Andreaea depressinervis	15–20	Net CO_2 uptake	Davey and Rothery, 1997
Andreaea gainii	10–15	Net CO_2 uptake	Davey and Rothery, 1997; Block et al., 2009
Brachytecium austro-salebrosum	15	Net CO_2 uptake	Davey and Rothery, 1997; Block et al., 2009
Bryum antarcticum (Hennediella heimii)	19	O_2 evolution	Rastorfer, 1970
Bryum argenteum	15	Net CO_2 uptake	Green et al., 2000
	≥20	O_2 evolution	Smith, 1999
	25	O_2 evolution	Rastorfer, 1970
Bryum pseudotriquetrum	12.0	Net CO_2 uptake	Pannewitz et al., 2005
	≥20	O_2 evolution	Ino, 1990; Smith, 1999
	25–30	Net CO_2 uptake	Present study
	25	ETR	Present study (data of 2012)
	24	ETR	Present study (data of 2015)
Bryum subrotundifolium	13.7	Net CO_2 uptake	Pannewitz et al., 2005
Calliergon sarmentosum	≥20	Net CO_2 uptake	Davey and Rothery, 1997
Cephaloziella varians	≥20	O_2 evolution	Newsham, 2010
Ceratodon purpureus	6.6	Net CO_2 uptake	Pannewitz et al., 2005
	≥20	O_2 evolution	Smith, 1999
	< 15	Net CO_2 uptake	Ino, 1990
	≥20	Net CO_2 uptake	Davey and Rothery, 1997
	≥20	Net CO_2 uptake	Block et al., 2009
	25–30	Net CO_2 uptake	Present study
	21	ETR	Present study
Chorisodontium aciphyllum	10–20	Net CO_2 uptake	Davey and Rothery, 1997
	24	ETR	Present study
Drepanocladus uncinatus	≥20	Net CO_2 uptake	Davey and Rothery, 1997
Marchantia berteroana	15	Net CO_2 uptake	Davey and Rothery, 1997; Block et al., 2009
Polytrichum alpestre	15–20	Net CO_2 uptake	Davey and Rothery, 1997
Polytrichum alpinum	10	Net CO_2 uptake	Davey and Rothery, 1997
	26	ETR	Present study
Polytrichum strictum	10	Net CO_2 uptake	Longton, 1988a
Racomitrium austro-georgicum	10–20	Net CO_2 uptake	Davey and Rothery, 1997
	10	Net CO_2 uptake	Block et al., 2009
*Sanionia uncinata**	≥20	Net CO_2 uptake	Block et al., 2009
	24	ETR	Present study
Schistidium antarctici	5–10	Net CO_2 uptake	Kappen et al., 1989
	30	O_2 evolution	Wilson, 1990
	0–10	Net CO_2 uptake	Davey and Rothery, 1997
	10	Net CO_2 uptake	Block et al., 2009
	20–25	Net CO_2 uptake	Present study
	19–23	ETR	Present study
Tortula saxicola	10–20	Net CO_2 uptake	Davey and Rothery, 1997
Warnstorfia sarmentosum	≥20	Net CO_2 uptake	Block et al., 2009

**Data from Nakatsubo (2002) and Kallio and Heinonen (1975) are excluded because measurements of over-watered samples were clearly performed at non-steady state conditions and variation of CO_2 assimilation might be associated with a variation of interstitial water during measurements rather than with temperature.*

inhibit respiration at low temperatures, rather than having lower optimum temperatures for photosynthesis.

Another factor that affects carbon gain of many bryophytes is water availability (Proctor, 1982). Both dehydration and an excessive interstitial water content can inhibit photosynthesis (Smith, 1982; Rice et al., 2011; Wagner et al., 2013; **Supplementary Figure 5**). Our study confirms that water content also influences the temperatures experienced by Antarctic mosses, with excess water buffering the extremes (decreasing their maximum and increasing their minimum temperatures) and, therefore, reducing the time when mosses experience temperatures higher than 14°C. This can be explained by the high specific heat capacity of water, as has been suggested previously by Pannewitz et al. (2005) and Block et al. (2009) for Antarctic mosses and by soil researchers (Campbell et al., 1995). Thus, provided the moss cells remain hydrated and the drier the interstitial environment, the wider the window for positive net CO_2 assimilation, since maximum temperatures are closer to optimum for photosynthesis and minimum temperatures are enough to substantially inhibit the loss of CO_2 by respiration (see **Figure 4** for estimation of carbon balance). However, the optimum water content for maximum carbon gain of these Antarctic mosses is still unknown, and the interaction of limitations by both temperature and water content should be analyzed in the future.

The effect of low temperatures on photosynthesis in the Antarctic species studied here was mainly driven by diffusional limitations, rather than biochemical ones, as has been reported for Antarctic vascular plants (Sáez et al., 2018) and by other important environmental stresses such as water stress (Flexas

et al., 2018; Nadal and Flexas, 2018). Variable responses of g_m to temperature have been reported for vascular plants (Bernacchi et al., 2002; von Caemmerer and Evans, 2015; Xiong et al., 2015; Huang et al., 2017). In our study, *B. pseudotriquetrum*, *S. antarctici*, and *C. purpureus* showed increasing g_m with temperature and only in the latter was a decline at supra-optimal temperatures observed. The components of mesophyll conductance that rule its response to temperature are not well understood (Shrestha et al., 2019). Since the temperature coefficient reported for CO_2 diffusion in pure water (Q_{10} = 1.25, Jähne et al., 1987) is lower than that reported for g_m in vascular plants (1.8-2.2) (Bernacchi et al., 2002; Yamori et al., 2006), physical diffusion alone cannot explain the variation of g_m with temperature. Instead, it has been hypothesised that CO_2 diffusion through both liquid phase (cell wall, cytosol and chloroplast stroma) and membranes (plasmic and chloroplastic, facilitated by protein transporters, i.e., aquaporins) is affected by temperature (Bernacchi et al., 2002; Evans and Von Caemmerer, 2013; Walker et al., 2013; von Caemmerer and Evans, 2015). The Q_{10} of *C. purpureus*, *B. pseudotriquetrum*, and *S. antarctici* was calculated as 6.6, 1.38, and 1.31, respectively. This suggests that the role of any facilitated process for CO_2 diffusion is highly variable and is enhanced more in *C. purpureus* than in either the latter two species or the reported vascular plants.

At suboptimal temperatures photosynthesis in these mosses is unlikely to be able to utilise all the absorbed light producing an energy imbalance. Despite this, photoprotective heat dissipation (here estimated by NPQ) decreased when saturating light was combined with short-term exposure to suboptimal temperatures (optimum temperatures for NPQ = 20°C for *S. antarctici* and 25°C for *C. purpureus* and *B. pseudotriquetrum*). Lovelock et al. (1995) reported similar results for *S. antarctici*, which decreased NPQ (expressed as q_N) after 2 h of 5–0°C. Only when temperatures were below the freezing point of -7°C (Melick and Seppelt, 1992), was de-epoxidation-independent (dithiothreitol insensitive) NPQ significantly increased, as has been observed for Antarctic lichens (Barták et al., 2007). Previous research into short-term changes of NPQ with suboptimal temperature (both cold and heat stress) in mesic and Antarctic species has reported various results including: (1) an increase in maximum steady-state NPQ (Xu et al., 1999; Hendrickson et al., 2004; Sinsawat et al., 2004; D'Ambrosio et al., 2006; Savitch et al., 2009; Sharkey and Zhang, 2010), (2) a decrease of NPQ, as in our study, (Bilger and Björkman, 1991; Fracheboud and Leipner, 2003; Corcuera et al., 2005; Lambrev et al., 2007; Pérez-Torres et al., 2007; Wang et al., 2009), or (3) invariable NPQ (Pérez-Torres et al., 2007). No correlation between de-epoxidase state or zeaxanthin concentration and NPQ at low temperatures (Xu et al., 1999; D'Ambrosio et al., 2006) and a decrease in the percentage of NPQ inhibited by dithiothreitol (Xu et al., 1999) suggest that the NPQ that is enhanced in the short-term at low temperatures could consist of zeaxanthin-independent heat dissipation (Johnson et al., 2009) and/or photoinactivated PSII reaction centre heat dissipation (Krause and Weis, 1988; Lee et al., 2001; Ivanov et al., 2003). This would be consistent with the fact that violaxanthin de-epoxidase is inhibited at low temperatures (Bilger and Björkman, 1991; Szilágyi et al., 2007) and the associated lower electron transport rates will compromise the generation of ΔpH, which is required for activation of violaxanthin de-epoxidase and inhibition of zeaxanthin epoxidase (Gilmore, 1997; Goss et al., 2008). The decrease of NPQ at low temperatures in our study could be a direct consequence of the inhibitory processes described above and not/less related with zeaxanthin-independent or photoinactivated PSII heat dissipation. Furthermore, maintenance of a constitutive zeaxanthin/lutein pool, which is a common mechanism for enhancing NPQ during cold-hardening (Haldimann et al., 1996; Leipner et al., 1997; Faria et al., 1998; Venema et al., 2000; Caffarri et al., 2005; Ivanov et al., 2006; Sáez et al., 2019; but see also Savitch et al., 2002), is less likely to be present in these short-term experiments. East Antarctic *S. antarctici*, *C. purpureus*, and *B. pseudotriquetrum* exhibit high zeaxanthin content in the field as a result of this cold hardening (Lovelock and Robinson, 2002). However, the presence of zeaxanthin alone without the generation of a pH gradient is insufficient to induce fluorescence quenching (Bilger and Björkman, 1991; Hurry et al., 1997; Hwang et al., 2003; Goss et al., 2008). Thus, questions remain about the photoprotection role of high zeaxanthin levels in Antarctic mosses at low temperatures, given they are not associated with the build-up of a transthylakoidal ΔpH which normally induces regulated heat dissipation. Screening and/or antioxidant roles of zeaxanthin may need to be considered (Havaux et al., 2007; Solovchenko, 2010). However, at high temperatures and high light levels, a high and sustained concentration of zeaxanthin could help to enable a rapid photoprotection response of heat dissipation at temperatures close to the photosynthetic optima.

CONCLUSION

We conclude that Antarctic mosses are not psychrophilic plants, since their photosynthetic optima occur at relatively high temperatures. A positive carbon gain can be maintained providing respiration is strongly inhibited at low temperatures. However, the interaction of limitations by both temperature and water require further study, since the moisture of the moss environment influences the temperatures at which they metabolize. At low temperatures, NPQ was not enhanced and the decline in photosynthesis was largely caused by an increase of diffusional limitations, which also suggests the existence of facilitated and variable processes for CO_2 diffusion in these mosses.

DATA AVAILABILITY STATEMENT

All data associated with this manuscript is available in Australian Antarctic Data Centre (https://data.aad.gov.au/metadata/records/AAS_4046_TempOptima_Frontiers_Perera-Castro).

AUTHOR CONTRIBUTIONS

AP-C designed and conducted all photosynthetic experiments in the laboratory in Australia and wrote the first draft of the manuscript. MW, SR, AC-K, and GZ collected samples and conducted experiments at the Instituto Antártico Chileno Profesor Julio Escudero Station. EM, JW, JT, and JB-A collected samples and performed Windmill Island experiments. AP-C, JF, and MA performed analyses. All authors contributed to the article and approved the submitted version.

FUNDING

The Spanish Ministry of Education, Culture and Sport (MECD) supported pre-doctoral fellowship (FPU-02054) awarded to AP-C. Research on King George Island was supported by the Instituto Antártico Chileno (INACH Grant RT-2716 to AC-K) and the National Fund for Scientific and Technological Development (FONDECYT grant 118745 to AC-K). Research at Casey was supported by Australian Antarctic Science Grant 4046 (SR). Research was funded by ARC DP110101714 (SR) and DP180100113 (SR) and a Global Challenges Project Grant ECO-Antarctica (GC Project ID 91 to MW).

ACKNOWLEDGMENTS

We are grateful for Station logistic support provided at both Profesor Julio Escudero Station (INACH) and at Casey Station (Australian Antarctic Division). We also thank Dr. Elena Baraza for her help with statistical analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01178/full#supplementary-material.

REFERENCES

Barták, M., Váczi, P., Hájek, J., and Smykla, J. (2007). Low-temperature limitation of primary photosynthetic processes in Antarctic lichens *Umbilicaria antarctica* and *Xanthoria elegans*. *Polar Biol.* 31, 47–51. doi: 10.1007/s00300-007-0331-x

Bernacchi, C. J., Portis, A. R., Nakano, H., Von Caemmerer, S., and Long, S. P. (2002). Temperature response of mesophyll conductance. Implications for the determination of Rubisco enzyme kinetics and for limitations to photosynthesis in vivo. *Plant Physiol.* 130, 1992–1998. doi: 10.1104/pp.008250

Biermsa, M. E., Convey, P., Wyber, R., Robinson, S. A, Dowton, M., Vijver, B. V., et al (2020). Latitudinal biogeographic structuring in the globally distributed moss Ceratodon purpureus. *Front. Plant Sci.*

Bilger, W., and Björkman, O. (1990). Role of the xanthophyll cycle in photoprotection elucidated by measurements of light-induced absorbance changes, fluorescence and photosynthesis in leaves of *Hedera canariensis*. *Photosynth. Res.* 25, 173–185. doi: 10.1007/BF00033159

Bilger, W., and Björkman, O. (1991). Temperature dependence of violaxanthin de-epoxidation and non-photochemical fluorescence quenching in intact leaves of *Gossypium hirsutum* L. and *Malva parviflora L. Planta* 184, 226–234. doi: 10.1007/BF00197951

Block, W., Smith, R. I. L., and Kennedy, A. D. (2009). Strategies of survival and resource exploitation in the Antarctic fellfield ecosystem. *Biol. Rev.* 84, 449–484. doi: 10.1111/j.1469-185X.2009.00084.x

Bramley-Alves, J., King, D. H., Robinson, S. A., and Miller, R. E. (2014). "Dominating the Antarctic environment: bryophytes in a time of change," in *Photosynthesis in bryophytes and early land plants*. Eds. D. T. Hanson and S. K. Rice (Dordrecht: Springer), 309–324.

Bramley-Alves, J., Wanek, W., French, K., and Robinson, S. A. (2015). Moss $\delta^{13}C$: an accurate proxy for past water environments in polar regions. *Glob. Change Biol.* 21, 2454–2464. doi: 10.1111/gcb.12848

Bunce, J. A. (2008). Acclimation of photosynthesis to temperature in *Arabidopsis thaliana* and *Brassica oleracea*. *Photosynthetica* 46, 517–524. doi: 10.1007/s11099-008-0088-7

Caffarri, S., Frigerio, S., Olivieri, E., Righetti, P. G., and Bassi, R. (2005). Differential accumulation of Lhcb gene products in thylakoid membranes of *Zea mays* plants grown under contrasting light and temperature conditions. *Proteomics* 5, 758–768. doi: 10.1002/pmic.200402008

Campbell, G. S., Jungbauer, J. D. Jr., Bristow, K. L., and Hungerford, R. D. (1995). Soil temperature and water content beneath a surface fire. *Soil Sci.* 159, 363–374. doi: 10.1097/00010694-199506000-00001

Carrasco, J., and González, M. (2007). *Climatología de la Península Antártica y de la base presidente Eduardo Frei Montalva* (Santiago: Dirección Meteorológica de Chile).

Carriquí, M., Roig-Oliver, M., Brodribb, T. J., Coopman, R., Gill, W., Mark, K., et al. (2019). Anatomical constraints to nonstomatal diffusion conductance and photosynthesis in lycophytes and bryophytes. *New Phyt.* 222, 1256–1270. doi: 10.1111/nph.15675

Clarke, L. J., Robinson, S. A., Hua, Q., Ayre, D. J., and Fink, D. (2012). Radiocarbon bomb spike reveals biological effects of Antarctic climate change. *Global Change Biol.* 18, 301–310. doi: 10.1111/j.1365-2486.2011.02560.x

Convey, P., and Smith, R. I. L. (2005). "Responses of terrestrial Antarctic ecosystems to climate change," in *Plants and Climate Change*. Eds. J. Rozema, R. Aerts and H. Cornelissen (Dordrecht: Springer), 1–12.

Convey, P., Chown, S. L., Clarke, A., Barnes, D. K., Bokhorst, S., Cummings, V., et al. (2014). The spatial structure of Antarctic biodiversity. *Ecol. Monogr.* 84, 203–244. doi: 10.1890/12-2216.1

Convey, P., Coulson, S. J., Worland, M. R., and Sjöblom, A. (2018). The importance of understanding annual and shorter-term temperature patterns and variation in the surface levels of polar soils for terrestrial biota. *Polar Biol.* 41, 1587–1605. doi: 10.1007/s00300-018-2299-0

Corcuera, L., Morales, F., Abadia, A., and Gil-Pelegrín, E. (2005). The effect of low temperatures on the photosynthetic apparatus of *Quercus ilex* subsp. *ballota* at its lower and upper altitudinal limits in the Iberian peninsula and during a single freezing-thawing cycle. *Trees* 19, 99–108. doi: 10.1007/s00468-004-0368-1

Davey, M. C., and Rothery, P. (1997). Interspecific variation in respiratory and photosynthetic parameters in Antarctic bryophytes. *New Phyt.* 137, 231–240. doi: 10.1046/j.1469-8137.1997.00805.x

de Mendiburu, F. (2009). *Una herramienta de análisis estadístico para la investigación agrícola. [Thesis]* (Lima: Universidad Nacional de Ingeniería).

Deltoro, V. I., Calatayud, Á., Morales, F., Abadía, A., and Barreno, E. (1999). Changes in net photosynthesis, chlorophyll fluorescence and xanthophyll cycle interconversions during freeze-thaw cycles in the Mediterranean moss *Leucodon sciuroides*. *Oecologia* 120, 499–505. doi: 10.1007/s004420050883

Dilks, T. J. K., and Proctor, M. C. F. (1975). Comparative experiments on temperature responses of bryophytes: assimilation, respiration and freezing damage. *J. Bryol.* 8, 317–336. doi: 10.1179/jbr.1975.8.3.317

D'Ambrosio, N., Arena, C., and De Santo, A. V. (2006). Temperature response of photosynthesis, excitation energy dissipation and alternative electron sinks to carbon assimilation in *Beta vulgaris* L. *Environ. Exp. Bot.* 55, 248–257. doi: 10.1016/j.envexpbot.2004.11.006

Edwards, J. A., and Smith, R. I. (1988). Photosynthesis and respiration of *Colobanthus quitensis* and *Deschampsia antarctica* from the maritime Antarctic. *Brit. Antarct. Surv. B.* 81, 43–63.

Ensminger, I., Busch, F., and Huner, N. P. (2006). Photostasis and cold acclimation: sensing low temperature through photosynthesis. *Physiol. Plantarum* 126, 28–44. doi: 10.1111/j.1399-3054.2006.00627.x

Evans, J. R., and Von Caemmerer, S. (2013). Temperature response of carbon isotope discrimination and mesophyll conductance in tobacco. *Plant Cell Environ.* 36, 745–756. doi: 10.1111/j.1365-3040.2012.02591.x

Faria, T., Silvério, D., Breia, E., Cabral, R., Abadia, A., Abadia, J., et al. (1998). Differences in the response of carbon assimilation to summer stress (water deficits, high light and temperature) in four Mediterranean tree species. *Physiol. Plantarum* 102, 419–428. doi: 10.1034/j.1399-3054.1998.1020310.x

Flexas, J., Barbour, M. M., Brendel, O., Cabrera, H. M., Carriquí, M., Diaz-Espejo, A., et al. (2012). Mesophyll diffusion conductance to CO_2: an unappreciated central player in photosynthesis. *Plant Sci.* 193, 70–84. doi: 10.1016/j.plantsci.2012.05.009

Flexas, J., Diaz-Espejo, A., Gago, J., Gallé, A., Galmés, J., Gulías, J., et al. (2014). Photosynthetic limitations in Mediterranean plants: a review. *Environ. Exp. Bot.* 103, 12–23. doi: 10.1016/j.envexpbot.2013.09.002

Flexas, J., Carriquí, M., and Nadal, M. (2018). Gas exchange and hydraulics during drought in crops: who drives whom? *J. Exp. Bot.* 69, 3791–3795. doi: 10.1093/jxb/ery235

Font, M., and Galmés, J. (2016). *Temperature response of in vitro Rubisco kinetics in bryophytes and ferns. [Poster]. 17th International Congress on Photosynthesis Research, Maastricht*.

Fortuin, J. P. F., and Oerlemans, J. (1990). Parameterization of the annual surface temperature and mass balance of Antarctica. *Ann. Glaciol.* 14, 78–84. doi: 10.3189/S0260305500008302

Fracheboud, Y., and Leipner, J. (2003). "The application of chlorophyll fluorescence to study light, temperature, and drought stress," in *Practical applications of chlorophyll fluorescence in plant biology*. Eds. J. R. DeEll and P. M. A. Toivonen (Boston: Springer), 125–150.

Furness, S. B., and Grime, J. P. (1982). Growth rate and temperature responses in bryophytes: II. A comparative study of species of contrasted ecology. *J. Ecol.* 1, 525–536. doi: 10.2307/2259920

Gago, J., Carriquí, M., Nadal, M., Clemente-Moreno, M. J., Coopman, R. E., Fernie, A. R., et al. (2019). Photosynthesis optimized across land plant phylogeny. *Trends Plant Sci.* 24, 947–958. doi: 10.1016/j.tplants.2019.07.002

Genty, B., Briantais, J. M., and Baker, N. R. (1989). The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *BBA-Gen. Subj.* 990, 87–92. doi: 10.1016/S0304-4165(89)80016-9

Gilmore, A. M. (1997). Mechanistic aspects of xanthophyll cycle-dependent photoprotection in higher plant chloroplasts and leaves. *Physiol. Plantarum* 99, 197–209. doi: 10.1111/j.1399-3054.1997.tb03449.x

Glime, J. M. (2011). "Ecological and physiological effects of changing climate on aquatic bryophytes," in *Bryophyte Ecology and Climate Change*. Eds. Z. Tuba, N. G. Slack and L. R. Stark (Cambridge: Cambridge University Press), 93–114.

Goss, R., Opitz, C., Lepetit, B., and Wilhelm, C. (2008). The synthesis of NPQ-effective zeaxanthin depends on the presence of a transmembrane proton gradient and a slightly basic stromal side of the thylakoid membrane. *Planta* 228, 999–1009. doi: 10.1007/s00425-008-0800-7

Grassi, G., and Magnani, F. (2005). Stomatal, mesophyll conductance and biochemical limitations to photosynthesis as affected by drought and leaf ontogeny in ash and oak trees. *Plant Cell Environ.* 28, 834–849. doi: 10.1111/j.1365-3040.2005.01333.x

Green, T. G. A., Schroeter, B., and Seppelt, R. D. (2000). "Effect of temperature, light and ambient UV on the photosynthesis of the moss Bryum argenteum Hedw., in continental Antarctica," in *Antarctic Ecosystems: Models for Wider Ecological Understanding*. Eds. W. Davidson, C. Howard-Williams and P. Broady (Christchurch: The Caxton Press), 165–170.

Green, T. A., Kulle, D., Pannewitz, S., Sancho, L. G., and Schroeter, B. (2005). UV-A protection in mosses growing in continental Antarctica. *Polar Biol.* 28, 822–827. doi: 10.1007/s00300-005-0011-7

Haldimann, P., Fracheboud, Y., and Stamp, P. (1996). Photosynthetic performance and resistance to photoinhibition of *Zea mays* L. leaves grown at sub-optimal temperature. *Plant Cell Environ.* 19, 85–92. doi: 10.1111/j.1365-3040.1996.tb00229.x

Harley, P. C., Loreto, F., Di Marco, G., and Sharkey, T. D. (1992). Theoretical considerations when estimating the mesophyll conductance to CO_2 flux by analysis of the response of photosynthesis to CO_2. *Plant Physiol.* 98, 1429–1436. doi: 10.1104/pp.98.4.1429

Havaux, M., Dall'Osto, L., and Bassi, R. (2007). Zeaxanthin has enhanced antioxidant capacity with respect to all other xanthophylls in *Arabidopsis* leaves and functions independent of binding to PSII antennae. *Plant Physiol.* 145, 1506–1520. doi: 10.1104/pp.107.108480

He, X., He, K. S., and Hyvönen, J. (2016). Will bryophytes survive in a warming world? *Perspect. Plant Ecol.* 19, 49–60. doi: 10.1016/j.ppees.2016.02.005

Hendrickson, L., Förster, B., Furbank, R. T., and Chow, W. S. (2004). Processes contributing to photoprotection of grapevine leaves illuminated at low temperature. *Physiol. Plantarum* 12, 272–281. doi: 10.1111/j.0031-9317.2004.0324.x

Hirano, M., Satoh, K., and Katoh, S. (1981). The effect on photosynthetic electron transport of temperature-dependent changes in the fluidity of the thylakoid membrane in a thermophilic blue-green alga. *BBA-Bioenergetics* 635, 476–487. doi: 10.1016/0005-2728(81)90107-9

Holaday, A. S., Martindale, W., Alred, R., Brooks, A. L., and Leegood, R. C. (1992). Changes in activities of enzymes of carbon metabolism in leaves during exposure of plants to low temperature. *Plant Physiol.* 98, 1105–1114. doi: 10.1104/pp.98.3.1105

Hovenden, M. J., Jackson, A. E., and Seppelt, R. D. (1994). Field photosynthetic activity of lichens in the Windmill Islands oasis, Wilkes Land, continental Antarctica. *Physiol. Plantarum* 90, 567–576. doi: 10.1111/j.1399-3054.1994.tb08816.x

Huang, G., Zhang, Q., Wei, X., Peng, S., and Li, Y. (2017). Nitrogen can alleviate the inhibition of photosynthesis caused by high temperature stress under both steady-state and flecked irradiance. *Front. Plant Sci.* 8:945:945. doi: 10.3389/fpls.2017.00945

Huner, N. P., Öquist, G., Hurry, V. M., Krol, M., Falk, S., and Griffith, M. (1993). Photosynthesis, photoinhibition and low temperature acclimation in cold tolerant plants. *Photosynth. Res.* 37, 19–39. doi: 10.1007/BF02185436

Hurry, V., Anderson, J. M., Chow, W. S., and Osmond, C. B. (1997). Accumulation of zeaxanthin in abscisic acid-deficient mutants of Arabidopsis does not affect chlorophyll fluorescence quenching or sensitivity to photoinhibition in vivo. *Plant Physiol.* 113, 639–648. doi: 10.1104/pp.113.2.639

Hwang, H. J., Xu, C. C., Moon, B. Y., and Lee, C. H. (2003). Recovery from low-temperature photoinhibition is related to dephosphorylation of phosphorylated CP29 rather than zeaxanthin epoxidation in rice leaves. *J. Plant Biol.* 46, 122–129. doi: 10.1007/BF03030441

Ino, Y. (1990). Field measurement of net photosynthesis of mosses at Langhovde, East Antarctica. *Ecol. Res.* 5, 195–205. doi: 10.1007/BF02346991

Ivanov, A. G., Sane, P., Hurry, V., Król, M., Sveshnikov, D., Huner, N. P., et al. (2003). Low-temperature modulation of the redox properties of the acceptor side of photosystem II: photoprotection through reaction centre quenching of excess energy. *Physiol. Plantarum* 119, 376–383. doi: 10.1034/j.1399-3054.2003.00225.x

Ivanov, A. G., Krol, M., Sveshnikov, D., Malmberg, G., Gardeström, P., Hurry, V., et al. (2006). Characterization of the photosynthetic apparatus in cortical bark chlorenchyma of Scots pine. *Planta* 223, 1165–1177. doi: 10.1007/s00425-005-0164-1

Jähne, B., Heinz, G., and Dietrich, W. (1987). Measurement of the diffusion coefficients of sparingly soluble gases in water. *J. Geophys. Res.-Oceans* 92, 10767–10776. doi: 10.1029/JC092iC10p10767

Johnson, M. P., Pérez-Bueno, M. L., Zia, A., Horton, P., and Ruban, A. V. (2009). The zeaxanthin-independent and zeaxanthin-dependent qE components of nonphotochemical quenching involve common conformational changes within the photosystem II antenna in *Arabidopsis*. *Plant Physiol.* 149, 1061–1075. doi: 10.1104/pp.108.129957

Kallio, P., and Heinonen, S. (1975). "CO2 exchange and growth of Rhacomitrium lanuginosum and Dicranum elongatum," in *Fennoscandian Tundra Ecosystems Ecological Studies (Analysis and Synthesis)*. Ed. F. E. Wiegolaski (Berlin: Springer), 138–148.

Kappen, L., Smith, R. L., and Meyer, M. (1989). Carbon dioxide exchange of two ecodemes of *Schistidium antarctici* in continental Antarctica. *Polar Biol.* 9, 415–422. doi: 10.1007/BF00443227

Kennedy, A. D. (1993). Water as a limiting factor in the Antarctic terrestrial environment: a biogeographical synthesis. *Arctic Alpine Res.* 25, 308–315. doi: 10.2307/1551914

King, D. H. (2017). "Monitoring Antarctic bryophyte communities in a time of change. [Ph D Thesis]," (Wollongong: University of Wollongong).

King, D. H., Wasley, J., Ashcroft, M. B., Ryan-Colton, E., Lucieer, A., Chisholm, L. A., and Robinson, S. A. (2020). Semi-automated analysis of digital photographs for monitoring East Antarctic vegetation. *Front. Plant Sci.* doi: 10.3389/fpls.2020.00766

Körner, C. (2003). *Alpine plant life. Functional Plant Ecology of High Mountain Ecosystems* (Berlin: Springer).

Krall, J. P., and Edwards, G. E. (1992). Relationship between photosystem II activity and CO_2 fixation in leaves. *Physiol. Plantarum* 86, 180–187. doi: 10.1111/j.1399-3054.1992.tb01328.x

Krause, G. H., and Weis, E. (1988). "The photosynthetic apparatus and chlorophyll fluorescence. An introduction," in *Applications of chlorophyll fluorescence in photosynthesis research, stress physiology, hydrobiology and remote sensing*. Ed. H. K. Lichtenthaler (Dordrecht: Springer), 3–11.

Lambrev, P. H., Tsonev, T., Velikova, V., Georgieva, K., Lambreva, M. D., Yordanov, I., et al. (2007). Trapping of the quenched conformation associated with non-photochemical quenching of chlorophyll fluorescence at low temperature. *Photosynth. Res.* 94, 321. doi: 10.1007/s11120-007-9216-7

Lee, H. Y., Hong, Y. N., and Chow, W. S. (2001). Photoinactivation of photosystem II complexes and photoprotection by non-functional neighbours in *Capsicum annuum* L. leaves. *Planta* 212, 332–342. doi: 10.1007/s004250000398

Leipner, J., Fracheboud, Y., and Stamp, P. (1997). Acclimation by suboptimal growth temperature diminishes photooxidative damage in maize leaves. *Plant Cell Environ.* 20, 366–372. doi: 10.1046/j.1365-3040.1997.d01-76.x

Lembrechts, J. J., and Lenoir, J. (2020). Microclimatic conditions anywhere at any time! *Global Change Biol.* 26, 337–339. doi: 10.1111/gcb.14942

Lembrechts, J. J., Lenoir, J., Roth, N., Hattab, T., Milbau, A., Haider, S., et al. (2019). Comparing temperature data sources for use in species distribution models: From in-situ logging to remote sensing. *Global Ecol. Biogeogr.* 28, 1578–1596. doi: 10.1111/geb.12974

Lobo, F. D. A., De Barros, M. P., Dalmagro, H. J., Dalmolin, Â. C., Pereira, W. E., De Souza, E. C., et al. (2013). Fitting net photosynthetic light-response curves with Microsoft Excel—a critical look at the models. *Photosynthetica* 51, 445–456. doi: 10.1007/s11099-013-0045-y

Longton, R. E. (1974). Microclimate and biomass in communities of the *Bryum* association on Ross Island, Continental Antarctica. *Bryologist* 77, 109–127. doi: 10.2307/3241549

Longton, R. E. (1988a). Adaptations and strategies of polar bryophytes. *Bot. J. Linn. Soc* 98, 253–268. doi: 10.1111/j.1095-8339.1988.tb02429.x

Longton, R. E. (1988b). *Biology of polar bryophytes and lichens* (Melbourne: Cambridge University Press).

Lovelock, C. E., and Robinson, S. A. (2002). Surface reflectance properties of Antarctic moss and their relationship to plant species, pigment composition and photosynthetic function. *Plant Cell Environ.* 25, 1239–1250. doi: 10.1046/j.1365-3040.2002.00916.x

Lovelock, C. E., Jackson, A. E., Melick, D. R., and Seppelt, R. D. (1995). Reversible photoinhibition in Antarctic moss during freezing and thawing. *Plant Physiol.* 109, 955–961. doi: 10.1104/pp.109.3.955

Lucieer, A., Turner, D., King, D. H., and Robinson, S. A. (2014). Using an Unmanned Aerial Vehicle (UAV) to capture micro-topography of Antarctic moss beds. *Int. J. Appl. Earth Obs.* 27, 53–62. doi: 10.1016/j.jag.2013.05.011

Matsuda, T. (1968). Ecological study of the moss community and microorganisms in the vicinity of Syowa Station, Antarctica. *JARE Sci. Rep. Ser. E Biol.* 29, 1–58.

Maxwell, K., and Johnson, G. N. (2000). Chlorophyll fluorescence—a practical guide. *J. Exp. Bot.* 51, 659–668. doi: 10.1093/jexbot/51.345.659

Melick, D. R., and Seppelt, R. D. (1992). Loss of soluble carbohydrates and changes in freezing point of Antarctic bryophytes after leaching and repeated freeze-thaw cycles. *Antarc. Sci.* 4, 399–404. doi: 10.1017/S0954102092000592

Melick, D. R., and Seppelt, R. D. (1994). Seasonal investigations of soluble carbohydrates and pigment levels in Antarctic bryophytes and lichens. *Bryologist* 1, 13–19. doi: 10.2307/3243343

Míguez, F., Fernández-Marín, B., Becerril, J. M., and García-Plazaola, J. I. (2015). Activation of photoprotective winter photoinhibition in plants from different environments: a literature compilation and meta-analysis. *Physiol. Plantarum* 155, 414–423. doi: 10.1111/ppl.12329

Nadal, M., and Flexas, J. (2018). "Mesophyll conductance to CO2 diffusion: effects of drought and opportunities for improvement," in *Water Scarcity and Sustainable Agriculture in Semiarid Environment*. Eds. I. F. García-Tejero and V. H. Durán-Zuazo (London: Elsevier), 403–438.

Nakatsubo, T. (2002). Predicting the impact of climatic warming on the carbon balance of the moss *Sanionia uncinata* on a maritime Antarctic island. *J. Plant Res.* 115, 99–106. doi: 10.1007/s102650200014

Newsham, K. K. (2010). The biology and ecology of the liverwort *Cephaloziella varians* in Antarctica. *Antarc. Sci.* 22, 131–143. doi: 10.1017/S0954102009990630

Ochyra, R., Lewis, S., and Bednarek-Ochyra, H. (2008). *The illustrated moss flora of Antarctica* (Cambridge: Cambridge University Press).

Pannewitz, S., Green, T. A., Maysek, K., Schlensog, M., Seppelt, R., Sancho, L. G., et al. (2005). Photosynthetic responses of three common mosses from continental Antarctica. *Antarc. Sci.* 17, 341–352. doi: 10.1017/S0954102005002774

Pearce, D. A. (2008). Climate change and the microbiology of the Antarctic Peninsula region. *Sci. Pro.* 91, 203–217. doi: 10.3184/003685008X332534

Peat, H. J., Clarke, A., and Convey, P. (2007). Diversity and biogeography of the Antarctic flora. *J. Biogeog.* 34, 132–146. doi: 10.1111/j.1365-2699.2006.01565.x

Pérez-Torres, E., Bravo, L. A., Corcuera, L. J., and Johnson, G. N. (2007). Is electron transport to oxygen an important mechanism in photoprotection? Contrasting responses from Antarctic vascular plants. *Physiol. Plantarum* 130, 185–194. doi: 10.1111/j.1399-3054.2007.00899.x

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D.R Core Team (2019). *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3*, 1–143.

Proctor, M. C. F. (1982). "Physiological ecology: water relations, light and temperature responses, carbon balance," in *Bryophyte ecology*. Ed. A. J. E. Smith (Dordrecht: Springer), 333–381.

Qiu, C., Ethier, G., Pepin, S., Dubé, P., Desjardins, Y., and Gosselin, A. (2017). Persistent negative temperature response of mesophyll conductance in red raspberry (*Rubus idaeus* L.) leaves under both high and low vapour pressure deficits: a role for abscisic acid? *Plant Cell Environ.* 40, 1940–1959. doi: 10.1111/pce.12997

R Core Team (2015). *R: A language and environment for statistical computing* (Vienna, Austria.: R Foundation for Statistical Computing). Available at: http://www.R-project.org/ (Accessed 24 May 2016).

Rastorfer, J. R. (1970). Effects of light intensity and temperature on photosynthesis and respiration of two East Antarctic mosses, *Bryum argenteum* and *Bryum antarcticum*. *Bryologist* 1, 544–556. doi: 10.2307/3241493

Rice, S. K., Neal, N., Mango, J., and Black, K. (2011). "Modeling bryophyte productivity across gradients of water availability using canopy form-function relationships," in *Bryophyte ecology and climate change*. Eds. Z. Tuba, N. G. Slack and L. R. Stark (Cambridge: Cambridge University Press), 441–457.

Ritchie, R. J. (2008). Fitting light saturation curves measured using modulated fluorometry. *Photosynth. Res.* 96, 201–215. doi: 10.1007/s11120-008-9300-7

Robinson, S. A., and Waterman, M. J. (2014). "Sunsafe bryophytes: photoprotection from excess and damaging solar radiation," in *Photosynthesis in bryophytes and early land plants*. Eds. D. Hanson and S. Rice (Dordrecht: Springer), 113–130.

Robinson, S. A., Wasley, J., Popp, M., and Lovelock, C. E. (2000). Desiccation tolerance of three moss species from continental Antarctica. *Funct. Plant Biol.* 27, 379–388. doi: 10.1071/PP99133

Robinson, S. A., King, D. H., Bramley-Alves, J., Waterman, M. J., Ashcroft, M. B., Wasley, J., et al. (2018). Rapid change in East Antarctic terrestrial vegetation in response to regional drying. *Nat. Clim. Change* 8, 879–884. doi: 10.1038/s41558-018-0280-0

Robinson, S. A., Klekociuk, A. R., King, D. H., Rojas, M. P., Zúñiga, G. E., and Bergstrom, D. M. (2020). The 2019/2020 summer of Antarctic heatwaves. *Glob. Change Biol.* 26, 3178–3180. doi: 10.1111/gcb.15083

Sáez, P. L., Galmés, J., Ramírez, C. F., Poblete, L., Rivera, B. K., Cavieres, L. A., et al. (2018). Mesophyll conductance to CO_2 is the most significant limitation to photosynthesis at different temperatures and water availabilities in Antarctic vascular species. *Environ. Exp. Bot.* 156, 279–287. doi: 10.1016/j.envexpbot.2018.09.008

Sáez, P. L., Rivera, B. K., Ramírez, C. F., Vallejos, V., Cavieres, L. A., Corcuera, L. J., et al. (2019). Effects of temperature and water availability on light energy utilization in photosynthetic processes of *Deschampsia antarctica*. *Physiol. Plantarum* 165, 511–523. doi: 10.1111/ppl.12739

Sage, R. F., and Kubien, D. S. (2007). The temperature response of C3 and C4 photosynthesis. *Plant Cell Environ.* 30, 1086–1106. doi: 10.1111/j.1365-3040.2007.01682.x

Sage, R. F. (2002). Variation in the k cat of Rubisco in C3 and C4 plants and some implications for photosynthetic performance at high and low temperature. *J. Exp. Bot.* 53, 609–620. doi: 10.1093/jexbot/53.369.609

Savitch, L. V., Leonardos, E. D., Krol, M., Jansson, S., Grodzinski, B., Huner, N. P. A., et al. (2002). Two different strategies for light utilization in photosynthesis in relation to growth and cold acclimation. *Plant Cell Environ.* 25, 761–771. doi: 10.1046/j.1365-3040.2002.00861.x

Savitch, L. V., Ivanov, A. G., Gudynaite-Savitch, L., Huner, N. P., and Simmonds, J. (2009). Effects of low temperature stress on excitation energy partitioning and photoprotection in *Zea mays*. *Funct. Plant Biol.* 36, 37–49. doi: 10.1071/FP08093

Scafaro, A. P., Von Caemmerer, S., Evans, J. R., and Atwell, B. J. (2011). Temperature response of mesophyll conductance in cultivated and wild *Oryza* species with contrasting mesophyll cell wall thickness. *Plant Cell Environ.* 34, 1999–2008. doi: 10.1111/j.1365-3040.2011.02398.x

Schenker, R., and Block, W. (1986). Micro-arthropod activity in three contrasting terrestrial habitats on Signy Island, maritime Antarctic. *Brit. Antarct. Surv. B.* 71, 31–43.

Sharkey, T. D., and Zhang, R. (2010). High temperature effects on electron and proton circuits of photosynthesis. *J. Integr. Plant Biol.* 52, 712–722. doi: 10.1111/j.1744-7909.2010.00975.x

Shrestha, A., Song, X., and Barbour, M. M. (2019). The temperature response of mesophyll conductance, and its component conductances, varies between species and genotypes. *Photosynth. Res.* 15, 1–18. doi: 10.1007/s11120-019-00622-z

Sinsawat, V., Leipner, J., Stamp, P., and Fracheboud, Y. (2004). Effect of heat stress on the photosynthetic apparatus in maize (*Zea mays* L.) grown at control or high temperature. *Environ. Exp. Bot.* 52, 123–129. doi: 10.1016/j.envexpbot.2004.01.010

Smith, E. L. (1936). Photosynthesis in relation to light and carbon dioxide. *P. Natl. Acad. Sci. U.S.A.* 22, 504–511. doi: 10.1073/pnas.22.8.504

Smith, A. J. E. (1982). *Bryophyte Ecology* (London: Chapman and Hall).

Smith, R. I. L. (1988). "Recording bryophyte microclimate in remote and severe environments," in *Methods in bryology*. Ed. J. M. Glime (Mainz: The Hattori Botanical Laboratory), 275–284.

Smith, R. I. L. (1996). "Terrestrial and freshwater biotic components of the western Antarctic Peninsula," in *Foundations for Ecological Research West of the Antarctic Peninsula*. Eds. R. M. Ross, E. Hoffmann II and L. B. Quetin (Washington, D.C.: American Geophysical Union), 15–59.

Smith, R. I. L. (1999). Biological and environmental characteristics of three cosmopolitan mosses dominant in continental Antarctica. *J. Veg. Sci.* 10, 231–242. doi: 10.2307/3237144

Solovchenko, A. (2010). "Screening pigments: general questions," in *Photoprotection in Plants*. Ed. A. Aolovchenko (Berlin: Springer), 9–31.

Szilágyi, A., Sommarin, M., and Åkerlund, H. E. (2007). Membrane curvature stress controls the maximal conversion of violaxanthin to zeaxanthin in the violaxanthin cycle—influence of α-tocopherol, cetylethers, linolenic acid, and temperature. *BBA-Biomembranes* 1768, 2310–2318. doi: 10.1016/j.bbamem.2007.06.001

Ubierna, N., Gandin, A., Boyd, R. A., and Cousins, A. B. (2017). Temperature response of mesophyll conductance in three C_4 species calculated with two methods: ^{18}O discrimination and *in vitro* V_{pmax}. *New Phytol.* 214, 66–80. doi: 10.1111/nph.14359

Valentini, R., Epron, D., De Angelis, P., Matteucci, G., and Dreyer, E. (1995). *In situ* estimation of net CO_2 assimilation, photosynthetic electron flow and photorespiration in Turkey oak (*Q. cerris* L.) leaves: diurnal cycles under different levels of water supply. *Plant Cell Environ.* 18, 631–640. doi: 10.1111/j.1365-3040.1995.tb00564.x

Van't, Hoff, J. H. (1884). *E'tudes de Dynamique Chimique* (Amsterdam: Frederik Muller & Co).

Venema, J. H., Villerius, L., and van Hasselt, P. R. (2000). Effect of acclimation to suboptimal temperature on chilling-induced photodamage: comparison between a domestic and a high-altitude wild *Lycopersicon* species. *Plant Sci.* 152, 153–163. doi: 10.1016/S0168-9452(99)00228-9

von Caemmerer, S., and Evans, J. R. (2015). Temperature responses of mesophyll conductance differ greatly between species. *Plant Cell Environ.* 38, 629–637. doi: 10.1111/pce.12449

Wagner, S., Zotz, G., Salazar Allen, N., and Bader, M. Y. (2013). Altitudinal changes in temperature responses of net photosynthesis and dark respiration in tropical bryophytes. *Ann. Bot.* 111, 455–465. doi: 10.1093/aob/mcs267

Walker, B., Ariza, L. S., Kaines, S., Badger, M. R., and Cousins, A. B. (2013). Temperature response of in vivo Rubisco kinetics and mesophyll conductance in *Arabidopsis thaliana*: comparisons to *Nicotiana tabacum*. *Plant Cell Environ.* 36, 2108–2119. doi: 10.1111/pce.12166

Walton, D. W. H. (1982). The Signy Island terrestrial reference sites. XV. Microclimate monitoring 1972-74. *Brit. Antarct. Surv. B.* 55, 111–126.

Wang, L. J., Loescher, W., Duan, W., Li, W. D., Yang, S. H., and Li, S. H. (2009). Heat acclimation induced acquired heat tolerance and cross adaptation in different grape cultivars: relationships to photosynthetic energy partitioning. *Funct. Plant Biol.* 36, 516–526. doi: 10.1071/FP09008

Warren, C. R., and Dreyer, E. (2006). Temperature response of photosynthesis and internal conductance to CO_2: results from two independent approaches. *J. Exp. Bot.* 57, 3057–3067. doi: 10.1093/jxb/erl067

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Software* 40, 1–29. http://www.jstatsoft.org/v40/i01/. doi: 10.18637/jss.v040.i01

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (New York: Springer-Verlag).

Wilson, J. W. (1957). Observations on the temperatures of arctic plants and their environment. *J. Ecol.* 1, 499–531. doi: 10.2307/2256933

Wilson, M. E. (1990). Morphology and photosynthetic physiology of *Grimmia antarctici* from wet and dry habitats. *Polar Biol.* 10, 337–341. doi: 10.1007/BF00237820

Xiong, D., Liu, X. I., Liu, L., Douthe, C., Li, Y., Peng, S., et al. (2015). Rapid responses of mesophyll conductance to changes of CO_2 concentration, temperature and irradiance are affected by N supplements in rice. *Plant Cell Environ.* 38, 2541–2550. doi: 10.1111/pce.12558

Xu, C. C., Jeon, Y. A., and Lee, C. H. (1999). Relative contributions of photochemical and non-photochemical routes to excitation energy dissipation in rice and barley illuminated at a chilling temperature. *Physiol. Plantarum* 107, 447–453. doi: 10.1034/j.1399-3054.1999.100411.x

Yamori, W., Noguchi, K., Hanba, Y. T., and Terashima, I. (2006). Effects of internal conductance on the temperature dependence of the photosynthetic rate in spinach leaves from contrasting growth temperatures. *Plant Cell Physiol.* 47, 1069–1080. doi: 10.1093/pcp/pcj077

Yang, Y. J., Zhang, S. B., and Huang, W. (2018). Chloroplastic ATP synthase alleviates photoinhibition of photosystem I in tobacco illuminated at chilling temperature. *Front. Plant Sci.* 9:1648:1648. doi: 10.3389/fpls.2018.01648

Yin, X., Sun, Z., Struik, P. C., and Gu, J. (2011). Evaluating a new method to estimate the rate of leaf respiration in the light by analysis of combined gas exchange and chlorophyll fluorescence measurements. *J. Exp. Bot.* 62, 3489–3499. doi: 10.1093/jxb/err038

Zotz, G., Büde, B., Meyer, A., Zellner, H., and Lange, O. L. (1997). Water relations and CO_2 exchange of tropical bryophytes in a lower montane rain forest in Panama. *Bot. Acta* 110, 9–17. doi: 10.1111/j.1438-8677.1997.tb00605.x

Zúñiga, P. E. (2016). *Respuesta fisiológica y enzimática en musgos antárticos creciendo bajo calentamiento pasivo de largo plazo expuestos a shock térmico. [Ph D Thesis]* (Concepción: Universidad de Concepción).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 19 August 2020
doi: 10.3389/fpls.2020.01274

Different Families of Retrotransposons and DNA Transposons Are Actively Transcribed and May Have Transposed Recently in *Physcomitrium* (*Physcomitrella*) *patens*

Pol Vendrell-Mir[1], *Mauricio López-Obando*[2], *Fabien Nogué*[3]* *and Josep M. Casacuberta*[1]*

[1] Centre for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG, Barcelona, Spain, [2] Department of Plant Biology, Swedish University of Agricultural Sciences, The Linnean Centre of Plant Biology in Uppsala, Uppsala, Sweden, [3] Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, Versailles, France

Similarly to other plant genomes of similar size, more than half of the genome of *P. patens* is covered by Transposable Elements (TEs). However, the composition and distribution of *P. patens* TEs is quite peculiar, with Long Terminal Repeat (LTR)-retrotransposons, which form patches of TE-rich regions interleaved with gene-rich regions, accounting for the vast majority of the TE space. We have already shown that RLG1, the most abundant TE in *P. patens*, is expressed in non-stressed protonema tissue. Here we present a non-targeted analysis of the TE expression based on RNA-Seq data and confirmed by qRT-PCR analyses that shows that, at least four LTR-RTs (RLG1, RLG2, RLC4 and tRLC5) and one DNA transposon (*PpTc2*) are expressed in *P. patens*. These TEs are expressed during development or under stresses that *P. patens* frequently faces, such as dehydratation/rehydratation stresses, suggesting that TEs have ample possibilities to transpose during *P. patens* life cycle. Indeed, an analysis of the TE polymorphisms among four different *P. patens* accessions shows that different TE families have recently transposed in this species and have generated genetic variability that may have phenotypic consequences, as a fraction of the TE polymorphisms are within or close to genes. Among the transcribed and mobile TEs, tRLC5 is particularly interesting as it concentrates in a single position per chromosome that could coincide with the centromere, and its expression is specifically induced in young sporophyte, where meiosis takes place.

Keywords: ***Physcomitrium (Physcomitrella) patens*, transposable element, transcription, genetic variability, centromere**

OPEN ACCESS

Edited by:
Meenu Kapoor,
Guru Gobind Singh Indraprastha University, India

Reviewed by:
Matej Lexa,
Masaryk University, Czechia
Rita Sharma,
Jawaharlal Nehru University, India

***Correspondence:**
Fabien Nogué
fabien.nogue@inrae.fr
Josep M. Casacuberta
josep.casacuberta@cragenomica.es

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *21 February 2020*
Accepted: *05 August 2020*
Published: *19 August 2020*

Citation:
Vendrell-Mir P, López-Obando M, Nogué F and Casacuberta JM (2020) Different Families of Retrotransposons and DNA Transposons Are Actively Transcribed and May Have Transposed Recently in Physcomitrium (Physcomitrella) patens. Front. Plant Sci. 11:1274. doi: 10.3389/fpls.2020.01274

INTRODUCTION

Mosses are one of the oldest groups of land plants, forming a sister clade with vascular plants (Leebens-Mack et al., 2019). Since the demonstration, in 1997, that gene targeting *via* homologous recombination was possible in *Physcomitrium* (*Physcomitrella*) *patens* (Schaefer and Zrÿd, 2001) this moss has become a leading plant model for answering essential questions in life sciences and in particular for understanding the evolution of biological processes of land plants. The draft of the *P. patens* genome was published in 2008 (Rensing et al., 2008), and a chromosome-scale assembly of the *P. patens* genome has been published (Lang et al., 2018), highlighting the similarities and differences with other plant genomes. Transposable Elements (TEs) account for the 57% of the 462,3 Mb of the assembled *P. patens* genome. This TE coverage is not very different from that of other plant genomes of similar size (Tenaillon et al., 2010). On the contrary, the distribution of TEs in *P. patens* is unusual as compared to other plants. TE-rich regions alternate with gene-rich regions all along the *P. patens* chromosomes (Lang et al., 2018) whereas in most plant genomes TEs accumulate in pericentromeric heterochromatic region on each chromosome. Interestingly, in spite of the general patchy TE distribution, a family of retrotransposons of the *copia* superfamily, RLC5 (comprised of full length, from now on RLC5, and truncated, tRLC5, elements), clusters at a single location in each chromosome that could correspond to the centromere (Lang et al., 2018). The TE-rich regions distributed all along the chromosomes are mainly composed of a single family of LTR-retrotransposons of the *gypsy* superfamily named RLG1 (Lang et al., 2018). RLG1 integrase contains a chromodomain, a type of protein domain that has been previously found To direct retrotransposon integration into heterochromatin (Gao et al., 2008), suggesting that RLG1 could target heterochromatic TE islands for integration. Although most TE copies are located in heterochromatic TE islands, gene-rich regions also contain some TE copies, with some of them that inserted recently and are polymorphic between the Gransden and Villersexel accessions (Lang et al., 2018). Moreover, the RLG1 retrotransposon is transcribed in *P. patens* protonema cells, suggesting that it can transpose during *P. patens* development (Vives et al., 2016; Lang et al., 2018). Although these data suggest that TE activity may have shaped the genome of *P. patens* and may continue to generate variability that potentially impact *P. patens* evolution, the global analysis of the capacity of *P. patens* TEs to be expressed and transpose is still lacking. Here we present an unbiased analysis of TE expression in *P. patens* based on RNA-Seq analyses and confirmed by qRT-PCR, that has allowed uncovering the developmentally or stress-related expression of different TE families, including class I (retrotransposons) and class II (DNA transposons) TEs. The data presented here reinforce the idea that TEs have shaped the genome of *P. patens* and show that they continue to drive its evolution.

MATERIALS AND METHODS

RNA-Seq Data Used

RNA-Seq data were obtained from the *P. patens* Gene Atlas library (Perroud et al., 2018). In particular, we used RNA-Seq data obtained from stress-treated tissues (protoplasts, ammonium treatment, de- and rehydration, heat stress, and UV-B), different developmental stages, including protonemata in BCD, BCDA or in Knopp medium, protonemata in liquid and solid medium, gametophores, leaflets, and sporophytes (green and brown stages) and some hormonal treatments (Auxin, ABA or the Jasmonic acid precursor OPDA). A complete list of the data set used can be found in **Supplementary Table 1**.

Transposable Element Transcriptome Assembly and Quantification

All selected reads where trimmed by quality using BBduk (https://sourceforge.net/projects/bbmap/). Reads mapping to the chloroplast, mitochondria or rRNA were discarded from the analysis. The remaining reads were mapped to the transposable element annotation (Hiss et al., 2017) using Bowtie2 (Langmead, 2013). All the reads that mapped were extracted using Samtools (Li et al., 2009). These reads were assembled to contigs using Trinity (Grabherr et al., 2011). In order to characterize and filter the assemblies, we aligned them to the TE library described in (Lang et al., 2018) using BLASTn (Altschul et al., 1990) with an e-value cutoff of 10^{-5}. For transcripts corresponding to class I TEs, we kept only those showing alignments longer than 1000 nt. Manual inspection allowed discarding assemblies corresponding to poorly annotated TEs (i.e. repetitive genes like Leucine-Rich Repeat genes), solo LTR or chimeric TEs. The potentially coding domains of the selected assemblies were identified by a CDD-search, which allowed defining the orientation of the potentially expressed TEs (Marchler-Bauer et al., 2015).

In order to estimate the levels of expression of the elements corresponding to the selected assemblies, RNA-Seq reads were mapped to the selected assemblies using bowtie2 and only the reads potentially corresponding to sense transcripts were kept. To quantify the expression the number of mapping reads was normalized by the length of the assembly (Kb) and the total amount of trimmed reads for each condition without aligning the reads to the genome. The normalized expression data of each transcript and the sequence of the selected transcripts can be found in **Supplementary Table 1**.

Plant Material

P. patens Gransden accession was used for all the samples used, with exception of the protonema vs sporophytes induction test where the *P. patens* Reute accession (Hiss et al., 2017) was used.

Protonemata were fragmented and plated on BCDAT medium overlaid with a cellophane disk in long-day conditions (16 h light 15 W m^{-2} to 8 h darkness) at 24°C for 7 days. Samples were collected at day 7 after 4 h of light. All the samples were frozen in liquid nitrogen immediately after harvesting and were kept at −80°C.

Protoplasts were isolated from 6 days old protonemal cultures after 30 min incubation in 1% driselase (Sigma D8037), 0.48 M mannitol. The suspension was filtered through two superposed sieves of 80 and 40 μm. Protoplasts where sedimented by low-speed centrifugation (600*g* for 5 min) and washed in 0.48 M mannitol.

The ABA treatment was performed as previously described (Perroud et al., 2018). Briefly, protonemal cultures were grown for 6 days on a cellophane disk on BCD medium. At day 6, the cellophane disks containing the protonemata tissues were transferred to BCD medium as control or to BCD containing 50 μM abscisic acid (Sigma A1049) for 24 h before harvesting.

Sporophyte RNA was obtained from Reute *P. patens*. Seven days old regenerated tissue from two consecutive rounds of a week old grinded material grown on solid BCDAT medium covered with cellophane was used as starting material. Six similar size small dots of moss tissue were plated in a 25 mm height petri dish (WVR international) containing BCD solid medium. They were grown for 40 days at 30 μmol m^{-2} s^{-1} constant white light regime and 25°C in a Sanyo MLR chamber. Then, plants were transferred to a Sanyo MLR chamber at an 8-h to 16-h light-dark cycle, 30 μmol m^{-2} s^{-1} light intensity and 15°C for reproductive gametangia induction and sporophyte development. After 20 days of post-reproductive induction (dpri), plants were submerged overnight in water to increase fertilization. Sporophyte samples were collected at 45 dpri showing a green round shape developmental stage. Each sporophyte was dissected under a Leica MZ16 stereomicroscope. Gametophyte tissue was discarded as much as possible and sporophyte was quickly frozen in liquid nitrogen. 40 dissected sporophytes were collected and used for RNA extraction.

RNA Extraction and cDNA Production

Sporophyte RNA was obtained using the QIAGEN RNeasy mini kit following manufacturer's protocol. DNA was removed by treating the samples with Ambion™ DNAseI kit (AM2222) following the manufacturer's protocol. For all other tissues, RNA extraction and DNAse treatment was done using the Maxwell® RSC Plant Kit (Promega). 500 ng of total RNA was used to synthetize the first-strand cDNA using the SuperScript™ III reverse transcriptase (Thermofisher).

qRT-PCR

Quantitative real-time PCR were done in 96-well plates using the Roche LightCycler II instrument. SYBER Green I Master Mix (Roche Applied Science), primers at 1 μm and 1/20 dilution of the cDNA obtained from the reverse transcription were used for the qRT PCR. Each sample was run per triplicate with negative reverse transcriptase and non-template controls. The amplification conditions were: 95°C for 5 min, followed by 95°C for 10 s, 56°C for 10 s, and 72°C for 10 s, ending with the melting curve to check the specificity of the qRT-PCR. The housekeeping gene adenine phosphoribosyl transferase (APT) (Schaefer et al., 2010) was used to normalize the qRT-PCR results.

The primers used to check TE expression were designed using the Primer3plus software (Untergasser et al., 2012). The list of the primers used in this study can be found in **Supplementary Table S2**.

Detection of Potentially Expressed TE Copies in the Genome and LTR-Retrotransposon Age Estimation

The TE copies most similar to the RNA assemblies, potentially representing the expressed elements, were identified by aligning the assemblies to the genome using Blastn with an e-value cutoff of $10E^{-90}$. However, in many cases the RNA assembly is obtained from the assembly of reads potentially generated by the expression of similar but different copies, and therefore, this approach may not be suitable. In order to identify the subset of elements potentially expressed in those cases, we also searched for elements showing a similarity of 80% over 80% of the sequence of the assembly. In those cases, we estimated the age of the subset of elements most similar to the assembled transcript and compared it to the age of all the complete elements of the same family annotated in the genome. To do that, we estimated first the Kimura two-parameter distance (Kimura, 1980) between the two Long Terminal Repeats (LTRs) and estimated the age using the formula $T = K/2 \times r$, where T = time of divergence, K = divergence and r = substitution rate (Bowen and McDonald, 2001). Taking into account an estimated substitution rate of 9E−09 (Rensing et al., 2007).

Transposable Element Polymorphisms Annotation

The publicly available DNA-seq resequencing data of three accessions of *P. patens* (Kaskaskia, SRX2234698; Reute, SRX1528135 and Villersexel, SRX030894) was used to look for TE polymorphisms with respect to the Gransden reference genome. Paired-end reads were mapped to the reference genome using BWA SW (Li and Durbin, 2009). TE insertions were detected using PoPoolationTE2 (Kofler et al., 2016) using the separate mode. To perform the analysis we kept only the non-reference insertions (insertions absent from the Gransden reference genome) predicted with a zygosity of at least 0.7. To establish the distance of these insertions to the closer genes the polymorphic TEs positions were intersected with that of the annotated genes using bedtools (Quinlan and Hall, 2010) using the function closestBed.

Phylogenetic Analyses

To look for sequences similar to *P. patens* TEs in other genomes we first performed a blastn search against the complete NCBI nucleotide database. As this only retrieve sequences with significant similarity to RLG1 element we complemented this search with a blastx search of the *P. patens* TEs first against the complete NCBI non-redundant protein sequence database excluding *P. patens* and subsequently, in order to increase the chance to detect plant sequences, to the NCBI green plant database (taxid:33090). We performed the tblastx with the default parameters with a maximum target sequence of 250. The most similar sequence for each species was chosen as

representative of the species. All the protein sequences were aligned using Mafft (Katoh and Standley, 2013) and trimmed using TrimAl (Capella-Gutiérrez et al., 2009). A phylogenetic tree was constructed using FastTree (Price et al., 2010) and visualized in iTOL (Letunic and Bork, 2019).

RESULTS

A New Approach to Measure the Expression of *P. patens* Transposable Elements

More than half of the *P. patens* genome (57%) is occupied by TEs, a figure that is similar to that of other genomes of similar size (Tenaillon et al., 2010). As an example, the *P. patens* TE content is similar to that of two other genomes of similar sizes and for which the TE content has been annotated using the same REPET package (Flutre et al., 2011), such as rice (46.6%) (Ou et al., 2019) and melon (45,2%) (Castanera et al., 2020). However, *P. patens* has a very different TE composition as compared with these two genomes. Indeed, class II TEs account for 21.06% of the rice genome and 15.42% of the melon genome, in *P. patens* they only represent 6% of the genome (**Figure 1**). More strikingly, a single retrotransposon family, RLG1 accounts for almost half (47.44%) of the genome space occupied by class I elements (Lang et al., 2018). RLG1 is actively expressed in non-stressed protonema cells, and it may have transposed recently during *P. patens* evolution, as some of its copies are polymorphic between *P. patens* Gransden and Villersexel ecotypes (Vives et al., 2016; Lang et al., 2018).

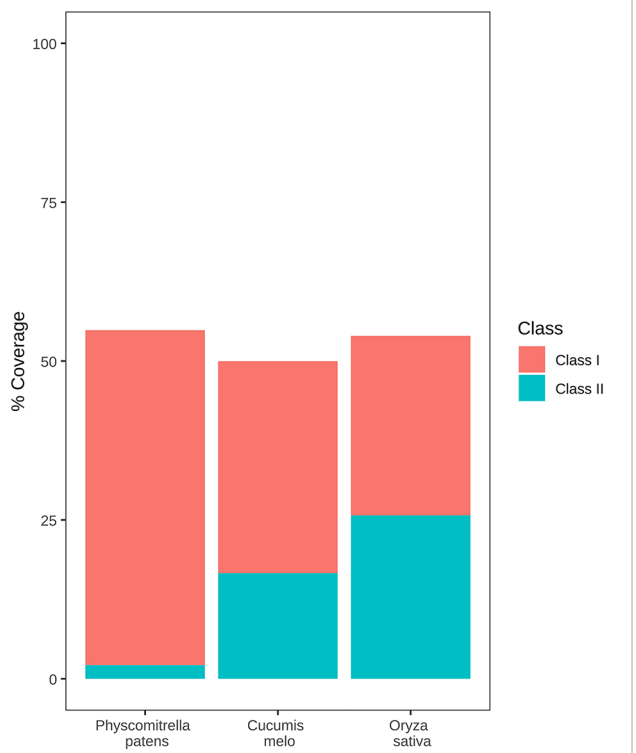


FIGURE 1 | TE content of the *P. patens*, rice, and melon genomes. Genome coverage of class 1 and class 2 TEs is shown as red and blue boxes respectively.

RLG1 copies are concentrated in TE-rich heterochromatic islands and RLG1 transposition has therefore a limited capacity to induce gene variability. In order to explore the possibility that other TE families, apart from RLG1, could be expressed in particular developmental stages or stress situations, and could therefore generate new variability in gene regions, we took advantage of the large collection of *P. patens* RNA-Seq data available from the recently published *P. patens* gene atlas (Perroud et al., 2018), which includes data from different developmental stages and stress conditions. In addition of complete TEs, eukaryote genomes, and in particular those of plants, usually contain large amounts of defective and truncated elements that may be included in transcripts that are not the result of a genuine TE expression (Anderson et al., 2019). These transcripts can be sense or antisense with respect to the TE orientation and may in some cases participate in TE regulation, but cannot be considered as productive TE transcripts potentially involved in transposition. In *P. patens*, as it is common in eukaryote genomes and in particular in plants (Hoen et al., 2015; Bennetzen and Park, 2018), the fragmented and degenerated copies of TEs outnumber the complete and potentially functional copies. As a consequence, a quantification of the level of expression based on the number of RNA-Seq reads mapping to all TE-related sequences can lead to an overestimation of the expression of the different TE families. We have therefore decided to follow a strategy based on the detection of potentially complete transcripts obtained from an assembly of RNA-Seq reads, similar to what has previously been described for the analysis of the expression of human TEs (Guffanti et al., 2018). We used Trinity RNA-seq *de novo* assembly (Grabherr et al., 2011) to assemble reads showing similarity to annotated TEs (Lang et al., 2018). The 696 assemblies obtained were blasted back to the TE annotation to classify them. The vast majority (94%) of these 696 assemblies showed similarity to LTR-RT annotations, and an important fraction of them (72%) were short (less than 1000 nt) and corresponded to fragments of LTR-RTs, such as the LTRs. As an example, the assembly TRINITY_DN331_c0_g1 showed high sequence similarity to the LTR of RLC5 elements. A search for the genomic sequence most similar to that of the assembly identified a RLC5 solo-LTR located in the downstream proximal region of the Pp3c4_32070 gene annotation (**Supplementary Figure 1**). Interestingly, an analysis of the expression data available from the *P. patens* gene atlas (Perroud et al., 2018) showed that both the RLC5 solo-LTR and the Pp3c4_32070 annotated are specifically induced in gametophores treated with ABA, which strongly suggests that this solo-LTR is expressed as a consequence of read-through transcription from the gene promoter. In order to eliminate assemblies corresponding to the expression of fragments of LTR-RTs, and taking into account that

typical complete LTR-RTs are several kb long, we discarded all the LTR-RT assemblies shorter than 1,000 nt. The remaining 172 transcripts were analyzed for the potential presence of regions coding for the typical class I and class II TE protein domains and their alignments to annotated TE sequences were manually inspected to discard those showing similarities to poorly annotated transposable elements, and truncated or chimeric elements. As an example, **Supplementary Figure 2** shows the analysis of TRINITY_DN99_c0_g1_i5 that corresponds to a complex region containing different degenerated TE fragments that seem to be transcribed as a single transcription unit. Among the 22 assemblies retained, some corresponded to the antisense strand of annotated TEs. After manual inspection, some of these were shown to correspond to LINE elements (see **Supplementary Figure 3** for an example). These transcripts may participate in the control (e.g. silencing) of TE expression but cannot be considered as genuine TE transcription. The assemblies corresponding to potential antisense transcripts were discarded. An analysis of the remaining assemblies showed that they corresponded to 9 different potentially complete annotated TEs and were selected for further analysis.

Both Retrotransposons and DNA Transposons Are Expressed in *P. patens*

The analysis of the transcript assemblies showed that they correspond to 9 different *P. patens* TEs: 2 LTR retrotransposons of the *gypsy* superfamily (RLG1 and RLG2), two of the *copia* superfamily (RLC4 and RLC5), with one of them potentially corresponding to the two different forms of RLC5, the full-length and the truncated form (RLC5/tRLC5) and two different DNA TEs belonging to the Mariner superfamily, that were not properly annotated in the *P. patens* TE annotation (Lang et al., 2018), but had been previously identified as *PpTc1* and *PpTc2* (Liu and Yang, 2014). In addition, the manual inspection of the alignments of the transcript assemblies with the annotated TEs allowed refining the annotation of two elements annotated as unclassified non-LTR retrotransposon that we could identify as a potentially expressed complete LINEs (LINE-1 and LINE-2). The RNA-Seq reads obtained from the RNAs generated by the expression of a TE family show a certain degree of sequence variability, and therefore, they are not all of them identical to the assembly that represents the complete RNA of the family. On the other hand, this assembly is in most cases not identical to any of the of the TE copies of that particular TE family. This suggests that, for most TE families, different elements are concomitantly expressed and that the RNA assembly should be considered as a consensus of the expressed RNAs.

These results suggest that different families of both retrotransposons and DNA transposons are transcribed in *P. patens*.

P. patens Contains TEs Closely Related to Fungal TEs

A preliminary characterization of the two Mariner-like elements found to be potentially expressed suggested that these elements were different from other plant Mariner-like elements, they being more closely related to fungal TEs of the Mariner superfamily. As this result was somehow surprising, we searched for sequences potentially corresponding to transposases of similar elements in the phylogenetically related liverwort *Marchantia polymorpha* and in well-characterized dicotyledonous and monocotyledonous plants such as Arabidopsis and rice. These searches did not retrieve significant hits, suggesting that these genomes do not contain sequences related to Mariner-like elements similar to those found in *P. patens*. A phylogenetic analysis of the potential transposases of Mariner-like sequences present in public databases more similar to those of the two Mariner-like elements found in *P. patens*, and including other Mariner-like sequences from plants, shows that the *P. patens* elements are closely related to elements found in fungal genomes, and are not related to *Marchantia polymorpha* or other plant sequences (**Figure 2**). These results may indicate a horizontal transfer of these TEs from fungi. In order to explore whether other TEs may have also experienced a similar phenomenon, we extended the phylogenetic analysis performed for the two Mariner-like elements to the other *P. patens* TE families here described. These analyses showed that, in contrast to what happens for the two Mariner-like elements, databases contain plant sequences with significant similarity to the rest of TE families here described. However, the phylogenetic analyses performed show that whereas the trees obtained for *P. patens* RLG2, RLC4, LINE-1 and LINE-2 retrotransposons are congruent with the phylogenetic relationships of the species, this is less obvious for RLG1 and tRLC5 (**Supplementary Figures 4–8**). This may suggest that, in addition to the two Mariner-like elements, other *P. patens* TEs may have been transferred horizontally from fungal species.

Developmental and Stress-Related Expression of *P. patens* TEs

The availability of RNA-Seq data from different developmental stages and stress conditions (Perroud et al., 2018) allowed us to perform and unbiased analysis of the patterns of expression of the different transcribed *P. patens* TEs. We have previously shown that RLG1 is expressed in non-stressed protonema cells and its expression is reduced in protonema-derived protoplasts. RLG1 seems thus to be repressed by stress, in clear contrast with the stress-related expression of most TEs, as already discussed (Vives et al., 2016; Lang et al., 2018). Here we confirm that RLG1 is expressed in protonema, its expression increasing as the protonema develops and decreasing when gametophores develop, and is repressed in protoplasts (**Figures 3** and **4**). On the other hand, RLG1 does not seem to be expressed in other tissues and it is repressed by several of the stresses analyzed, in particular by heat shock and UV-B light (**Figures 3** and **4**). We confirmed the RLG1 expression in protonema cells and its repression in protonema-derived protoplasts by qRT-PCR (**Supplementary Figure 9**). A comparison of the RLG1 assembled RNA with all the RLG1 genomic copies suggests that only a subset of the RLG1 elements is expressed (**Table 1**). An analysis of the putative ages of these elements, by analyzing the sequence differences between the two LTRs of each element,

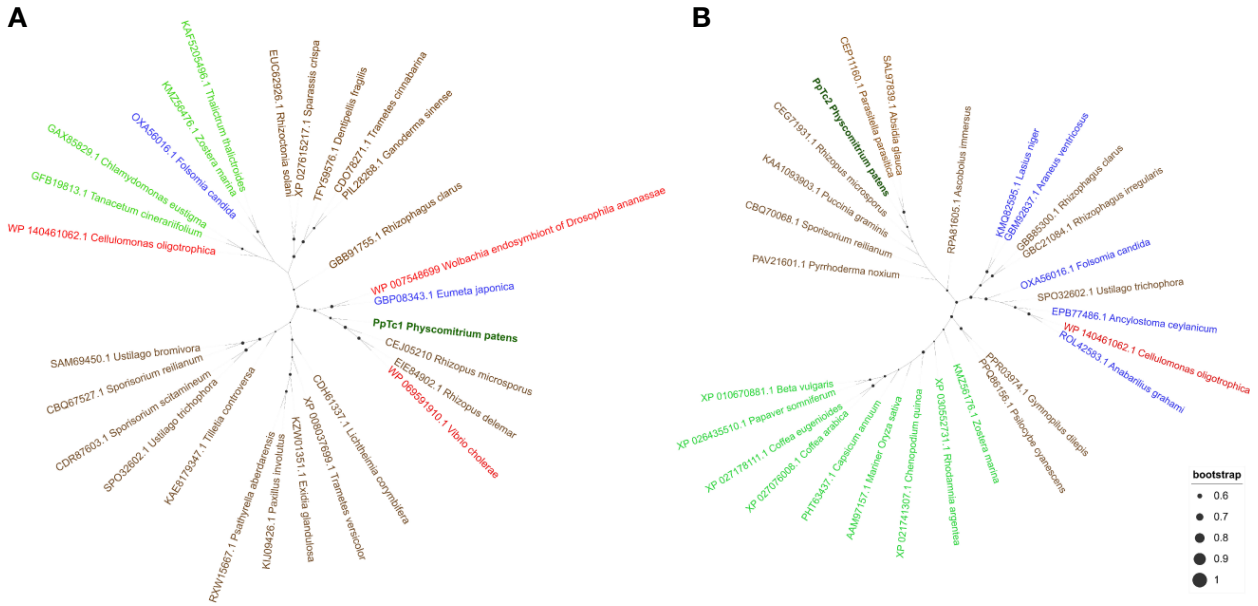


FIGURE 2 | Phylogenetic analysis of the transposases potentially encoded by *PpTc1* **(A)** and *PpTc2* **(B)** with those potentially encoded by plant, fungal, animals and bacterial Mariner-like elements. *P. patens* sequences are shown in dark green, plant sequences in light green, fungal sequences in brown, animal sequences in blue and bacterial sequences in red.

suggests that only the youngest RLG1 elements are transcribed (**Figure 5A**).

RLG1 is the TE expressed at the highest level in *P. patens* but, as already mentioned, we show here that other TEs are also expressed during *P. patens* development or under particular environmental conditions. The second Gypsy-like LTR-RT family found to be expressed, RLG2, is also expressed in protonema cells, and its expression increases in gametophores (**Figure 3**). On the other hand, the expression of RLG2 is strongly induced by ABA and heat stress in protonema, and repressed when gametophores are submitted to dehydratation and rehydratation (**Figure 4**). We confirmed the induction of RLG2 expression by ABA by qRT-PCR analyses (**Supplementary Figure 10**). Similarly, to RLG1, the comparison of the RLG2 assembled RNA with the RLG2 genomic copies shows that only the youngest RLG2 elements are transcribed in the conditions tested (**Table 1** and **Figure 5B**).

The two *copia* retrotransposon families found here to be expressed, show low levels of expression during *P. patens* development. RLC4 seems to be particularly expressed in gametophores, whereas tRLC5 seems to be more expressed in sporophytes. RLC4 expression seems to be repressed in most stress conditions, although the levels of expression are very low in all cases.

tRLC5 is a particularly interesting family of TEs, as tRLC5 copies have been proposed to mark the centromere and participate in the centromeric function (Lang et al., 2018). The data presented suggest that tRLC5 may be particularly expressed in green sporophytes (**Figure 3**). In order to confirm this pattern of expression we performed qRT-PCR experiments. As the Gransden ecotype produces few sporophytes, which makes it difficult to analyze sporophyte-specific expression, we used Reute tissues, as this ecotype produces many more sporophytes in laboratory conditions (Hiss et al., 2017). This analysis confirmed that tRLC5 expression is induced in young sporophytes (**Supplementary Figure 11**). A comparison of the tRLC5 assembled RNA with the tRLC5 genomic copies suggests that only the youngest tRLC5 elements are transcribed (**Table 1** and **Figure 5C**).

LINE-1 seems to be expressed at a very low level in all conditions and we have not detected any relevant change in expression upon stress (not shown). On the other hand, LINE-2 is also expressed at a low level in most tissues but shows an increased expression in sporophytes and germinating spores (**Figure 3**). A comparison of the LINE-2 assembled RNA with the genomic copies suggests that the expressed LINE-2 is located in the close vicinity of an annotated gene (Pp3c16_3270) and the mapping of the RNA seq reads to this region suggests that LINE-2 could be expressed as the result of a readthrough transcription of this gene (**Supplementary Figure 12**). Indeed, although there are some minor differences, the patterns of expression of Pp3c16_3270 and LINE-2 during development or under particular stress conditions are mostly coincident (not shown).

Finally, of the two Mariner-like elements analyzed, only *PpTc1* is expressed in non-stressed tissues, with a particularly high expression in gametophores and leaflets (**Figure 4**), but both *PpTc1* and *PpTc2* are strongly induced by stress. *PpTc1* expression is particularly induced by heat stress, whereas *PpTc2* is only expressed after ABA induction or after dehydration or rehydration of gametophores (**Figure 4**). A comparison of the two Mariner-like assembled RNAs with their genomic copies identified the two elements potentially transcribed. Both

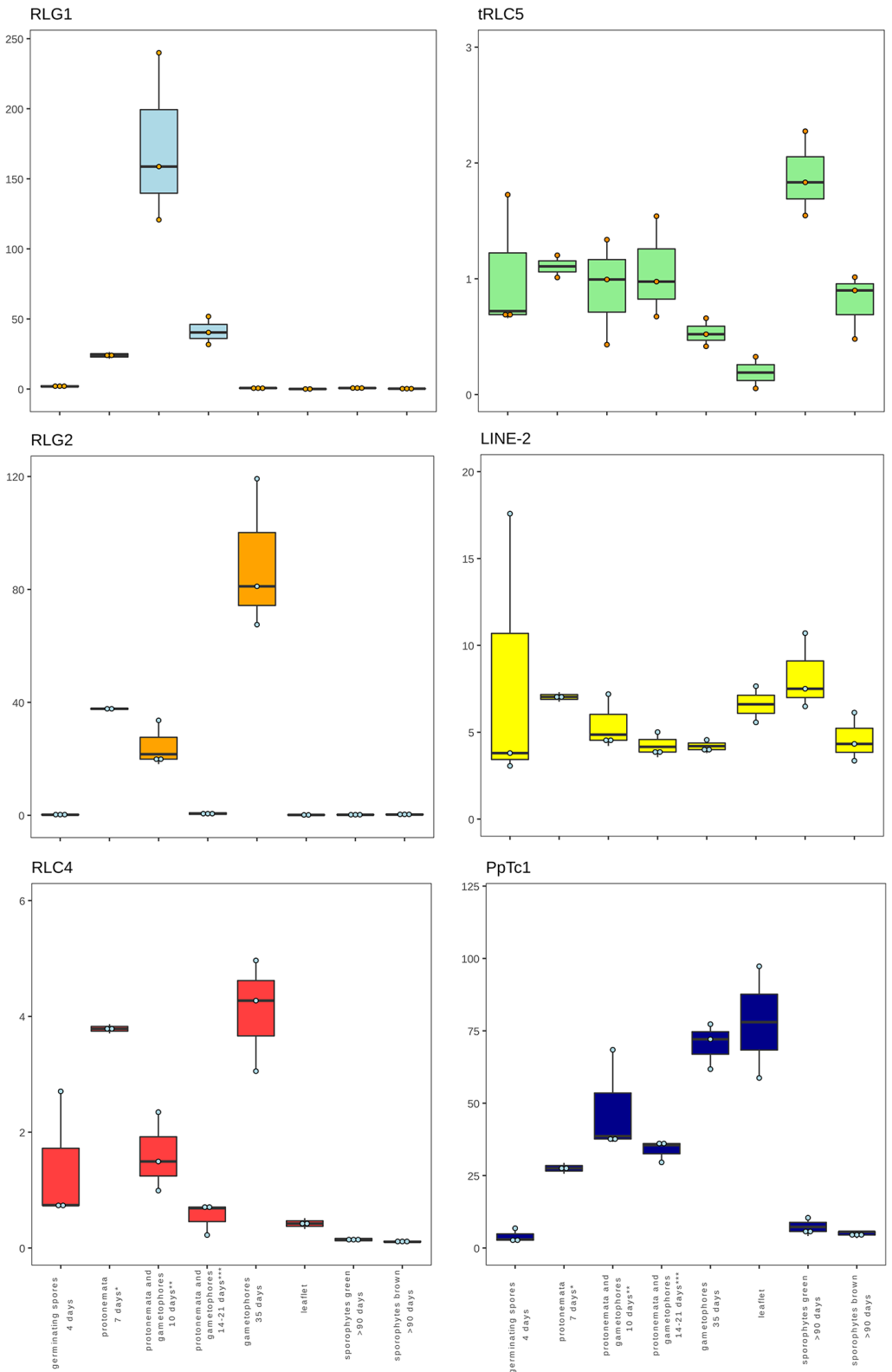


FIGURE 3 | Developmental expression of *P. patens* TEs. Normalized TE expression (see methods) in different developmental conditions selected from the *P. patens* Gene Atlas library (Perroud et al., 2018).

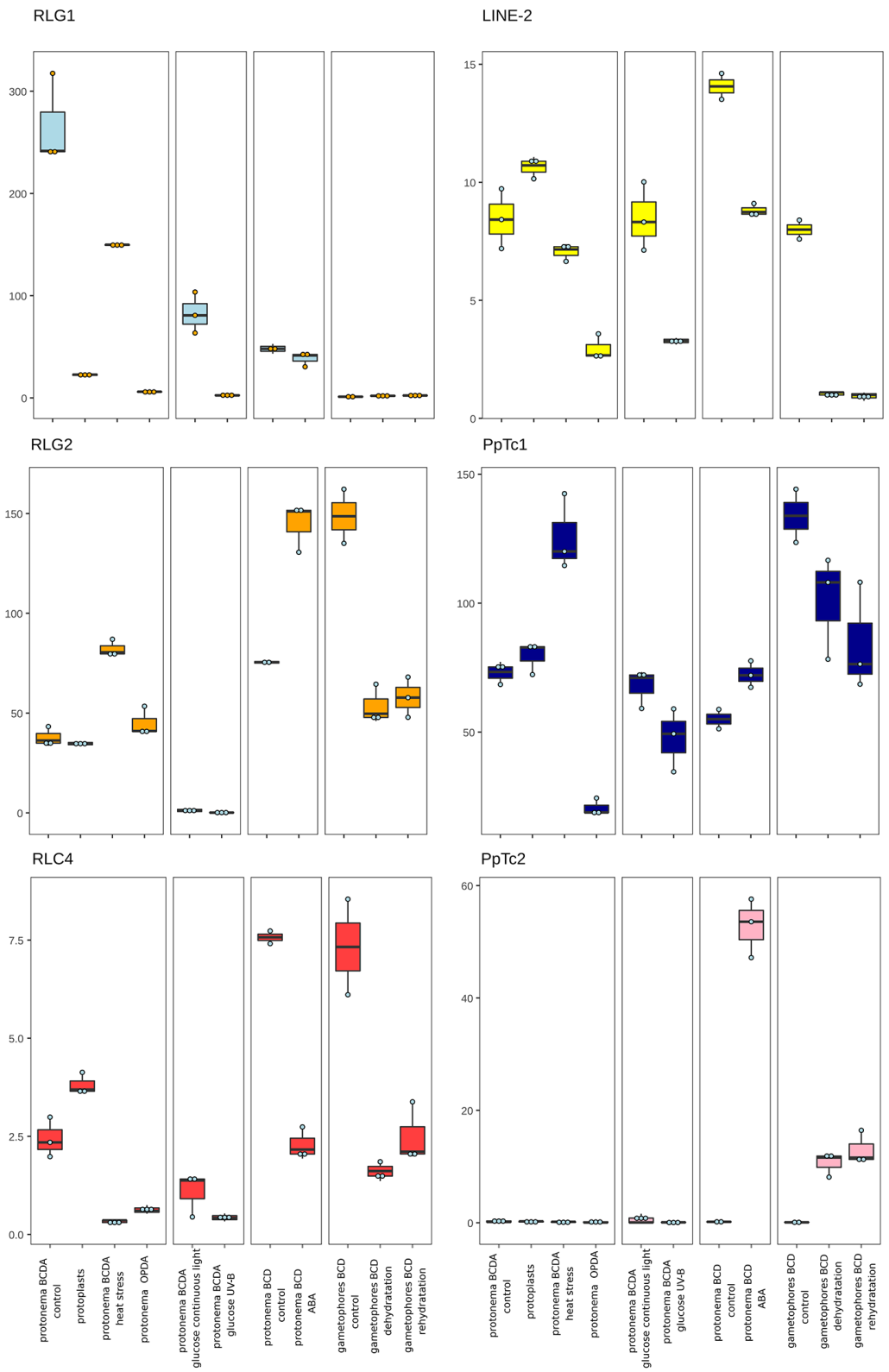


FIGURE 4 | *P. patens* TE expression under stress conditions. Normalized TE expression (see methods) under different stress conditions selected from the *P. patens* Gene Atlas library (Perroud et al., 2018).

TABLE 1 | Total number of complete elements (total), number of elements showing 80% identity over 80% of the length of the corresponding RNA assembly (80/80) for each indicated TE family.

	Total	80/80
RLG1	5092	3636
RLG2	529	25
RLC4	96	75
tRLC5	332	88
PpTc2	22	22

elements are located close to a gene, and the analysis of the patterns of expression of both genes provides information on the possible expression of the two Mariner-like elements. In the case of *PpTc1* the TE is only expressed in the conditions where the gene (Pp3c20_23510V3.1) is expressed (**Supplementary Figure 13**), which suggests that the expression detected for *PpTc1* could be the result of read-through transcription from the neighboring gene. On the contrary, the expression of *PpTc2* and the gene located nearby (Pp3c9_17220V3.1) do not overlap. Indeed, only *PpTc2*, and not the gene located nearby, is expressed in gametophores submitted to dehydration and rehydration and its expression is strongly induced in protonema treated with ABA which is not the case for the close by gene (**Supplementary Figure 14**). We confirmed the induction of *PpTc2* by ABA by qRT-PCR (**Supplementary Figure 15**). Therefore, whereas we cannot rule out the possibility that *PpTc1* expression could be the result of a readthrough expression from a neighboring gene, the transcript corresponding to *PpTc2* seems to be the result of a genuine TE transcription. Moreover, the sequence variability of the RNA-Seq reads corresponding to *PpTc2*, suggests that other *PpTc2* elements may also be expressed. Indeed, although the *PpTc2* copy located in the vicinity of the Pp3c9_17220V3.1 gene is almost identical to the RNA assembly (99.4%), other PpTc2 copies also show high similarity to the assembly (**Table 1**) and may also be expressed.

TE Mobility During Recent *P. patens* Evolution

The transcription of a copy of the TE in case of retrotransposons, and/or of the proteins necessary to mobilize the element, is the first and obligatory step of TE transposition. Therefore, the transcription of the different TEs reported here suggests that different TEs may have recently moved during *P. patens* evolution. We have already reported that this is indeed the case for RLG1, as RLG1 elements are polymorphic between the Gransden and Villersexel accessions. Here we decided to expand the analysis for possible insertion polymorphisms to all *P. patens* TEs using data from 4 different *P. patens* accessions, Reute, Kaskaskia, Villersexel, and the one from which the reference genome has been obtained, Gransden. To this end we used PopoolationTE2 to look for TE polymorphisms among these accessions using paired-end short-read resequencing data from Reute, Kaskaskia, Villersexel, that we mapped to the Gransden reference genome.

We found an important number of RLG1 polymorphisms in the three analyzed accessions with respect to Gransden (**Table 2**).

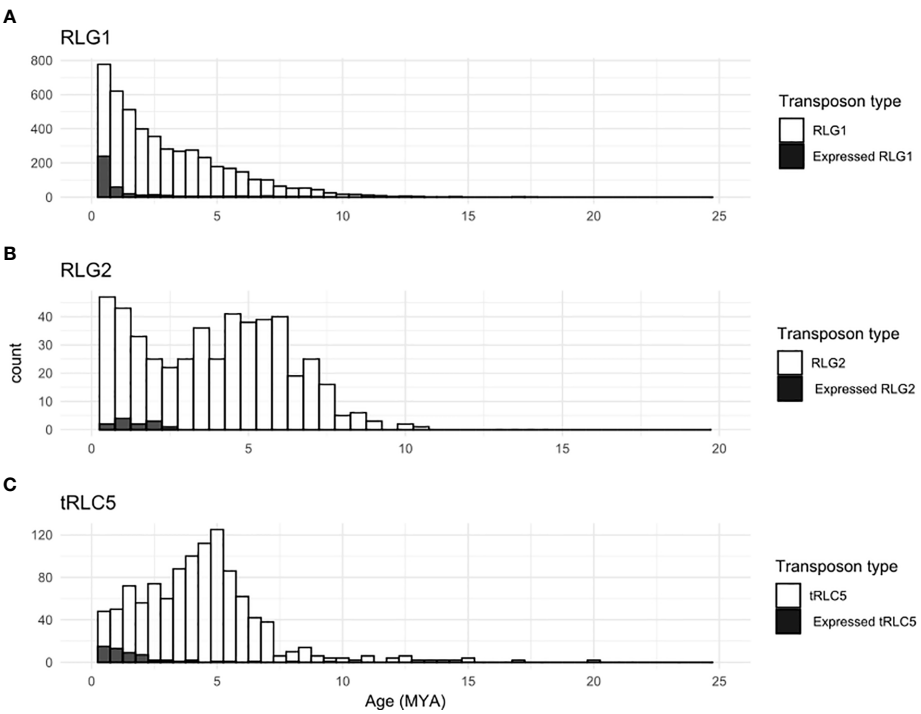


FIGURE 5 | Relative age of expressed TEs. Kimura-2-parameter distance between the two LTRs of all elements (white bars) or of elements similar to the corresponding RNA assembly, and therefore potentially expressed (black bars) belonging to the RLG1 **(A)**, RLG2 **(B)** and tRLC5 **(C)** families.

The number of polymorphisms in Reute was much smaller than in the two other accessions, which is in accordance with the close genetic relationship between Gransden and Reute (Hiss et al., 2017). Interestingly, in addition to polymorphisms related to RLG1 elements, we also detected polymorphic insertions of RLG2, RLG3, tRLC5/RLC5 and *PpTc1* (**Table 2**). In general, the number of polymorphisms is higher in Villersexel and smaller in Reute, as seen for RLG1.

The number of polymorphic insertions was particularly high for RLG3 and tRLC5/RLC5. In order to start analyzing the potential impact of the polymorphic insertions described here in the phenotypic differences between the four *P. patens* ecotypes, we analyzed the locations of the polymorphic TE insertions (**Supplementary Table 3**) and found that 20% of them are located close to genes, with potential consequences on their coding capacity or expression (**Table 3**).

DISCUSSION

The Challenging Analysis of TE Transcription

Different programs to measure TE transcription from NGS data exist (Jin et al., 2015; Lerat et al., 2017). These programs usually rely on mapping RNA-Seq reads to a TE annotation or a consensus of a TE family. Although these programs can be very useful for certain genomes and particular TE families, they may not be adequate in others. Indeed, most eukaryote genomes, and in particular those of plants, contain an important number of fragmented or degenerated TE copies in addition to full copies of TEs. As the TE fragments can also be included in transcripts, and outnumber the complete copies (Hoen et al., 2015; Bennetzen and Park, 2018), an estimation of the expression of TEs that would not discriminate between transcripts corresponding to TE fragments or to complete elements will overestimate the expression of certain families and will lead to erroneous results. This is what we came across when starting to analyze the expression of *P. patens* TEs. As an example, as already explained, among the short assemblies discarded there was one (TRINITY_DN331_c0_g1) corresponding to a RLC5 solo-LTR. An analysis of the RNA-Seq reads matching this assembly showed their specific accumulation in ABA-treated protonema cells and in gametophores under dehydration/rehydration stress. The results presented here show that the RLC5 solo-LTR is expressed as the result of read-through transcription from the ABA-induced Pp3c4_32070 gene located just upstream of it. An analysis of RLC5 expression based solely on mapping RNA-Seq reads to the TE annotation would have led to the wrong conclusion that RLC5 is induced by ABA and drought stresses. On the contrary, the approach described here, which is similar to the one previously described for the analysis of the expression of human TEs (Guffanti et al., 2018), allows for the assessment of the expression of RNAs corresponding to complete elements potentially resulting from genuine TE transcription.

Different Retrotransposon and DNA Transposon Families Are Transcribed in *P. patens*

The results presented here show that at least four LTR-RTs (RLG1, RLG2, RLC4 and tRLC5) and one DNA transposon (*PpTc2*) are expressed in *P. patens*. Among those, RLG1 and RLG2 are highly expressed during normal *P. patens* development, RLG1 being expressed mainly in protonema tissues whereas the expression of RLG2 is increased in gametophores. RLC4 seems also to be expressed in gametophores, albeit at a low level, and tRLC5 is expressed in young sporophytes. Therefore, during *P. patens* development, there is an important expression of different transposons. In addition, although RLG1 seems to be repressed by most stresses, different TEs are activated by stress. RLG2 is overexpressed under heat shock and ABA treatment, and *PpTc2* is induced by ABA and by dehydration and rehydration treatments. Mosses are known to be tolerant to dehydration and rehydration (Cuming et al., 2007; Cui et al., 2012), which, together with the associated changes of temperature, are part of their natural lifestyle. The dehydration/rehydration stresses and the ABA treatment, known to mediate the responses to those stresses (Cuming et al., 2007), and to some extent heat stress, could thus be considered as part of the normal development of *P. patens* or, at least, frequent stresses *P. patens* has to face.

Recent Mobilization of *P. patens* TEs

The expression of different TEs in normal *P. patens* growing conditions could allow the mobilization of TEs and the

TABLE 3 | Distance of polymorphic TE insertions to genes.

Accession	Inside Genes	< 1 kb closest gene	> 1 kb closest gene	Total
Reute	4	4	19	27
Kaskaskia	12	27	140	179
Villersexel	22	34	247	303
Total	38	65	406	509

TABLE 2 | TE polymorphisms in the different *P. patens* accessions.

Accession	RLG1	RLG2	RLG3	RLC4	tRLC5/RLC5	LINE-1	LINE-2	PpTc1	PpTc2	Total
Reute	18	0	4	0	5	0	0	0	0	27
Kaskaskia	147	0	15	0	17	0	0	0	0	179
Villersexel	229	2	48	2	21	0	0	1	0	303
Total	394	2	67	2	43	0	0	1	0	509

generation of genetic variability that could potentially affect gene expression/function in this haploid species. The analysis presented here shows that many TE insertions are polymorphic between different *P. patens* accessions. Indeed, we have detected an important number of polymorphic insertions of RLG1, RLG3 and tRLC5/RLC5 elements. The high number of polymorphisms related to RLG3 is intriguing as we did not detected expression. RLG3 may therefore be expressed under different environmental conditions not tested here. Alternatively, RLG3 may have lost the ability to transcribe and transpose recently during evolution. In all cases, the highest number of polymorphisms with respect to the Gransden accession is found in Villersexel and the lowest in Reute, which is in accordance with the number of SNPs these accessions show with respect to the Gransden reference genome (Lang et al., 2018). We have also found a small number of polymorphic insertions of RLG2, RLC4 and *PpTc1*. The number of detected TE polymorphisms with respect to the Gransden reference genome in these accessions is probably underestimated, as none of the programs available to look for TE polymorphisms, including the one used here, can detect polymorphic TE insertion sitting in repetitive regions (Vendrell-Mir et al., 2019). In any case, the polymorphisms detected here illustrate the potential of TEs to generate genetic variability in *P. patens*. Moreover, an important fraction of the polymorphisms detected are within or close (less than 1 Kb) to a gene, which suggests that TE movement may have impacted gene coding or gene regulation, and therefore may have contributed to the phenotypic variability of *P. patens*.

The Heterochromatic tRLC5 Elements Are Transcribed in Sporophytes

In addition to generate new alleles or new gene regulations, TEs are also involved in chromosome structure and function. In plants, TEs have been shown to provide origins of replication in heterochromatic regions (Sequeira-Mendes et al., 2019), and are frequently part of centromeres (Lermontova et al., 2015). Different retrotransposon have been found to specifically accumulate in the centromeres of the green algae *Coccomyxa subellipsoidea* (Blanc et al., 2012) or the liverwort *M. polymorpha* (Diop et al., 2020) were they could support centromere function. Interestingly, tRLC5 was previously proposed to mark the centromere and participate to the centromere function in *P. patens* (Lang et al., 2018). We show here that tRLC5 is transcribed in *P. patens*. In spite of its heterochromatic nature, centromere sequences have been shown to be transcribed in yeast, animals and plants and this transcription seems vital for the maintenance of the centromere chromatin identity and in several aspects of centromere function (Chan and Wong, 2012; Perea-Resa and Blower, 2018). Young sporophytes are a key developmental stage of *P. patens* where meiosis takes place (Charlot et al., 2014). We show here that most meiosis-specific genes (Mercier et al., 2015) are highly induced in green sporophytes (**Supplementary Figure 16**), the developmental stage where tRLC5 is expressed. It has been proposed that demethylation of centromeric DNA during meiosis may allow the transcription of centromeric sequences, which could serve as markers recognized by other factors and allow centromere assembly (Liu et al., 2015). The expression of tRLC5 in the centromere, at the moment meiosis takes place, could thus play a role in centromere assembly and function during this key process. On the other hand, the transcription pattern of tRLC5, specifically activated in young sporophytes, is reminiscent of the expression of the Athila retrotransposon of Arabidopsis, which also concentrates in the centromere and is expressed in the pollen grain (Keith Slotkin, 2010). It has been proposed that TE expression in the vegetative nurse cells of the pollen may allow re-establishing its silencing in the sperm cells (Martínez et al., 2016). The expression of tRLC5 in the sporophyte could also fulfill a similar role. Further experimental work will be required to explore any of these two non-exclusive hypotheses.

Are Some of the *P. patens* TE Families the Result of a Horizontal Transfer from Fungal Species?

In addition to the characterization of the transcriptional activity of *P. patens* TEs, the work presented also allowed us to better characterize two Mariner-like elements. These *P. patens* elements, that are transcribed and mobile, are more closely related to fungal elements than to any Mariner-like element found in plants, suggesting that they may have been horizontally transmitted from fungi. Interestingly, another *P. patens* Mariner-like element already described was also shown to be closely similar to fungal TEs (Castanera et al., 2016), which suggest that the horizontal transfer of Mariner-like elements from fungi to *P. patens* may have been a frequent event. The Mariner TE family is ubiquitous in the genomes of virtually all extant eukaryotic species and seem to be particularly prone to horizontal transfer, probably because they contain a transcriptionally promiscuous "blurry" promoter (Palazzo et al., 2019). Early land plants were aided by mutualistic interactions with fungi and these symbiotic interactions with fungi have been maintained in some bryophytes such as *M. polymorpha* (Humphreys et al., 2010). Surprisingly, although *P. patens* contains the strigolactone signaling pathway, which induce mycorrhizal signaling, it has not been shown to establish mycorrhizal interactions (Delaux et al., 2013; Field and Pressel, 2018; Rensing, 2018). The potential horizontal transfer of Mariner-like elements could be a remnant of this lost interaction, although an ulterior close contact between *P. patens* and different fungi may have also be at the origin of these horizontal transfers. It is interesting to note that *P. patens* is the only plant that shares with fungi the traces of past infections of giant virus relatives (Maumus et al., 2014), which also highlights the close relationship with fungi that *P. patens* has maintained during its recent evolution.

CONCLUSION

In summary, the results presented here show that TEs have an important activity in *P. patens*, with the transcriptional activation of different TE families in normal *P. patens* growing conditions, suggesting that TEs may have shaped *P. patens* genome and may continue to contribute to its function, including adaptation to stresses and the intraspecific genetic variability.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

JC and FN conceived the project. PV-M performed all the experiments. ML-O obtained the RNA from green sporophytes. PV-M, FN, and JC drafted the manuscript. All authors contributed to the article and approved the submitted versión.

FUNDING

This work was supported by grants from the Ministerio de Economia y Competitividad (AGL2016-78992-R) to JC and from the Investissement d'Avenir program of the French National Agency of Research for the project GENIUS (ANR-11-BTBR-0001_GENIUS) to FN. The work at CRAG is supported by a Spanish Ministry of Economy and Competitivity grant for the Center of Excellence Severo Ochoa 2016–2019 (SEV-2015-0533); the IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). PV-M holds a FPI (Formación de Personal Investigador) fellowship from the Spanish Ministerio de Economia y Competitividad.

ACKNOWLEDGMENTS

We thank all the members of JC's lab for their critical reading of the manuscript. ML-O would like to thank Eva Sundberg for their support during his post-doctoral stay at SLU. We also wish to thank Moaine Elbaidouri (U. Perpignan) for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01274/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Anderson, S. N., Stitzer, M. C., Zhou, P., Ross-Ibarra, J., Hirsch, C. D., and Springer, N. M. (2019). Dynamic patterns of transcript abundance of transposable element families in maize. *G3 Genes Genomes Genet.* 9, 3673–3682. doi: 10.1534/g3.119.400431

Bennetzen, J. L., and Park, M. (2018). Distinguishing friends, foes, and freeloaders in giant genomes. *Curr. Opin. Genet. Dev.* 49, 49–55. doi: 10.1016/j.gde.2018.02.013

Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D. D., et al. (2012). The genome of the polar eukaryotic microalga Coccomyxa subellipsoidea reveals traits of cold adaptation. *Genome Biol.* 13. doi: 10.1186/gb-2012-13-5-r39

Bowen, N. J., and McDonald, J. F. (2001). Drosophila euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res.* 11, 1527–1540. doi: 10.1101/gr.164201

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., et al. (2016). Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PloS Genet.* 12. doi: 10.1371/journal.pgen.1006108

Castanera, R., Ruggieri, V., Pujol, M., Garcia-Mas, J., and Casacuberta, J. M. (2020). An Improved Melon Reference Genome With Single-Molecule Sequencing Uncovers a Recent Burst of Transposable Elements With Potential Impact on Genes. *Front. Plant Sci.* 10, 1815. doi: 10.3389/fpls.2019.01815

Chan, L. F., and Wong, L. H. (2012). Transcription in the maintenance of centromere chromatin identity. *Nucleic Acids Res* 40, 11178–11188. doi: 10.1093/nar/gks921

Charlot, F., Chelysheva, L., Kamisugi, Y., Vrielynck, N., Guyon, A., Epert, A., et al. (2014). RAD51B plays an essential role during somatic and meiotic recombination in Physcomitrella. *Nucleic Acids Res.* 42, 11965–11978. doi: 10.1093/nar/gku890

Cui, S., Hu, J., Guo, S., Wang, J., Cheng, Y., Dang, X., et al. (2012). Proteome analysis of Physcomitrella patens exposed to progressive dehydration and rehydration. *J. Exp. Bot.* 64, 711–726. doi: 10.1093/jxb/err296

Cuming, A. C., Cho, S. H., Kamisugi, Y., Graham, H., and Quatrano, R. S. (2007). Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, Physcomitrella patens. *New Phytol.* 176, 275–287. doi: 10.1111/j.1469-8137.2007.02187.x

Delaux, P. M., Séjalon-Delmas, N., Bécard, G., and Ané, J. M. (2013). Evolution of the plant-microbe symbiotic "toolkit". *Trends Plant Sci.* 18, 298–304. doi: 10.1016/j.tplants.2013.01.008

Diop, S.II, Subotic, O., Giraldo-Fonseca, A., Waller, M., Kirbis, A., Neubauer, A., et al. (2020). A pseudomolecule-scale genome assembly of the liverwort Marchantia polymorpha. *Plant J.* 101, 1378–1396. doi: 10.1111/tpj.14602

Field, K. J., and Pressel, S. (2018). Unity in diversity: structural and functional insights into the ancient partnerships between plants and fungi. *New Phytol.* 220, 996–1011. doi: 10.1111/nph.15158

Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PloS One* 6, 1. doi: 10.1371/journal.pone.0016526

Gao, X., Hou, Y., Ebina, H., Levin, H. L., and Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res.* 18, 359–369. doi: 10.1101/gr.7146408

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Guffanti, G., Bartlett, A., Klengel, T., Klengel, C., Hunter, R., Glinsky, G., et al. (2018). Novel Bioinformatics Approach Identifies Transcriptional Profiles of Lineage-Specific Transposable Elements at Distinct Loci in the Human Dorsolateral Prefrontal Cortex. *Mol. Biol. Evol.* 35, 2435–2453. doi: 10.1093/molbev/msy143

Hiss, M., Meyberg, R., Westermann, J., Haas, F. B., Schneider, L., Schallenberg-Rüdinger, M., et al. (2017). Sexual reproduction, sporophyte development and molecular variation in the model moss Physcomitrella patens: introducing the ecotype Reute. *Plant J.* 90, 606–620. doi: 10.1111/tpj.13501

Hoen, D. R., Hickey, G., Bourque, G., Casacuberta, J., Cordaux, R., Feschotte, C., et al. (2015). A call for benchmarking transposable element annotation methods. *Mob. DNA* 6, 13. doi: 10.1186/s13100-015-0044-6

Humphreys, C. P., Franks, P. J., Rees, M., Bidartondo, M.II, Leake, J. R., and Beerling, D. J. (2010). Mutualistic mycorrhiza-like symbiosis in the most ancient group of land plants. *Nat. Commun.* 1 (1), 1–7. doi: 10.1038/ncomms1105

Jin, Y., Tam, O. H., Paniagua, E., and Hammell, M. (2015). TEtranscripts: A package for including transposable elements in differential expression analysis

of RNA-seq datasets. *Bioinformatics* 31 (22), 3593–3599. doi: 10.1093/bioinformatics/btv422

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi: 10.1093/molbev/mst010

Keith Slotkin, R. (2010). The epigenetic control of the athila family of retrotransposons in Arabidopsis. *Epigenetics* 5, 483–490. doi: 10.4161/epi.5.6.12119

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.

Kofler, R., Gómez-Sánchez, D., and Schlötterer, C. (2016). PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. *Mol. Biol. Evol.* 33, 2759–2764. doi: 10.1093/molbev/msw137

Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93. doi: 10.1111/tpj.13801

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357.

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679. doi: 10.1038/s41586-019-1693-2

Lerat, E., Fablet, M., Modolo, L., Lopez-Maestre, H., and Vieira, C. (2017). TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. *Nucleic Acids Res.* 45, 17. doi: 10.1093/nar/gkw953

Lermontova, I., Sandmann, M., Mascher, M., Schmit, A. C., and Chabouté, M. E. (2015). Centromeric chromatin and its dynamics in plants. *Plant J.* 83, 4–17. doi: 10.1111/tpj.12875

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, 256–259. doi: 10.1093/nar/gkz239

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liu, Y., and Yang, G. (2014). Tc1-like transposable elements in plant genomes. *Mob. DNA* 5. doi: 10.1186/1759-8753-5-17

Liu, Y., Su, H., Zhang, J., Liu, Y., Han, F., and Birchler, J. A. (2015). Dynamic epigenetic states of maize centromeres. *Front. Plant Sci.* 6, 904. doi: 10.3389/fpls.2015.00904

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, 222–226. doi: 10.1093/nar/gku1221

Martínez, G., Panda, K., Köhler, C., and Slotkin, R. K. (2016). Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. *Nat. Plants* 2 (4), 1–8. 16030. doi: 10.1038/nplants.2016.30

Maumus, F., Epert, A., Nogué, F., and Blanc, G. (2014). Plant genomes enclose footprints of past infections by giant virus relatives. *Nat. Commun.* 5, 4268. doi: 10.1038/ncomms5268

Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N., and Grelon, M. (2015). The Molecular Biology of Meiosis in Plants. *Annu. Rev. Plant Biol.* 66, 297–327. doi: 10.1146/annurev-arplant-050213-035923

Ou, S., Su, W., Liao, Y., Chougule, K., Ware, D., Peterson, T., et al. (2019). Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined , Comprehensive Pipeline. *Genome Biol.* 20, 275. doi: 10.1186/s13059-019-1905-y

Palazzo, A., Lorusso, P., Miskey, C., Walisko, O., Gerbino, A., Marobbio, C. M. T., et al. (2019). Transcriptionally promiscuous "blurry" promoters in Tc1/mariner transposons allow transcription in distantly related genomes. *Mob. DNA* 10:13. doi: 10.1186/s13100-019-0155-6

Perea-Resa, C., and Blower, M. D. (2018). Centromere Biology: Transcription Goes on Stage. *Mol. Cell. Biol.* 38. doi: 10.1128/mcb.00263-18

Perroud, P. F., Haas, F. B., Hiss, M., Ullrich, K. K., Alboresi, A., Amirebrahimi, M., et al. (2018). The Physcomitrella patens gene atlas project: large-scale RNA-seq based expression data. *Plant J.* 95, 168–182. doi: 10.1111/tpj.13940

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PloS One* 5. doi: 10.1371/journal.pone.0009490

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rensing, S. A., Ick, J., Fawcett, J. A., Lang, D., Zimmer, A., Van De Peer, Y., et al. (2007). An ancient genome duplication contributed to the abundance of metabolic genes in the moss Physcomitrella patens. *BMC Evol. Biol.* 7, 130. doi: 10.1186/1471-2148-7-130

Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69. doi: 10.1126/science.1150646

Rensing, S. A. (2018). Great moments in evolution: the conquest of land by plants. *Curr. Opin. Plant Biol.* 42, 49–54. doi: 10.1016/j.pbi.2018.02.006

Schaefer, D. G., and Zrÿd, J. P. (2001). The moss Physcomitrella patens, now and then. *Plant Physiol.* 127, 1430–1438. doi: 10.1104/pp.010786

Schaefer, D. G., Delacote, F., Charlot, F., Vrielynck, N., Guyon-Debast, A., Le Guin, S., et al. (2010). RAD51 loss of function abolishes gene targeting and de-represses illegitimate integration in the moss Physcomitrella patens. *DNA Repair (Amst).* 9, 526–533. doi: 10.1016/j.dnarep.2010.02.001

Sequeira-Mendes, J., Vergara, Z., Peiró, R., Morata, J., Aragüez, I., Costas, C., et al. (2019). Differences in firing efficiency, chromatin, and transcription underlie the developmental plasticity of the Arabidopsis DNA replication origins. *Genome Res.* 29, 784–797. doi: 10.1101/gr.240986.118

Tenaillon, M.II, Hollister, J. D., and Gaut, B. S. (2010). A triptych of the evolution of plant transposable elements. *Trends Plant Sci.* 15, 471–478. doi: 10.1016/j.tplants.2010.05.003

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 40, 115. doi: 10.1093/nar/gks596

Vendrell-Mir, P., Barteri, F., Merenciano, M., González, J., Casacuberta, J. M., and Castanera, R. (2019). A benchmark of transposon insertion detection tools using real data. *Mob. DNA* 10, 53. doi: 10.1186/s13100-019-0197-9

Vives, C., Charlot, F., Mhiri, C., Contreras, B., Daniel, J., Epert, A., et al. (2016). Highly efficient gene tagging in the bryophyte Physcomitrella patens using the tobacco (Nicotiana tabacum) Tnt1 retrotransposon. *New Phytol.* 212, 759–769. doi: 10.1111/nph.14152

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 August 2020
doi: 10.3389/fpls.2020.502359

Latitudinal Biogeographic Structuring in the Globally Distributed Moss *Ceratodon purpureus*

Elisabeth M. Biersma[1,2,3], Peter Convey[1], Rhys Wyber[4], Sharon A. Robinson[4], Mark Dowton[5], Bart van de Vijver[6,7], Katrin Linse[1], Howard Griffiths[2] and Jennifer A. Jackson[8]*

[1] Biodiversity, Evolution and Adaptation Team, British Antarctic Survey, Cambridge, United Kingdom, [2] Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom, [3] Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, [4] School of Earth, Atmospheric and Life Sciences, University of Wollongong, NSW, Australia, [5] School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, NSW, Australia, [6] Research Department, Botanic Garden Meise, Meise, Belgium, [7] Ecosystem Management (ECOBE), Department of Biology, University of Antwerp, Antwerp, Belgium, [8] Ecosystems Team, British Antarctic Survey, Cambridge, United Kingdom

OPEN ACCESS

Edited by:
Jeffrey Graham Duckett, Natural History Museum, United Kingdom

Reviewed by:
Alexander Zizka, German Centre for Integrative Biodiversity Research (iDiv), Germany
Santiago Martín-Bravo, Universidad Pablo de Olavide, Spain

***Correspondence:**
Elisabeth M. Biersma
elisabeth.biersma@snm.ku.dk

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *02 October 2019*
Accepted: *13 August 2020*
Published: *28 August 2020*

Citation:
Biersma EM, Convey P, Wyber R, Robinson SA, Dowton M, van de Vijver B, Linse K, Griffiths H and Jackson JA (2020) Latitudinal Biogeographic Structuring in the Globally Distributed Moss Ceratodon purpureus. Front. Plant Sci. 11:502359. doi: 10.3389/fpls.2020.502359

Biogeographic patterns of globally widespread species are expected to reflect regional structure, as well as connectivity caused by occasional long-distance dispersal. We assessed the level and drivers of population structure, connectivity, and timescales of population isolation in one of the most widespread and ruderal plants in the world — the common moss *Ceratodon purpureus*. We applied phylogenetic, population genetic, and molecular dating analyses to a global (n = 147) sampling data set, using three chloroplast loci and one nuclear locus. The plastid data revealed several distinct and geographically structured lineages, with connectivity patterns associated with worldwide, latitudinal "bands." These imply that connectivity is strongly influenced by global atmospheric circulation patterns, with dispersal and establishment beyond these latitudinal bands less common. Biogeographic patterns were less clear within the nuclear marker, with gene duplication likely hindering the detection of these. Divergence time analyses indicated that the current matrilineal population structure in *C. purpureus* has developed over the past six million years, with lineages diverging during the late Miocene, Pliocene, and Quaternary. Several colonization events in the Antarctic were apparent, as well as one old and distinct Antarctic clade, possibly isolated on the continent since the Pliocene. As *C. purpureus* is considered a model organism, the matrilineal biogeographic structure identified here provides a useful framework for future genetic and developmental studies on bryophytes. Our general findings may also be relevant to understanding global environmental influences on the biogeography of other organisms with microscopic propagules (e.g., spores) dispersed by wind.

Keywords: phylogeography, model organism, moss, spore, wind, bryophyte, global, Antarctica

INTRODUCTION

Bryophytes tend to show extensive distributions, often spanning multiple continents and/or both hemispheres (Schofield and Crum, 1972). In particular, the often extremely prolific spore production (Longton, 1997) combined with small spore size (generally ~10–20 μm diameter in mosses; Frahm, 2007) give many bryophytes strong potential for long distance dispersal. Mounting evidence from experimental (e.g. Sundberg, 2013) and phylogeographic studies (e.g. McDaniel and Shaw, 2005; Huttunen et al., 2008; Pokorny et al., 2011; Stenøien et al., 2011; Karlin et al., 2012; Piñeiro et al., 2012; Szövényi et al., 2012; Lewis et al., 2014a; Pisa et al., 2014; Kyrkjeeide et al., 2016; Biersma et al., 2017) has demonstrated the occurrence of long-distance dispersal in mosses with disjunct or widespread ranges, and suggests that the majority of bryophyte distribution patterns are underlain by relatively frequent short distance, and less frequent long distance, dispersal events (Heinrichs et al., 2009; Patiño and Vanderpoorten, 2018; Vigalondo et al., 2019).

Some bryophytes show a truly cosmopolitan or global distribution, similar to some microbial groups (Fontaneto, 2011). The capability for long-distance dispersal in such species is well illustrated by their frequent occurrence on geologically young islands (Convey et al., 2000; Vanderpoorten et al., 2007). Additionally, the speed and frequency with which some of these species colonize newly-erected buildings similarly illustrates the tremendous dispersal potential of such species (Forman, 2014). Yet, the extent to which the global distribution of cosmopolitan bryophytes predominantly reflects recent or ongoing long-distance dispersal events (e.g. over thousands of years), or a worldwide spread acquired over much longer timescales (e.g. millions of years), is poorly known.

Ceratodon purpureus (Hedw.) Brid., is one of the most widespread and ruderal moss species known, and can be found in an exceptionally wide geographic range from polar to tropical areas (Ochyra et al., 2008). It is commonly found in harsh, ruderal habitats such as concrete surfaces, buildings, roofs, sidewalks, recently burnt soil and barren glacial deposits (Jules and Shaw, 1994; Shaw and Beer, 1999), and has a high tolerance to drought, pollution, and trampling (Clarke and Robinson, 2008; Forman, 2014; Waterman et al., 2017). The species is commonly used as a model organism in genetic, physiological, and developmental studies, in particular for studying the evolution of developmental processes in bryophytes (e.g. McDaniel et al., 2007, and references therein; Szövényi et al., 2014). For this type of developmental research, good baseline knowledge on the evolutionary history and global biogeography of a species is fundamental, for instance, for underpinning interpretation of crossing experiments, trait mapping and marker discovery, and controlling for demographic or population effects.

The global genetic diversity in natural populations of *C. purpureus* was initially investigated by McDaniel and Shaw (2005) using a worldwide data set of n = 34 samples, including the chloroplast spacer *atpB-rbcL* and the nuclear genes adenosine kinase (*adk*) and phytochrome 2 (*phy2*). They found two distinct Northern Hemisphere clades, a Southern Hemisphere clade including some Northern Hemisphere specimens, and several more distantly related distinct lineages from equatorial regions, suggesting that migration between Australasian and Holarctic regions was more common than among equatorial regions. Overall, the study found that, while markers differed in implied population structure, the overall global population structure in *C. purpureus* was sparse, and provided evidence that migration is ongoing.

Here, we aim to gain a better understanding of the biogeography of *C. purpureus* and assess the level of connectivity between its globally widespread populations. Building on the previous biogeographic analysis of McDaniel and Shaw (2005), we considerably increased the global sampling (n = 147 samples; 257 newly obtained sequences), focusing primarily on three chloroplast markers, the *rps4* gene and *trnL-F* and *atpB-rbcL* spacers and, to a lesser degree, one nuclear marker, the Internal Transcribed Spacer (*ITS*). We tested for global latitudinal and longitudinal population structure, providing an overall assessment of biogeographic patterns in the species. Finally, we assessed the relative divergence time of matrilineal divergent populations, and evaluated timescales over which such widespread populations have been isolated, with a particular interest in those located in isolated, remote regions (e.g. Antarctica).

MATERIALS AND METHODS

Sampling and Molecular Methods

Moss samples for molecular analyses were obtained from herbaria (AAS, BR, BM, E, WOLL, and NY) and fresh collections (see text and **Supplementary Table S1** for sample information). We included sequences available on GenBank, and unpublished sequences from the Honours thesis of RW (Wyber, 2013), originally collected for microsatellite analyses (Clarke et al., 2008, 2009). All specimens were confirmed as *C. purpureus* by specialist bryologists. Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany), using a mortar and pestle and liquid nitrogen, following the manufacturer's instructions. As *C. purpureus* has very small gametophytes, multiple stems were extracted per sample to ensure sufficient DNA quantity. The *trnL-F* spacer, the *rps4* gene, *atpB-rbcL* spacer, and *ITS* were amplified using the primers trnLF-c and trnLF-f (Taberlet et al., 1991), trnS (Souza-Chies et al., 1997) and rps 5′ (Nadot et al., 1994), atpB1 and rbcL1 (Chiang et al., 1998) and ITS1 and ITS4 (White et al., 1990), respectively. PCR reactions were performed using the Taq PCR Core Kit (Qiagen GmbH, Hilden, Germany) with addition of $MgCl_2$ and Bovine Serum Albumin (BSA), using annealing temperatures of 60°C, 55°C, 53°C, and 55°C for *trnL-F*, *rps4*, *atpB-rbcL*, and *ITS*, respectively. PCR products were checked using gel electrophoresis. For *ITS*, many samples revealed messy or double bands, and for a small selection of samples bands were successfully excised and purified using the Wizard SV Gel and PCR clean up kit (Promega, USA), respectively. Forward and reverse sequencing was performed by LGC Genomics (Berlin, Germany) and the University of Wollongong sequencing facility.

Sequence Analysis

Forward and reverse sequences were manually checked and concatenated using CODONCODE ALIGNER v.5.0.2 (CodonCode Corp., Dedham, MA). Totals of 74, 94, 80, and 56 concatenated sequences of *trnL-F* (433–435 bp), *rps4* (589 bp), *atpB-rbcL* (623–627 bp) and *ITS* (674–816 bp) were generated, respectively. Sequences were deposited in GenBank as listed in **Supplementary Table S1**. Following McDaniel and Shaw (2005), and based on Hedderson et al. (2004), we selected sequences of *Cheilothela chloropus* (Brid.) Lindb. and three *Dicranum* species from GenBank as outgroups for phylogenetic chloroplast DNA (cpDNA) analyses. CpDNA regions were aligned using the Geneious aligner within GENEIOUS 9.0.4 (Biomatters, LTD, Auckland, NZ). *ITS* was aligned using PRANK (Löytynoja and Goldman, 2008), using default settings. Obvious misalignments were adjusted by eye. In the case of partially incomplete data, short sections at the ends of alignments were excluded. The number of variable and parsimony informative sites of all cpDNA regions were calculated using MEGA7 (Kumar et al., 2016). As McDaniel and Shaw (2005) found evidence for a possible selective sweep, we tested for positive selection in the coding gene *rps4* using the Z-test for synonymous *vs*. non-synonymous mutations, applying the Nei-Gojobori method with Jukes-Cantor correction, within MEGA7 (Kumar et al., 2016).

Three alternative alignments of *ITS* were created; (i) a full PRANK alignment, and with removal of ambiguously aligned or hypervariable regions using (ii) NOISY (Dress et al., 2008) and (iii) GBLOCKS (Castresana, 2000), using default settings. NOISY and GBLOCKS treatments resulted in alignments of length 588–678 and 514 bp, respectively. To test for recombination we applied recombination detection methods within the program RDP v4.71 (Martin et al., 2015) to the original *ITS* alignment, using default settings.

Phylogenetic and Population Genetic Analyses

For phylogenetic analyses the best-fitting models of evolution were investigated by locus, and by codon for *rps4*, with JMODELTEST-2.1.7 (Darriba et al., 2012) using the SPR tree topology search operation and AICc calculations. This provided the models TIM1+G for *atpB-rbcL* and HKY+G for *trnL-F*. The *rps4* marker was partitioned by codon position with model TPM1uf selected for the first two codon positions and JC for the third codon position.

To estimate population and species relationships, phylogenetic analyses were performed using Bayesian and Maximum likelihood (ML) methods. For individual cpDNA data sets indels were coded with simple indel coding (SIC; Simmons and Ochoterena, 2000) in SEQSTATE v1.0. We investigated four different types of data sets, including i) each cpDNA locus separately (including indel information, where present), ii) a concatenated data set including samples for which all of *rps4* and *atpB-rbcL* loci were complete, called the "*rps4* + *atpB-rbcL*" data set, and iii) a concatenated data set including only samples for which all three cpDNA loci had been sequenced, called the "concatenated cpDNA" data set. This data set was analyzed with and without indels. We also analyzed iv) a data set using all samples and cpDNA loci which had been sequenced, including those samples for which only one or two loci were available.

Bayesian analyses were performed using MRBAYES 3.2 (Ronquist et al., 2012). Runs were continued for 1.5×10^6 generations, sampling every 1.0×10^3 generations, and discarding the first 25% as burn-in. Convergence was assessed by confirming that split frequencies had an average standard deviation of < 0.01, and by using TRACER v.1.6 (Rambaut et al., 2014) to confirm that all parameters exceeded effective sample sizes (ESS) > 200. ML analyses were performed using RAxML-GUI v1.3.1 (Silvestro and Michalak, 2012; Stamatakis, 2014), using the "bootstrap + consensus" option (1000 iterations), applying models of evolution most similar to best fitting JMODELTEST-2.1.7 results in each case (e.g. GTR with or without the Gamma model of rate heterogeneity), and applying default settings. Maximum clade credibility trees were visualized using FIGTREE v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

To investigate possible species clusters within the "concatenated cpDNA" data set we used the web-based pairwise genetic distance-based Automatic Barcode Gap Discovery approach (ABGD; Puillandre et al., 2012), applying default settings. ABGD groups samples into hypothetical candidate species based on non-overlapping values of intra- and interspecific genetic distances. We investigated Prior "maximum divergence of intraspecific diversity" (P_{max}) values over a range of 0.001–0.05.

To examine phylogeographic structure, TCS networks were produced for all loci, and for the "concatenated cpDNA" data set within POPART (Leigh and Bryant, 2015) using default settings. Within *ITS*, haplotype networks were constructed using original, NOISY and GBLOCKS alignments.

Population Diversity and Demographic Analyses

To evaluate the cpDNA demographic history, we calculated genetic diversity indices, pairwise Kimura-2P distances, demographic and spatial expansion models, Tajima's D (Tajima, 1989) and Fu's Fs (Fu, 1997) neutrality tests for each cpDNA locus with 10000 permutations, using ARLEQUIN v3.5.1.2 (Excoffier and Lischer, 2010). Analyses were performed on the complete data set for each individual cpDNA marker, as well as on the total "concatenated cpDNA" data set. The former were analyzed to investigate population processes using the maximum available sample sizes; the latter using the maximum available co-segregating sites. In the "concatenated cpDNA" analyses we also considered clade substructure within *Ceratodon*, and performed demographic analyses on the overall data set (clades I–VII), and the large ABGD-inferred cluster as defined with P_{max} = 0.0017–0.0046 (clades III–VII) and P_{max} = 0.001 (clades IV–VII).

To investigate worldwide latitudinal and longitudinal geographic structure, we divided the "concatenated cpDNA" data set and the GBLOCKS filtered *ITS* data set into three

partitions based on geographical areas. The three *a priori* latitudinal partitions were based on the general atmospheric circulation cells (see Farmer and Cook, 2013) composed of i) 30°S–30°N: the area spanning the Hadley Cells between both Horse Latitudes on each side of the equator, forming one region where air is circulating (via the trade winds); ii) > 30° N; and iii) > 30°S: the Ferrel Cells and Polar Cells at the higher latitudes beyond both hemisphere's Horse Latitudes, the two other main regions where air is circulating (via the Westerlies and Polar easterlies). The *a priori* longitudinal partitions included regions between 30°W–165°W (Americas), 60°E–30° W (Europe/Africa), and 165°W–60°E (Australia/Asia). We conducted hierarchical AMOVA analysis of partitioning of genetic variation within and between regions in ARLEQUIN v3.5.1.2, and calculated F_{ST} and Φ_{ST} differentiation, using 10000 permutations. Here, F_{ST} estimates reflect differences in composition and frequency between regions, Φ_{ST} estimates reflect levels of evolutionary differentiation between different regions.

Divergence Time Estimation

To estimate the age of splits among populations we investigated divergence times within the "concatenated cpDNA" data set in BEAST v2.6.2 (Bouckaert et al., 2014). In the absence of suitable fossils we used a relaxed log normal clock with nucleotide substitution rate of 5.0×10^{-4} and standard deviation of 1.0×10^{-4} substitutions/site/my, respectively; a rate previously applied in bryophyte studies, and corresponding to the average absolute substitution rate of cpDNA across a wide range of land plants and algae (see Villarreal and Renner, 2014, and references therein). We included the same outgroups, models of evolution, and partitioning as described above. To investigate the impact of tree prior choice on divergence times in a data set that contains many population-level samples, as well as potentially different species, we explored two types of tree priors: i) a Coalescent Bayesian Skyline tree prior, and ii) a Yule tree prior. All other parameter settings were identical. MCMC chains were run for 1.0×10^8 generations, with parameters sampled every 10^3 generations. We combined log and tree files of three runs using LOGCOMBINER v2.6.2 with 10% burn-in. TRACER v.1.6 (Rambaut et al., 2014) was used to assess ESS > 200 for all estimated parameters with 10% burn-in. The maximum clade credibility tree was visualized using TREEANNOTATOR v1.8.2 (Drummond and Rambaut, 2007) and FIGTREE v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS

Molecular Sequence Data

The final sample set comprised 147 samples with a wide geographic spread across the globe (distribution shown in **Figure 1**; **Supplementary Table S1**). All cpDNA markers had low nucleotide diversity ($\pi > 0.01$) (**Supplementary Table S2**), as would be expected for these markers, which are more commonly used for species- and genus-level rather than population-level studies (Stech and Quandt, 2010). The "concatenated cpDNA" alignment comprised a minimum total combined length of 1,645 bp. Of the cpDNA markers, *atpB-rbcL* was most variable, followed by *rps4* and *trnL-F* (18, 18, and 10 variable sites, and 17, 12, and 7 parsimony informative (PI) sites, respectively). We detected no evidence for positive selection within *rps4* ($p > 0.05$), but note that low numbers of variable sites limits our ability to draw strong conclusions.

ITS had much greater genetic variation than the cpDNA markers, reflected in the numbers of variable (115) and PI sites (61), even after treatment with NOISY or GBLOCKS (83 and 41 variable sites and 60 and 20 PI sites, respectively). Many PCR-amplified samples provided a "clean" single band during

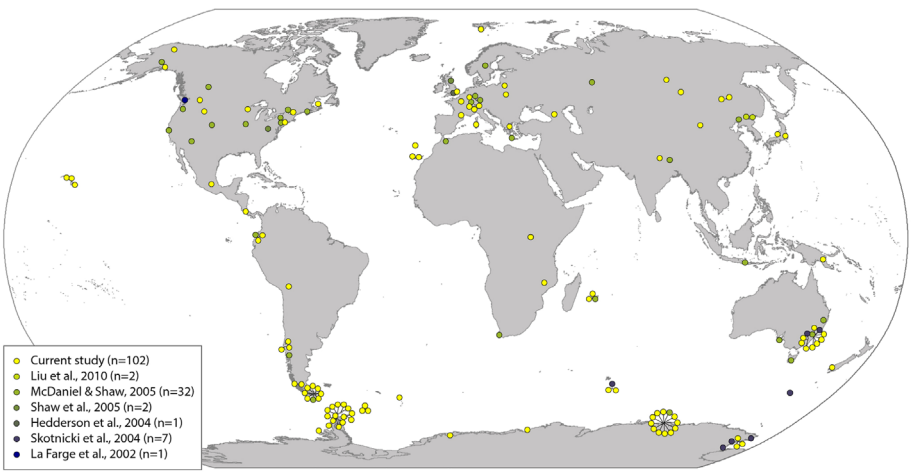


FIGURE 1 | Geographical distribution of *Ceratodon purpureus* samples. Colored dots refer to sequences generated in this study (yellow) and previously published studies (remaining colors; for more information see **Supplementary Table S1**).

electrophoresis (**Table S1**). Other samples yielded double or messy bands, and were not sequenced. Two sets of double bands were successfully excised and sequenced, and revealed multiple copies of *ITS* in the same specimens (samples from Hawaii and Australia, see Discussion for more detail). We did not investigate the occurrence of *ITS* copies further (e.g. through cloning) as *ITS* amplification of other contaminants e.g. fungi, is common in herbarium samples, and this approach would be better pursued through future studies with fresh rather than degraded herbarium material. Recombination tests revealed evidence of recombination signals in five of the seven tests performed (GENECONV, BOOTSCAN, MAXCHI, SISCAN and 3SEQ; Martin et al., 2015).

Phylogenetic and Population Genetic Analyses

Phylogenetic trees and haplotype networks of the "concatenated cpDNA" data set are presented in **Figures 2A, B**, respectively

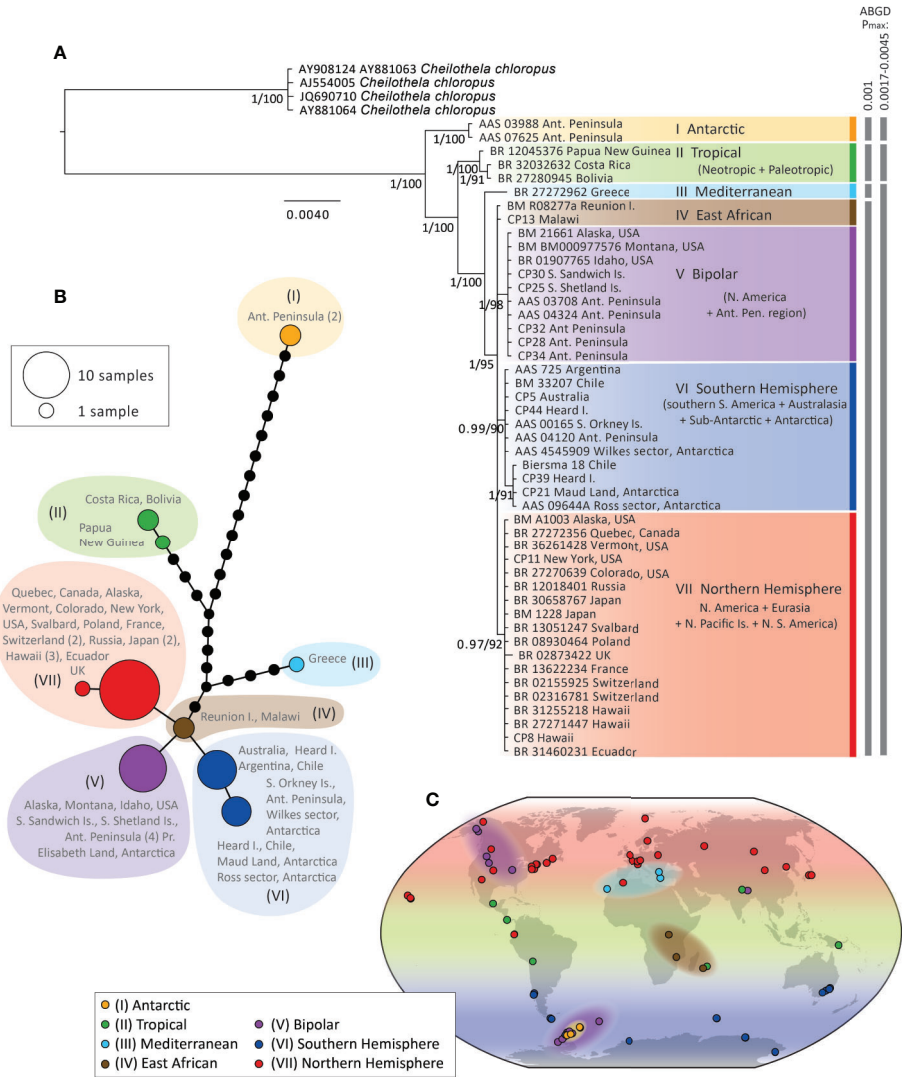


FIGURE 2 | Bayesian phylogeny **(A)** and haplotype network **(B)** for *Ceratodon purpureus* constructed with a concatenated cpDNA data set (*atpB-rbcL+rps4+trnL-F*). Posterior probabilities and bootstrap support are shown next to branches **(A)**. The scale bar represents the mean number of nucleotide substitutions per site. Biogeographic clade descriptions (I–VII) and ABGD species-clusters with different P_{max}-values are shown next to **(A)**. In **(B)** haplotype circle sizes and colors correspond to the number of specimens and clades I–VII, respectively. Branches represent mutations between haplotypes, with mutations shown as 1-step edges. **(C)** represents the sample locations and biogeographic regions of samples in the different clades (I–VII) as interpreted from the concatenated cpDNA data set **(A)**, as well as the placement of samples in phylogenies of single cpDNA markers (e.g. when samples were only represented by one or two single cpDNA marker(s); see also **Supplementary Figures S1A–D**). For example, while the Mediterranean clade (III) includes just one sample (from Greece) in the concatenated cpDNA data set **(A)**, clade III on the map in **(C)** includes one more sample from Greece (AY881059) and one from the Canary Islands (BM 27), based on the well-resolved placement of these samples in clade III in the Bayesian phylogenies of *atpB-rbcL* (**Figure S1A**) for the former, and *rps4* and *trnL-F* (**Figures S1C, D**) for the latter, respectively.

(for trees and haplotypes of individual loci and the "*rps4* + *atpB-rbcL*" data set see **Supplementary Figures S1** and **S2**). No topological conflicts were found between Bayesian and ML analyses. Phylogenetic relationships were similar for analyses of the "concatenated cDNA" data set with and without indels included (the former not shown).

Phylogenetic analyses of the "concatenated cpDNA" data set revealed seven highly supported clades (posterior probability (PP) > 0.97, bootstrap values > 90) (**Figure 2A**). These included, in order, an Antarctic clade (I), a tropical clade (II), a single-specimen lineage originating from Greece (III), and a polytomy consisting of an East African clade (IV), a bipolar clade (V), a Southern Hemisphere clade (VI) and a Northern Hemisphere clade (VII). All single-locus analyses also resolved the first three clades (I–III) (see **Supplementary Figures S1A–C**), while the latter four clades were resolved only by *atpB-rbcL* and/or *rps4*. The phylogeny with all sequenced cpDNA loci (data set iv) showed very limited resolution by comparison, with most samples falling as a single large polytomy (data not shown).

The geographic ranges spanned by each clade (I–VII) are shown by locus in **Supplementary Figures S1** and **S2**, and visualized in **Figure 2C**. Clade I only included specimens from the maritime Antarctic. Clade II included specimens from equatorial regions in the Palaeotropics and the Neotropics, including Papua New Guinea, Nepal, Réunion Island, Mexico, Costa Rica, and Bolivia. Clade III consisted of specimens from Greece and the Canary Islands. The East African group (IV) was not resolved by individual loci, but appeared ancestral within clades IV–VII (**Figures 2A, B**) and "*rps4* + *atpB-rbcL*" data set (**Supplementary Figure S1D**), and included specimens from Reunion Island, Malawi, and Uganda. The bipolar clade (V) (**Figures 2A–C**) included specimens from western North America, the sub-Antarctic, and Antarctic, as well as one specimen from India (this was only resolved in the *atpB-rbcL* marker; **Figure S1A**). The Southern Hemisphere clade (VI) included specimens from higher latitudes in the Southern Hemisphere (South America, Australia, the sub-Antarctic, and Antarctica). Finally, the Northern Hemisphere clade (VII) included specimens from the Holarctic, Hawaii, and one specimen from Ecuador.

The species delimitation method ABGD revealed three or four significant "barcoding gaps" in the concatenated cpDNA data set at P_{max} = 0.0017–0.0046 (separating I, II, and III–VII) and P_{max} = 0.001 (separating I, II, III, and IV–VII), respectively (see **Figure 2A**). Clades I and II were identified as different species clusters from the rest of the data set in all the P_{max}-values analyzed, while clade III was only regarded as a distinct species cluster at the lower P_{max}-value.

Haplotype networks of *ITS* (based on GBLOCKS, NOISY, and original alignments (**Figures 3A–C**); for more detail see **Supplementary Figure S3**) revealed that the same samples representing distinct clades I, II, and III in the cpDNA (see **Figure 2**) were also genetically distinct in *ITS*, with long branch-lengths in all *ITS* networks (**Figures 3A–C**; **Figure S3**). No distinct clustering of the remaining cpDNA-defined clades (V–VII) was apparent in the *ITS* networks. Multiple gene copies within the same sample were placed in widely separated regions in the haplotype networks (indicated with * in **Figure S3**).

Population Diversity and Demographic Analyses

F_{ST} and Φ_{ST} calculations of cpDNA revealed significant genetic differentiation between the latitudinal geographic areas (**Figure 4A**; **Supplementary Table S3**). *ITS* also revealed significant genetic differentiation in composition and frequency (F_{ST}) between all latitudinally divided regions, but only the northern (> 30°N) and southern region (> 30°S) showed significant genetic differentiation (Φ_{ST}) from one another. When divided based on longitudinal geographic regions, both cpDNA and *ITS* showed significant haplotypic differentiation (F_{ST}) between regions, but no significant evolutionary differentiation (Φ_{ST}) (**Figure 4A**; **Supplementary Table S4**).

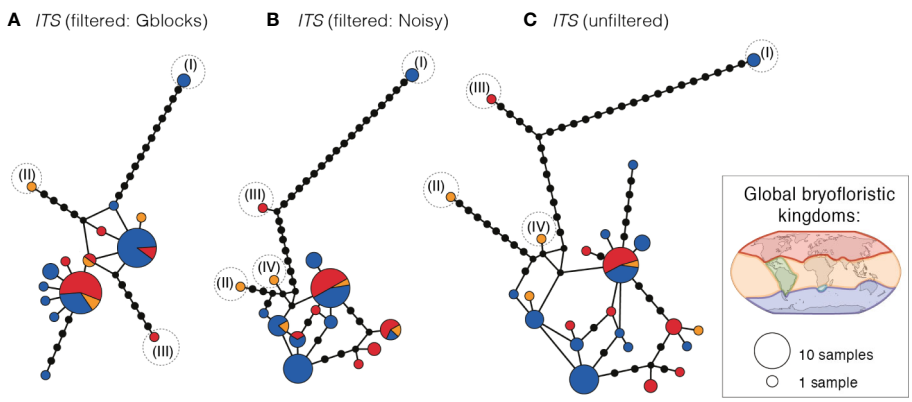


FIGURE 3 | Haplotype networks of *ITS* for *Ceratodon purpureus*, after treatment with **(A)** GBLOCKS, **(B)** NOISY or as **(C)** original data. Haplotype circle sizes and colors correspond to the number of specimens and globally recognized bryofloristic kingdoms (see legend; Schofield, 1992), respectively. Branches represent mutations between haplotypes, with mutations shown as 1-step edges. Numbers (I–IV) indicate the placement of the same samples falling in clades I–IV as resolved in the cpDNA data sets (see **Figure 2**).

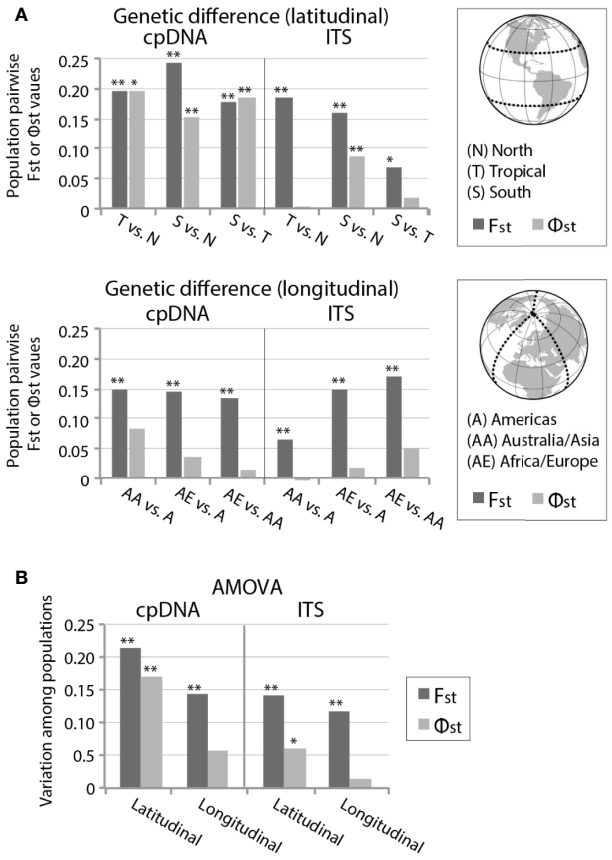


FIGURE 4 | (A) Pairwise F_{ST} and Φ_{ST} values and **(B)** analysis of molecular variance (AMOVA) within *Ceratodon purpureus* populations based on latitudinal and longitudinal geographically divided areas, including samples from the "concatenated cpDNA" and GBLOCKS filtered *ITS* data set. For details see **Tables S3–S5**. For geographically divided areas see legend in **(A)**. P-values are represented by * for $P < 0.05$ and ** for $P < 0.01$.

Latitudinal regionalized partitions of cpDNA revealed $\Phi_{ST} > F_{ST}$ in two population comparisons (tropical versus either north or south, **Figure 4A**; **Supplementary Table S4**), suggesting a significant phylogeographic signal. This was not seen in longitudinal regionalized partitions of cpDNA, or in the *ITS* data for either latitudinal or longitudinal partition.

Hierarchical AMOVA analyses of cpDNA estimated that 21% and 17% of haplotypic ($F_{ST} = 0.21$, $p < 0.01$) and genetic ($\Phi_{ST} = 0.17$, $p < 0.01$) differentiation was found within latitudinal bands, respectively, while longitudinal regionalization showed weaker haplotypic differentiation ($F_{ST} = 0.14$, $p < 0.01$) and no significant genetic differentiation ($\Phi_{ST} = 0.06$, $p = 0.07$) (**Figure 4B**; **Supplementary Table S5**). *ITS* revealed a less strong but similar pattern, with 14% and 6% of haplotypic ($F_{ST} = 0.14$, $p < 0.01$) and genetic ($\Phi_{ST} = 0.06$, $p < 0.05$) differentiation found among latitudinal geographic regions, while longitudinal regionalization showed a decreased haplotypic differentiation ($F_{ST} = 0.12$, $p < 0.01$) and no significant genetic differentiation ($\Phi_{ST} = 0.01$, $p = 0.22$) (**Figure 4B**; **Supplementary Table S5**).

The demographic expansion test on the concatenated cpDNA data set of all clades (I–VII) was significant, rejecting a demographic expansion (**Supplementary Table S2**). The spatial expansion test was non-significant, indicating that the genetic pattern may be consistent with a spatial expansion. Tajima's D was significantly negative in the full (I–VII) concatenated cpDNA data set, as well as the full data sets of *rps4* and *trnL-F*, suggesting that rare alleles were present in these loci at lower frequencies than expected, indicative of a selective sweep or past population expansion. However, Tajima's D was not significant when the most divergent lineages (I–II) were removed, suggesting the significant results were mostly influenced by cryptic population structure within *C. purpureus*. Furthermore, Fu's Fs was not significant for any data set, suggesting that overall cpDNA does not support a rapid past expansion of *C. purpureus* on a global scale.

Divergence Time Estimation

The divergence time analysis using a coalescent tree prior indicated that the ancestor of the *C. purpureus* clades originated in the mid- to late Miocene ~11.24 Mya (with 95% highest posterior density intervals (95HPD): 6.27–16.56 Mya; see **Table 1**, **Figure 5**), a time when the Antarctic (I) and tropical (II) clade diverged from the remaining clades (III–VII). The splits between clades III–VII and clades IV–VII (which were, depending on P_{max} value, delimited as one species using ABGD; see **Figure 2**) were dated to be 4.76 (95HPD: 2.00–7.78) Mya and 2.22 (95HPD: 0.76–3.93) Mya, respectively.

Implementing a Yule tree prior the analysis indicated that the ancestor of the *C. purpureus* clades originated in the late Miocene ~5.98 Mya (95HPD: 2.68–9.80 Mya; see **Table 1**, **Supplementary Figure S4**), and clade III diverged from the other clades ~4.66 Mya (95HPD: 1.95–7.85 Mya). The remaining clades (IV–VII) diverged ~3.81 Mya onwards (95HPD: 1.54–6.53 Mya). Regardless of the tree prior used, the remaining clades (IV–VII) diverged throughout the late Pliocene to mid-

TABLE 1 | Mean estimated time to most recent common ancestor (T_{MRCA}) (95% HDP lower–upper) for clades within the "concatenated cpDNA" data set (see **Figure 2A**) of *Ceratodon purpureus*, as calculated in BEAST.

Lineage T_{MRCA}	Mean ages (my) Coalescent Bayesian Skyline tree prior	Mean ages (my) Yule tree prior
Cheilothela chloropus + *Ceratodon purpureus*	62.78 (26.94–91.13)	7.18 (3.90–12.20)
I–VII	11.24 (6.27–16.56)	5.98 (2.68–9.80)
I–II	9.80 (4.86–14.85)	3.99 (0.92–7.11)
III–VII	4.76 (2.00–7.78)	4.66 (1.95–7.85)
IV–VII	2.22 (0.76–3.93)	3.81 (1.54–6.53)
I Antarctic	9.19×10^{-2} (2.21×10^{-7} to 0.36)	0.72 (1.43×10^{-6} to 2.22)
II Tropical	0.68 (2.97×10^{-3} to 2.06)	1.49 (0.07–3.45)
IV East-African	0.11 (2.47×10^{-7} to 0.45)	0.91 (2.33×10^{-6} to 2.56)
V Bipolar	0.33 (1.46×10^{-2} to 0.83)	1.99 (0.49–3.85)
VI Southern Hemisphere	0.55 (4.47×10^{-2} to 1.30)	2.23 (0.61–4.18)
VII Northern Hemisphere	0.61 (6.39×10^{-2} to 1.37)	2.55 (0.84–4.63)

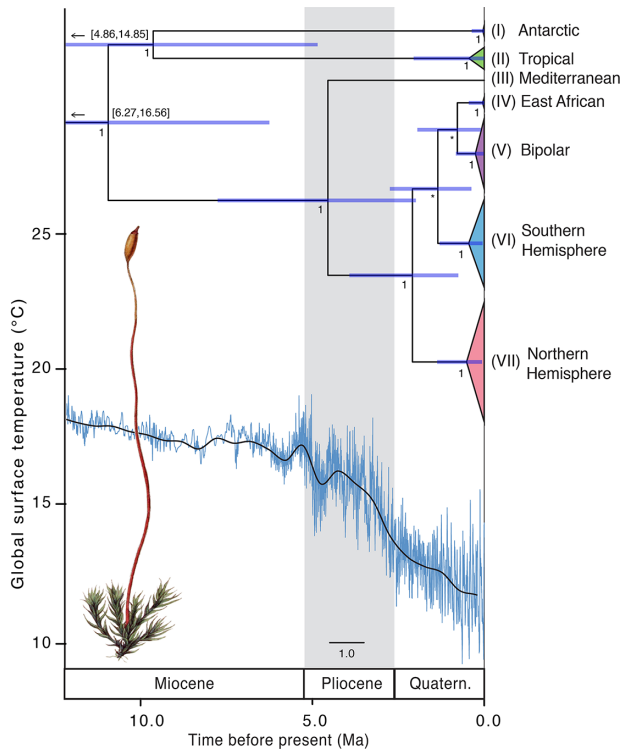


FIGURE 5 | Time-calibrated phylogeny of *Ceratodon purpureus*. The maximum clade credibility tree presents the median divergence time estimates for major lineages (**Figure 2A**) from a concatenated cpDNA data set (*atpB-rbcL+rps4+trnL-F*) using a coalescent tree prior. Node bars represent 95% height posterior distribution of age estimates. Posterior support (PP) values are shown below nodes, with PP < 0.5 provided as *. Global surface temperature estimates (blue and solid line representing temperature variations and a 500 kyr smoothed resolution, respectively), reproduced from Hansen et al. 2013, are provided below. Due to the old age (62.78 Mya; 95HDP 26.94–91.13 Mya;) the split of the outgroup *Cheilothela chloropus* and *Ceratodon purpureus* is not shown in this figure. Note that the lower part of the 95HPD range of the two oldest nodes is not shown, but given in numbers above the node bars. Scale bar indicates the number of substitutions per site. Illustration by Christiaan Sepp (Kops et al., 1868; Wikimedia Commons).

Quaternary, also a period of global cooling (Hansen et al., 2013; for a comparison with global surface temperature see **Figure 5**). All nodes had high posterior support (**Figure 5**), except for the node uniting clades IV and VI, and that separating V from IV +VI, relationships which were also not strongly supported in the phylogenetic analyses (**Figure 2A**). The phylogenies of BEAST (**Figures 5** and **S4**) and MrBayes (**Figure 2**) showed a topological inconsistency (with clade I being sister to the remainder of *Ceratodon* clades in the latter, while clades I and II were sister to each other in the former analyses). The analysis in BEAST incorporates a relaxed clock, which has a potential to improve phylogenetic accuracy but on the other hand can be less precise than the time-free approach taken in MrBayes, depending on the true underlying evolutionary pattern (Wertheim et al., 2010). Consequently, we decided to retain both topologies despite the minor biogeographic differences.

DISCUSSION

Old Lineages in Distinct Biogeographic Regions

The analysis of the cpDNA loci within *C. purpureus* revealed well-supported, several hundreds of thousands to multi-million-year old lineages derived from distinct global regions (**Figures 2** and **3**). This surprising finding implies that the global distribution of a ruderal, cosmopolitan species such as *C. purpureus* is mainly the result of a worldwide spread achieved by dispersal and establishment over hundreds of thousands to million-year timescales rather than high-frequency long-distance dispersal events, as would be expected for a highly ruderal species. The old age of the genus *Ceratodon* is also in line with previous age estimates implementing fossil ages (Laenen et al., 2014).

The matrilineal lineages were mainly strongly linked with latitude (**Figures 2C** and **4**; **Supplementary Tables S3–S5**). This global latitudinal structuring was particularly evident for cpDNA (significant F_{ST} and Φ_{ST}; **Figure 4**), and partly reflected in *ITS* (significant F_{ST}, while only one latitudinal comparison showed a significant Φ_{ST}). Biogeographic patterns were in line with those found by McDaniel and Shaw (2005), while the increased sampling and additional cpDNA loci considerably expanded the geographic extent and characterization of several clades (particularly newly-recognized clades I and VI).

All cpDNA loci (**Figures 2** and **3**) and the *ITS* marker (**Supplementary Figure S3**) revealed particularly strong differentiation of the first three clades identified (I–III). The ABGD species delimitation method supported a species complex with at least three (I, II, and III–VII) and possibly four (also dividing III from IV–VII) species, with the whole complex referred to as *C. purpureus sensu lato (s.l.)* hereafter.

Several taxonomic studies have previously also noted phenotypic differentiation within geographically separated populations of *C. purpureus* (Burley and Pritchard, 1990; Ochyra et al., 2008, and references therein), observations which regain credence based on the genetic differentiation of cpDNA and *ITS* regions within *C. purpureus s.l.*, and which are of relevance to future developmental studies using the species as a model species.

In their global revision of the *Ceratodon* genus, Burley and Pritchard (1990) identified four species and three subspecies based on a quantitative assessment of morphological characters. Although their work did not gain wide acceptance (Ochyra et al., 2008), the phylogeny and geographic regions of their described species and subspecies show a striking similarity with the genetic patterns obtained in our study. The overlap in geographic regions in the Holarctic occupied by *C. purpureus* ssp. *purpureus* with Northern Hemisphere clade VII suggests an agreement between morphological and genetic evidence. A similar resemblance was found in the occurrence of Holantarctic

C. purpureus ssp. *convolutes* (Reichardt) J.S. Burley and Southern Hemisphere clade VI. Although ssp. *convolutes* was suggested to be found in southern South America, the sub-Antarctic and Australia (Burley and Pritchard, 1990), we also find many specimens of this clade (VI) in Antarctica. The third described subspecies in Burley and Pritchard (1990), *C. purpureus* ssp. *stenocarpus* (B.S.G.) Dix., found in equatorial regions of the Old and New World, has a distribution consistent with that of tropical clade II.

Other lineages for which distinct overlap was found in both biogeography and phylogenies of work by Burley and Pritchard (1990) included: 1) the bipolar species consisting of the morphologically similar Antarctic *C. antarcticus* Card. and Arctic *C. purpureus* ssp. *arcticus* Kindb., showing an overlap with bipolar clade V; and 2) the distinct European (or Eurasian) *C. conicus* (Hamp.) Lindb., showing a possible overlap with clade III.

The morphological differences in *Ceratodon* between different geographic areas reported by Burley and Pritchard (1990) point towards a phenotypic response to local environmental conditions. It should be noted however, that we did not directly compare our samples with the exact samples used in Burley and Pritchard (1990), nor did we make detailed morphological re-assessments of our specimens, and therefore further morphological assessments are needed to clarify the status of these taxa. However, it is clear that several ancient and geographically distinct lineages within specimens morphologically assigned to *C. purpureus s.l.* exists, a pattern that might be similar in other species that are currently assumed to be globally distributed cosmopolitan species.

Concordance in Phylogeographic Patterns Between cpDNA and *ITS*

Biogeographic patterns of cpDNA were only partly mirrored by *ITS*. The latter revealed clear differentiation of the earlier diverging lineages (primarily clades I–III; **Figure 3**); however, the remaining clades could not be clearly distinguished, and *ITS* haplotypes could be found across geographically separated areas around the globe (**Figure 3**).

ITS and cpDNA revealed concordance in most of the population Φ_{ST} and F_{ST} comparisons according to latitude or longitude. Both genetic compartments revealed significant haplotypic differentiation (F_{ST}) in all latitudinal and longitudinal population comparisons. However, while cpDNA showed significant evolutionary differentiation (Φ_{ST}) in all latitudinal comparisons (implying a latitudinal biogeographic signal), this was only found in one latitudinal population comparison (> 30° N vs. > 30°S) in *ITS* (**Figure 4A**; **Supplementary Table S4**). Neither cpDNA nor *ITS* showed significant evolutionary differentiation (Φ_{ST}) in any of the longitudinal partition comparisons, implying that there is no strong longitudinal biogeographic signal.

This observed difference in geographical patterns and Φ_{ST} values in latitudinal population comparisons between cpDNA vs. nDNA could have several origins. The different nuclear pattern is unlikely to be attributed to migration of male gametes: the latter (unlike pollen in angiosperms) have an extremely limited dispersal capacity in mosses (Pressel and Duckett, 2019). The nuclear pattern is more likely to be caused by paralogy and evolution of pseudogenes. Previous research has shown *C. purpureus* has undergone an ancient nuclear genome duplication (Szövényi et al., 2014), a finding also shown in McDaniel and Shaw (2005) through the different gene copies of *adk* and *phy2*. In addition, *ITS* is known to have paralogous copies in mosses (Vanderpoorten et al., 2006). In our study we also found *ITS* had copies in several samples, and the two samples for which two different copies were sequenced were placed in widely spaced haplotypes in the haplotype network (**Figure S3**). In addition to multiple paralogous copies, the signal of the nDNA is likely influenced by the use of multiple moss shoots in the DNA extraction, and other factors such as recombination (shown in *adk* and *phy2* in McDaniel and Shaw, 2005, and *ITS* in this study), genetic interactions among nuclear loci (McDaniel et al., 2007; McDaniel et al., 2008), and possibly by processes such as hybridization among populations, selection or incomplete lineage sorting (i.e., cpDNA is haploid and therefore is subject to more rapid genetic drift than nuclear DNA, while the nuclear DNA has undergone genome duplication, shows high diversity and extensive recombination).

Previously, McDaniel and Shaw (2005) suggested the cpDNA to have undergone a selective sweep. Our results are consistent with this possibility, with current levels of diversity suggesting a selective sweep prior to the mid-Miocene, with subsequent cpDNA population structure likely reflecting patterns of wind dispersal since this time. In contrast the nuclear genome may reflect much older patterns of dispersal and connectivity over the Tertiary period.

Although our genealogical history of *ITS* is not complete (due to the difficulty of obtaining paralogous copies with degraded herbarium specimens), we highlight the importance of combining nuclear and plastid markers to unravel evolutionary histories, as both are part of a species' biology. We await future studies on the nuclear genome of *C. purpureus* (using fresh rather than herbarium samples) to advance knowledge on nDNA biogeographic patterns, as well as the extent, causes and consequences of recombination and genome doubling on the evolution in *C. purpureus* and other non-vascular plants.

Drivers of Dispersal and Establishment

We found clear biogeographic structuring dividing the main global populations of *C. purpureus s.l.* into physically distinct "latitudinal" geographic areas (see **Figures 2C** and **4**; **Supplementary Tables S3–S5**). This structuring is plausibly linked with atmospheric circulation patterns (see **Figure 6** for general global wind patterns overlaying the distributions of the Southern Hemisphere, Northern Hemisphere, tropical clades). Generally, global air masses, and thus particles that are carried within them, are retained within particular latitudinal bands or geographic regions. At higher latitudes in both hemispheres, the prevailing westerly winds (blue arrows; **Figure 6**) generally

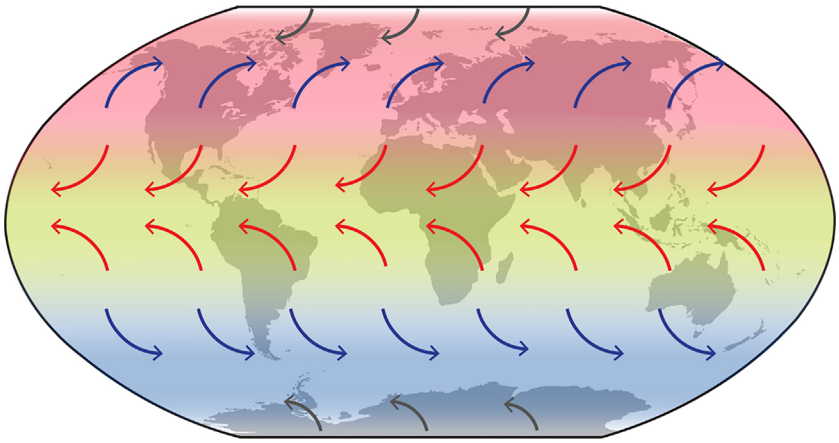


FIGURE 6 | Global wind patterns overlaying the distributions of the Southern Hemisphere (VI), Northern Hemisphere (VII), and tropical (II) clades (see **Figure 2A**). Red arrows: Trade Winds; blue arrows: Westerlies; grey arrows: Polar Easterlies.

isolate higher latitudes from equatorial regions. These could therefore be an isolating force for the Southern Hemisphere (VI) and Northern Hemisphere (VII) clades. Similarly, the easterly trade winds (red arrows; **Figure 6**) move mostly towards the equator, thus retaining entrained particles within the equatorial band. These could be an isolating force for clade II, found in tropical regions around the globe.

Similar isolating latitudinal distributions *via* wind currents have also been reported in aerobiology modeling studies (Muñoz et al., 2004; Wilkinson et al., 2012). Such atmospheric circulation patterns not only restrict the movement of particles within a given hemisphere or latitudinal band, but could, by doing so, be a driving isolating force triggering adaptation to particular climatic zones. Such genetic isolation in *C. purpureus* has been shown by McDaniel et al. (2007, 2008), who showed that considerable reproductive isolation exists between a tropical and temperate population (from Ecuador and the northern United States, respectively), including genetic incompatibilities within hybrids. If historic distribution of spore-dispersed organisms such as *Ceratodon* has been structured by atmospheric circulation, then climate change and ozone depletion, which are changing the position and strength of such wind belts (e.g. Polvani et al., 2011; Thompson et al., 2011; Turner et al., 2014; Robinson and Erickson, 2015; Robinson et al., 2020), will likely influence dispersal in the future. Increased ice melt is also predicted to increasingly open up new areas for colonization especially in polar and montane regions (e.g. Lee et al., 2017; Robinson et al., 2020).

Dispersal between hemispheres is generally limited due to atmospheric circulation patterns causing the Intertropical Convergence Zone across the equator. The distinct bipolar group (clade V; **Figure 2**) is nevertheless likely a product of at least one long-distance dispersal and establishment event across the equator. Bipolar disjunctions are a distribution pattern characteristic of many bryophytes (Ochyra et al., 2008), however recent molecular studies (e.g. Lewis et al., 2014a, 2017; Biersma et al., 2018b; Shaw et al., 2019) suggest that trans-equatorial movements are infrequent events that happen very occasionally over hundreds of thousands to multi-million year timescales. Such trans-equatorial dispersal could have been facilitated by sporadic air movements or zoochory (adventitious attachment to other organisms, e.g. birds). As nearly all clade V specimens belonged to the New World and the Antarctic region directly south of this (excepting a single specimen from India resolved to this clade in the *atpB-rbcL* marker; **Figure S1A**), migratory birds that link parts of these northern and southern regions in their annual migration are plausible vectors (cf. Lewis et al., 2014b; also a likely vector for bipolar vascular plant species, Villaverde et al., 2017). Whether the placement of the Indian sample in this bipolar clade is the result of a different long-distance dispersal event, a human induced introduction, or perhaps a matter of convergent evolution in the *atpB-rbcL* marker, requires further regional sampling and sequencing to investigate.

As *C. purpureus s.l.* is a ruderal species, characteristically found in a wide range of dry and disturbed habitats, an increase in environmental (e.g. glacial, fire-influenced and, more recently, anthropogenic) disturbances could have aided its spread across the globe. The main distribution of *C. purpureus s.l.*, occupying the vast areas of the temperate regions (particularly clades V-VII), became established throughout the late Pliocene and Quaternary. This was a period of high disturbance globally, including global cooling and repeated glacial periods (Hansen et al., 2013; see **Figure 5**). It is likely that repeated glacial disturbance provided favorable conditions for the spread and population expansion of *C. purpureus s.l.* in high latitude areas of both hemispheres. Additionally, the rise of modern flammable grass-, shrub- and

woodlands (late Miocene onwards with peak origins in late Pliocene; Bond, 2015) could have promoted its spread, as *C. purpureus s.l.* is also frequently found in fire-influenced habitats (Clément and Touffet, 1988; even called "fire moss" as a common name). Furthermore, in the recent post-quaternary period, the origin and expansion of urban environments provided major sources of anthropogenically influenced disturbance potentially favorable to *C. purpureus s.l.*

Multiple Antarctic Colonizations, Including an Ancient Lineage

C. purpureus s.l. is one of the most widespread moss species found across the Antarctic (Ochyra et al., 2008). Ochyra et al. (2008) noted considerable diversity in leaf size and morphology between Antarctic *C. purpureus s.l.* plants, suggesting the presence of multiple lineages within Antarctica, a finding confirmed here by the presence of multiple origins of Antarctic populations (clades I, V, and VI). This reveals that Antarctica is not as isolated as is often assumed for spore-dispersed organisms (also seen in the Antarctic moss *Chorisodontium aciphyllum* (Hook. f. & Wils.) Broth.; Biersma et al., 2018a). Recently, Pisa et al. (2014) also proposed at least three independent origins of the moss *Bryum argenteum* Hedw. in Antarctica, and even the vascular plant *Colobanthus quitensis* (with much larger seeds than moss spores) was also found to have at least two independent origins in the Antarctic Peninsula (Biersma et al., 2020).

According to the dating analysis, Antarctic clade I, whose members consist only of specimens from the Antarctic Peninsula and the South Orkney Islands, has been isolated since the late Miocene or early Pliocene (**Table 1**; **Figures 5** and **S4**). Although more extensive sampling may be required to fully assess whether clade I is limited to Antarctica, its apparent ancient isolation suggests it may be a remnant lineage that has survived past glaciations in the maritime Antarctic *in situ*. Its arrival could have coincided with relatively warm interglacials on the Antarctic Peninsula during the early Pliocene (particularly between ~4.5–4.4 and ~3.6–3.4 Mya; De Schepper et al., 2014). Recent climate and glaciological modeling studies have highlighted greater dynamism in glacial extent than previously considered possible throughout the early Pliocene and Pleistocene (Scherer et al., 2008; Naish et al., 2009; Pollard and DeConto, 2009; De Schepper et al., 2014; DeConto and Pollard, 2016), suggesting the possibility of ice-free local refugial areas persisting throughout these periods (e.g. as suggested by Fraser et al., 2014). Molecular, phylogenetic and biogeographic studies also suggest *in situ* survival for many groups of terrestrial fauna in Antarctica throughout the Quaternary, Neogene and even Paleogene (see Convey et al., 2008, 2009, 2020, and references therein). Recently, increased evidence has also been found of million-year persistence of the Antarctic flora, e.g. several species of *Schistidium* (Biersma et al., 2018b), and *B. argenteum* (Pisa et al., 2014). Here, our data indicate that at least one lineage (I) of *C. purpureus s.l.* may also have had a long-term Antarctic presence *in situ*.

DATA AVAILABILITY STATEMENT

Sequences are uploaded to GenBank with accessions MN542517–MN542603, MN542604–MN542650, MN552307–MN552379 and MN556618–MN556666 (see **Supplementary Table S1**). Phylogenetic matrixes (MrBayes and BEAST input files), tree files and POPART input files are uploaded in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

PC, EB, RW, and SR conceived the study, with details further developed by JJ, MD, KL, and HG. SR, MD, and RW conceived and executed a small pilot study. EB, BV, and RW conducted the majority of herbarium sampling. EB and RW carried out the molecular work. EB, with guidance from JJ, conducted the analyses and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by a Natural Environment Research Council (NERC) PhD studentship (ref NE/K50094X/1), NERC-CONICYT (NE/P003079/1) and Carlsberg Foundation (CF18-0267) to EB and NERC core funding to the BAS Biodiversity, Evolution, and Adaptation Team, and Australian Research Council (ARC) Discovery Project DP180100113 funding to SR and PC.

ACKNOWLEDGMENTS

We are grateful for comments of Stuart McDaniel, Matthew G. Johnson and Alain Vanderpoorten, which greatly improved earlier versions of this article. We thank Johanna Turnbull, Laurence Clarke, Alison Downing, Marc Lebouvier, Ryszard Ochyra, Stuart McDaniel, Helena Korpeläinen, Angelica Casanova-Katny, Graham Bell and curators of herbaria BR, BM, AAS, WOLL, E, NY for providing specimens or DNA. We are grateful for the support of the French Polar Institute (program 136), the Australian Antarctic Division and Instituto Antartico Chileno for support to collect specimens. We thank Laura Gerrish (British Antarctic Survey) for preparing **Figures 1** and **2C**. This research contributes to the Scientific Committee on Antarctic Research "State of the Antarctic Ecosystem" program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.502359/full#supplementary-material

REFERENCES

Biersma, E. M., Jackson, J. A., Hyvönen, J., Koskinen, S., Linse, K., Griffiths, H., et al. (2017). Global biogeographic patterns in bipolar moss species. *R. Soc Open Sci.* 4:170147. doi: 10.1098/rsos.170147

Biersma, E. M., Jackson, J. A., Bracegirdle, T. J., Griffiths, H., Linse, K., and Convey, P. (2018a). Low genetic variation between South American and Antarctic populations of the bank–forming moss *Chorisodontium aciphyllum* (Dicranaceae). *Polar. Biol.* 41, 599–610. doi: 10.1007/s00300-017-2221-1

Biersma, E. M., Jackson, J. A., Stech, M., Griffiths, H., Linse, K., and Convey, P. (2018b). Molecular data suggest long-term *in situ* Antarctic persistence within Antarctica's most speciose plant genus, *Schistidium*. *Front. Ecol. Evol.* 6:77. doi: 10.3389/fevo.2018.00077

Biersma, E. M., Torres-Díaz, C., Molina-Montenegro, M. A., Newsham, K. K., Vidal, M. A., Collado, G. A., et al. (2020). Multiple post-glacial colonisation events of the Antarctic pearlwort *Colobanthus quitensis* (Caryophyllaceae) reveal the recent arrival of native Antarctic vascular flora. *J. Biogeogr.* 00, 1–11. doi: 10.1111/jbi.13843

Bond, W. J. (2015). Fires in the Cenozoic: a late flowering of flammable ecosystems. *Front. Plant Sci.* 5:749. doi: 10.3389/fpls.2014.00749

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, e1003537. doi: 10.1371/journal.pcbi.1003537

Burley, J. S., and Pritchard, N. M. (1990). Revision of the genus *Ceratodon* (Bryophyta). *Harv. Pap. Bot.* 2, 17–76.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chiang, T.-Y., Schaal, B. A., and Peng, C.-I. (1998). Universal primers for amplification and sequencing a noncoding spacer between the atpB and rbcL genes of chloroplast DNA. *Bot. Bull. Acad. Sinica.* 39, 245–250.

Clarke, L. J., Ayre, D. J., and Robinson, S. A. (2008). Somatic mutation and the Antarctic ozone hole. *J. Ecol.* 96, 378–385. doi: 10.1111/j.1365-2745.2007.01347.x

Clarke, L. J., and Robinson, S. A. (2008). Cell wall-bound ultraviolet-screening compounds explain the high ultraviolet tolerance of the Antarctic moss, Ceratodon purpureus. *New Phytologist* 179 (3), 776–783. doi: 10.1111/j.1469-8137.2008.02499.x

Clarke, L. J., Robinson, S. A., and Ayre, D. J. (2009). Genetic structure of Antarctic populations of the moss *Ceratodon purpureus*. *Antarct. Sci.* 21, 51–58. doi: 10.1017/S0954102008001466

Clément, B., and Touffet, J. (1988). Le rôle des bryophytes dans la recolonisation des landes après incendie. *Cryptogamie Bryol. L.* 9 (4), 297–311.

Convey, P., Smith, R.II, Hodgson, D. A., and Peat, H. J. (2000). The flora of the South Sandwich Islands, with particular reference to the influence of geothermal heating. *J. Biogeogr.* 27, 1279–1295. doi: 10.1046/j.1365-2699.2000.00512.x

Convey, P., Gibson, J. A., Hillenbrand, C. D., Hodgson, D. A., Pugh, P. J., Smellie, J. L., et al. (2008). Antarctic terrestrial life - challenging the history of the frozen continent? *Biol. Rev. Camb. Philos. Soc* 83, 103–117. doi: 10.1111/j.1469-185X.2008.00034.x

Convey, P., Bindschadler, R., Di Prisco, G., Fahrbach, E., Gutt, J., Hodgson, D. A., et al. (2009). Antarctic climate change and the environment. *Antarct. Sci.* 21, 541–563. doi: 10.1017/S0954102009990642

Convey, P., Biersma, E. M., Casanova-Katny, A., and Maturana, C. S. (2020). "Refuges of Antarctic diversity," in *Past Antarctica*. Eds. M. Oliva and J. Ruiz-Fernández (Cambridge, MA: Academic Press), 181–200. doi: 10.1016/B978-0-12-817925-3.00010-0

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772. doi: 10.1038/nmeth.2109

De Schepper, S., Gibbard, P. L., Salzmann, U., and Ehlers, J. (2014). A global synthesis of the marine and terrestrial evidence for glaciation during the Pliocene Epoch. *Earth Sci. Rev.* 135, 83–102. doi: 10.1016/j.earscirev.2014.04.003

DeConto, R. M., and Pollard, D. (2016). Contribution of Antarctica to past and future sea-level rise. *Nature* 531, 591–597. doi: 10.1038/nature17145

Dress, A. W., Flamm, C., Fritzsch, G., Grunewald, S., Kruspe, M., Prohaska, S. J., et al. (2008). Noisy: identification of problematic columns in multiple sequence alignments. *Algorithms Mol. Biol.* 3:7. doi: 10.1186/1748-7188-3-7

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Excoffier, L., and Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Farmer, G. T., and Cook, J. (2013). "Atmospheric Circulation and Climate," in *Climate Change Science: A Modern Synthesis* (Dordrecht: Springer), 231–243. doi: 10.1007/978-94-007-5757-8_11

Fontaneto, D. (Ed.) (2011). *Biogeography of Microscopic Organisms: Is Everything Small Everywhere? (Systematics Association Special Volume Series)* (Cambridge: Cambridge University Press). doi: 10.1017/CBO9780511974878

Forman, R. T. T. (2014). "Urban habitat, vegetation, plants," in *Urban Ecology: Science of Cities*. Ed. R. T. T. Forman (Cambridge: Cambridge University Press), 205–240. doi: 10.1017/CBO9781139030472.010

Frahm, J. P. (2007). *"Diversity, dispersal and biogeography of bryophytes (mosses)." in Protist Diversity and Geographical Distribution, Topics in Biodiversity and Conservation* Vol. 8. Eds. W. Foissner and D. L. Hawksworth (Dordrecht: Springer), 43–50. doi: 10.1007/978-90-481-2801-3_4

Fraser, C.II, Terauds, A., Smellie, J., Convey, P., and Chown, S. L. (2014). Geothermal activity helps life survive glacial cycles. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5634–5639. doi: 10.1073/pnas.1321437111

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915–925.

Hansen, J., Sato, M., Russell, G., and Kharecha, P. (2013). Climate sensitivity, sea level and atmospheric CO_2. *Philos. Trans. R. Soc A.* 371, 20120294. doi: 10.1098/rsta.2012.0294

Hedderson, T. A., Murray, D. J., Cox, C. J., and Nowell, T. L. (2004). Phylogenetic relationships of haplolepideous mosses (Dicranidae) inferred from rps4 gene sequences. *Syst. Biol.* 29, 29–41. doi: 10.1600/036364404772973960

Heinrichs, J., Hentschel, J., Feldberg, K., Bombosch, A., and Schneider, H. (2009). Phylogenetic biogeography and taxonomy of disjunctly distributed bryophytes. *J. Syst. Evol.* 47, 497–508. doi: 10.1111/j.1759-6831.2009.00028.x

Huttunen, S., Hedenäs, L., Ignatov, M. S., Devos, N., and Vanderpoorten, A. (2008). Origin and evolution of the northern hemisphere disjunction in the moss genus *Homalothecium* (Brachytheciaceae). *Am. J. Bot.* 95, 720–730. doi: 10.3732/ajb.2007407

Jules, E. S., and Shaw, A. J. (1994). Adaptation to metal-contaminated soils in populations of the moss, *Ceratodon purpureus*: vegetative growth and reproductive expression. *Am. J. Bot.* 81, 791–797. doi: 10.1002/j.1537-2197.1994.tb15516.x

Karlin, E. F., Hotchkiss, S. C., Boles, S. B., Stenøien, H. K., Hassel, K., Flatberg, K.II, et al. (2012). High genetic diversity in a remote island population system: *sans sex*. *New Phytol.* 193, 1088–1097. doi: 10.1111/j.1469-8137.2011.03999.x

Kops, J., Hartsen, F. A., and van Eeden, F. W. (1868). Flora Batava, of Afbeeldingen en Beschrijving van Nederlandsche gewassen.. *XIII Deel* 13 (Amsterdam, the Netherlands: J. C. Sepp en Zoon).

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 3, 1870–1874. doi: 10.1093/molbev/msw054

Kyrkjeeide, M. O., Hassel, K., Flatberg, K.II, Shaw, A. J., Brochmann, C., and Stenøien, H. K. (2016). Long-distance dispersal and barriers shape genetic structure of peatmosses (*Sphagnum*) across the Northern Hemisphere. *J. Biogeogr.* 43, 1215–1226. doi: 10.1111/jbi.12716

Laenen, B., Shaw, B., Schneider, H., Goffinet, B., Paradis, E., Désamoré, A., et al. (2014). Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nat. Commun.* 5 (1), 1–6. doi: 10.1038/ncomms6134

Lee, J. R., Raymond, B., Bracegirdle, T. J., Chades, I., Fuller, R. A., Shaw, J. D., et al. (2017). Climate change drives expansion of Antarctic ice-free habitat. *Nature* 547, 49. doi: 10.1038/nature22996

Leigh, J. W., and Bryant, D. (2015). popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Lewis, L. R., Rozzi, R., and Goffinet, B. (2014a). Direct long-distance dispersal shapes a New World amphitropical disjunction in the dispersal-limited dung

moss *Tetraplodon* (Bryopsida: Splachnaceae). *J. Biogeogr.* 41, 2385–2395. doi: 10.1111/jbi.12385

Lewis, L. R., Behling, E., Gousse, H., Qian, E., Elphick, C. S., Lamarre, J. -F., et al. (2014b). First evidence of bryophyte diaspores in the plumage of transequatorial migrant birds. *PeerJ* 2, e424. doi: 10.7717/peerj.424

Lewis, L. R., Biersma, E. M., Carey, S. B., Holsinger, K., McDaniel, S. F., Rozzi, R., et al. (2017). Resolving the northern hemisphere source region for the long-distance dispersal event that gave rise to the South American endemic dung moss *Tetraplodon fuegianus*. *Am. J. Bot.* 104, 1651–1659. doi: 10.3732/ajb.1700144

Longton, R. E. (1997). Reproductive biology and life-history strategies. *Adv. Bryol.* 6, 101.

Löytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320, 1632–1635. doi: 10.1126/science.1158395

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003

McDaniel, S. F., and Shaw, A. J. (2005). Selective sweeps and intercontinental migration in the cosmopolitan moss *Ceratodon purpureus* (Hedw.) Brid. *Mol. Ecol.* 14, 1121–1132. doi: 10.1111/j.1365-294X.2005.02484.x

McDaniel, S. F., Willis, J. H., and Shaw, A. J. (2007). A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. *Genetics* 176, 2489–2500. doi: 10.1534/genetics.107.075424

McDaniel, S. F., Willis, J. H., and Shaw, A. J. (2008). The genetic basis of developmental abnormalities in interpopulation hybrids of the moss *Ceratodon purpureus*. *Genetics* 179, 1425–1435. doi: 10.1534/genetics.107.086314

Muñoz, J., Felicísimo, Á. M., Cabezas, F., Burgaz, A. R., and Martínez, I. (2004). Wind as a long-distance dispersal vehicle in the Southern Hemisphere. *Science* 304, 1144–1147. doi: 10.1126/science.1095210

Nadot, S., Bajon, R., and Lejeune, B. (1994). The chloroplast generps 4 as a tool for the study of Poaceae phylogeny. *Plant Syst. Evol.* 191, 27–38.

Naish, T., Powell, R., Levy, R., Wilson, G., Scherer, R., Talarico, F., et al. (2009). Obliquity-paced Pliocene West Antarctic ice sheet oscillations. *Nature* 458, 322–328. doi: 10.1038/nature07867

Ochyra, R., Smith, R. I. L., and Bednarek-Ochyra, H. (2008). *The illustrated moss flora of Antarctica* (Cambridge, UK: Cambridge University Press).

Patiño, J., and Vanderpoorten, A. (2018). Bryophyte biogeography. *Crit. Rev. Plant Sci.* 37, 175–209. doi: 10.1080/07352689.2018.1482444

Piñeiro, R., Popp, M., Hassel, K., Listl, D., Westergaard, K. B., Flatberg, K.II, et al. (2012). Circumarctic dispersal and long-distance colonization of South America: the moss genus *Cinclidium*. *J. Biogeogr.* 39, 2041–2051. doi: 10.1111/j.1365-2699.2012.02765.x

Pisa, S., Biersma, E. M., Convey, P., Patiño, J., Vanderpoorten, A., Werner, O., et al. (2014). The cosmopolitan moss *Bryum argenteum* in Antarctica: recent colonisation or in situ survival? *Polar. Biol.* 37, 1469–1477. doi: 10.1007/s00300-014-1537-3

Pokorny, L., Oliván, G., and Shaw, A. J. (2011). Phylogeographic patterns in two southern hemisphere species of *Calyptrochaeta* (Daltoniaceae, Bryophyta). *Syst. Biol.* 36, 542–553. doi: 10.1600/036364411X583529

Pollard, D., and DeConto, R. M. (2009). Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature* 458, 329–332. doi: 10.1038/nature07809

Polvani, L. M., Waugh, D. W., Correa, G. J. P., and Son, S.-W. (2011). Stratospheric ozone depletion: the main driver of twentieth-century atmospheric circulation changes in the Southern Hemisphere. *J. Clim.* 24, 795–812. doi: 10.1175/2010JCLI3772.1

Pressel, S., and Duckett, J. G. (2019). Do motile spermatozoids limit the effectiveness of sexual reproduction in bryophytes? Not in the liverwort *Marchantia polymorpha*. *J. Syst. Evol.* 57 (4), 371–381. doi: 10.1111/jse.12528

Puillandre, N., Lambert, A., Brouillet, S., and Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* 21, 1864–1877. doi: 10.1111/j.1365-294X.2011.05239.x

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). *Tracer v1.6*. Available at: http://beast.bio.ed.ac.uk/Tracer.

Robinson, S. A., and Erickson, D. J. „.IIIII (2015). Not just about sunburn – the ozone hole's profound effect on climate has significant implications for Southern Hemisphere ecosystems. *Glob. Change Biol.* 21, 515–527. doi: 10.1111/gcb.12739

Robinson, S. A., Klekociuk, A. R., King, D. H., Rojas, M. P., Zúñiga, G. E., and Bergstrom, D. M. (2020). The 2019/2020 summer of Antarctic heatwaves. *Glob. Change Biol.* 26, 3178–3180. doi: 10.1111/gcb.15083

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Scherer, R. P., Bohaty, S. M., Dunbar, R. B., Esper, O., Flores, J. A., Gersonde, R., et al. (2008). Antarctic records of precession-paced insolation-driven warming during early Pleistocene Marine Isotope Stage 31. *Geophys. Res. Lett.* 35, L03505. doi: 10.1029/2007GL032254

Schofield, W. B., and Crum, H. A. (1992). "Bryophyte distribution patterns", in *Bryophytes and Lichens in a Changing Environment*, ed. Bates, J. W., and Farmer, A. M. (Oxford: Clarendon Press).

Schofield, W. B., and Crum, H. A. (1972). Disjunctions in Bryophytes. *Ann. Missouri Bot. Gard.* 59, 174–202.

Shaw, J., and Beer, S. C. (1999). Life history variation in gametophyte populations of the moss *Ceratodon purpureus* (Ditrichaceae). *Am. J. Bot.* 86 (4), 512–521.

Shaw, A. J., Carter, B. E., Aguero, B., da Costa, D. P., and Crowl, A. A. (2019). Range change evolution of peat mosses (*Sphagnum*) within and between climate zones. *Glob. Change Biol.* 25 (1), 108–120. doi: 10.1111/gcb.14485

Silvestro, D., and Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol.* 12, 335–337. doi: 10.1007/s13127-011-0056-0

Simmons, M. P., and Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381. doi: 10.1093/sysbio/49.2.369

Souza-Chies, T. T., Bittar, G., Nadot, S., Carter, L., Besin, E., and Lejeune, B. (1997). Phylogenetic analysis of Iridaceae with parsimony and distance methods using the plastid gene rps4. *Plant Syst. Evol.* 204, 109–123.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stech, M., and Quandt, D. (2010). 20,000 species and five key markers: The status of molecular bryophyte phylogenetics. *Phytotaxa* 9, 196–228. doi: 10.11646/phytotaxa.9.1.11

Stenøien, H. K., Shaw, A. J., Shaw, B., Hassel, K., and Gunnarsson, U. (2011). North American origin and recent European establishments of the amphi-Atlantic peat moss *Sphagnum angermanicum*. *Evolution* 65, 1181–1194. doi: 10.1111/j.1558-5646.2010.01191.x

Sundberg, S. (2013). Spore rain in relation to regional sources and beyond. *Ecography* 36, 364–373. doi: 10.1111/j.1600-0587.2012.07664.x

Szövényi, P., Sundberg, S., and Shaw, A. J. (2012). Long-distance dispersal and genetic structure of natural populations: an assessment of the inverse isolation hypothesis in peat mosses. *Mol. Ecol.* 21, 5461–5472. doi: 10.1111/mec.12055

Szövényi, P., Perroud, P. F., Symeonidi, A., Stevenson, S., Quatrano, R. S., Rensing, S. A., et al. (2014). *De novo* assembly and comparative analysis of the *Ceratodon purpureus* transcriptome. *Mol. Ecol. Resour.* 15, 203–215. doi: 10.1111/1755-0998.12284

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17, 1105–1109.

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics* 123, 597–601.

Thompson, D. W., Solomon, S., Kushner, P. J., England, M. H., Grise, K. M., and Karoly, D. J. (2011). Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nat. Geosci.* 4, 741–749. doi: 10.1038/ngeo1296

Turner, J., Barrand, N. E., Bracegirdle, T. J., Convey, P., Hodgson, D. A., Jarvis, M., et al. (2014). Antarctic climate change and the environment: an update. *Polar. Rec.* 50, 237–259. doi: 10.1017/S0032247413000296

Vanderpoorten, A., Goffinet, B., and Quandt, D. (2006). "Utility of the internal transcribed spacers of the 18S-5.8 S-26S nuclear ribosomal DNA in land plant systematics with special emphasis on Bryophytes," in *Plant Genome: Biodiversity and Evolution, vol. 2, B, Lower groups*. Eds. A. K. Sharma and A. Sharma (Enfield, New Hampshire, USA: Science Publishers), 385–407.

Vanderpoorten, A., Rumsey, F., and Carine, M. (2007). Does Macaronesia exist? Conflicting signal in the bryophyte and pteridophyte floras. *Am. J. Bot.* 94, 625–639. doi: 10.3732/ajb.94.4.625

Vigalondo, B., Garilleti, R., Vanderpoorten, A., Patiño, J., Draper, I., Calleja, J. A., et al. (2019). Do mosses really exhibit so large distribution ranges? Insights from the integrative taxonomic study of the *Lewinskya affinis* complex (Orthotrichaceae, Bryopsida). *Mol. Phylogenet. Evol.* 140, 106598. doi: 10.1016/j.ympev.2019.106598

Villarreal, J. C., and Renner, S. S. (2014). A review of molecular-clock calibrations and substitution rates in liverworts, mosses, and hornworts, and a timeframe for a taxonomically cleaned-up genus *Nothoceros*. *Mol. Phylogenet. Evol.* 78, 25–35. doi: 10.1016/j.ympev.2014.04.014

Villaverde, T., Escudero, M., Martín-Bravo, S., Jiménez-Mejías, P., Sanmartín, I., Vargas, P., et al. (2017).). Bipolar distributions in vascular plants: A review. *Am. J. Bot.* 104 (11), 1680–1694. doi: 10.3732/ajb.1700159

Waterman, M. J., Nugraha, A. S., Hendra, R., Ball, G. E., Robinson, S. A., and Keller, P. A. (2017). Antarctic moss Biflavonoids show high antioxidant and ultraviolet-screening activity. *J. Nat. Prod.* 80 (8), 2224–2231. doi: 10.1021/acs.jnatprod.7b00085

Wertheim, J. O., Sanderson, M. J., Worobey, M., and Bjork, A. (2010). Relaxed molecular clocks, the bias–variance trade-off, and the quality of phylogenetic inference. *Syst. Biol.* 59, 1–8. doi: 10.1093/sysbio/syp072

White, T. J., Bruns, T., Lee, S., and Taylor, J. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics," in *PCR protocols: a guide to methods and applications*. Ed. M. A. Innis (New York: Academic Press), 315–322.

Wilkinson, D. M., Koumoutsaris, S., Mitchell, E. A. D., and Bey, I. (2012). Modelling the effect of size on the aerial dispersal of microorganisms. *J. Biogeogr.* 39, 89–97. doi: 10.1111/j.1365-2699.2011.02569.x

Wyber, R. A. (2013). *Phylogeny of Three East Antarctic Mosses. [PhD thesis]* (Wollongong, (Australia: University of Wollongong).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 September 2020
doi: 10.3389/fpls.2020.536862

Cell Division Patterns in the Peristomial Layers of the Moss Genus *Costesia:* Two Hypotheses and a Third Solution

Michael S. Ignatov[1,2]*, *Ulyana N. Spirina*[2,3], *Maria A. Kolesnikova*[2], *Juan Larraín*[4] *and Elena A. Ignatova*[1]

[1] Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia, [2] Tsitsin Main Botanical Garden, Russian Academy of Sciences, Moscow, Russia, [3] Faculty of Biology, Tver State University, Tver, Russia, [4] Instituto de Biología, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

The Chilean endemic genus *Costesia* belongs to the Gigaspermaceae, one of the most basal groups of arthrodontous mosses. While none of the species in this family has a peristome, earlier stages of sporophyte development often disclose its basic structure. The study of *Costesia* sporophytes at the early stages of development was conducted to identify possible similarities with *Diphyscium*, the genus sister to Gigaspermaceae plus all other arthrodontous mosses in the moss phylogenetic tree. *Diphyscium* shares a strongly unequal cell division pattern with the Dicranidae. In groups more closely related to *Diphyscium*, as it is the case of *Costesia*, this pattern is not known. Our study of *Costesia* found only irregular presence of slightly unequal cell divisions that may then be considered as a plesiomorphic state in peristomate mosses. The most frequently present pattern revealed in *Costesia* is common with the Polytrichaceae, a more basal moss group with nematodontous peristomes.

Keywords: Gigaspermaceae, peristome, bryophytes, haplolepideous, diplolepideous, morphogenesis

OPEN ACCESS

Edited by:
Jeffrey Graham Duckett, Natural History Museum, United Kingdom

Reviewed by:
Sean Rowan Edwards, The University of Manchester, United Kingdom
Jessica M. Budke, The University of Tennessee, United States

***Correspondence:**
Michael S. Ignatov
misha_ignatov@list.ru

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *21 February 2020*
Accepted: *18 August 2020*
Published: *04 September 2020*

Citation:
Ignatov MS, Spirina UN, Kolesnikova MA, Larraín J and Ignatova EA (2020) Cell Division Patterns in the Peristomial Layers of the Moss Genus Costesia: Two Hypotheses and a Third Solution. Front. Plant Sci. 11:536862. doi: 10.3389/fpls.2020.536862

HIGHLIGHTS

We looked for the presence of unequal cell divisions in peristomial layers of *Costesia*, one of the most basal lineages of arthrodontous mosses. Although found at the earliest stages, unequal cell divisions are not stably exhibited throughout the division process. The main developmental pattern in this genus unexpectedly links it to the nematodontous peristome type of the Polytrichaceae.

INTRODUCTION

The peristome is a structure at the mouth of moss capsules, which enhances the process of spore release by means of hygroscopic movements. The role of the peristome for Bryophyta classification is as important as that of flower and fruit structure for Magnoliophyta. Molecular phylogenetic data in general coincide with the classification based on the peristome structure, defining main subclasses (Newton et al., 2000; Frey and Stech, 2009; Goffinet et al., 2009; Shaw et al., 2011). Recent

comprehensive analysis by Liu et al. (2019) summarized molecular data and built a robust phylogenetic tree, where peristomate mosses are found in a grade of seven classes/subclasses and two terminal clades, the subclasses Bryidae (with diplolepideous alternate peristomes) and Dicranidae (with haplolepideous peristomes). These two subclasses include no less than 90% of the contemporary species diversity. Accordingly, the mosses of the grade from Takakiales to Timmiales (**Figure 1**), have much fewer species in the modern flora.

The peristomes of taxa not included in the Bryidae or the Dicranidae are more diverse and not as well studied compared to peristomes of these two terminal clades. Especially enigmatic is the transition from mosses with nematodontous peristomes, formed by entire cells, to mosses with arthrodontous peristomes. The latter are understood in recent literature as a group starting from *Diphyscium* (Shaw et al., 1987; Goffinet et al., 2009), although the distinction between arthrodontous and nematodontous mosses was treated differently. Originally, Mitten (1859) defined neamtodontous peristome as lacking conspicuous transverse cell wall remains, which was followed, among others, by Philibert (1884), translation by Taylor (1962). Fleischer (1904) was the first who grouped Polytricaceae and Tetraphidaceae together; however in the 'Fleischer-Brotherus system' in Die Natürlichen Pflanzenfamilien (Brotherus, 1925) Tetraphidaceae is placed near Bryaceae. Later Dixon (1932) included *Tetraphis* into nematodontous mosses, but together with Buxbaumiales, Diphysciales, Schistostegales, and Calomniales. Until recently, the Polytrichopsida was placed in phylogenetic systems ancestral to Tetraphidopsida (Goffinet et al., 2009). Morphology also supports the position of *Tetraphis* closer to arthrodontous mosses than to the Polytrichaceae (Ligrone and Duckett, 2011), and Shaw and Anderson (1988) concluded that the peristome development in *Tetraphis* is more similar to arthrodontous mosses than to the Polytrichaceae. However, the expanded analyses of Liu et al. (2019) found Tetraphidaceae in a more basal position than Polytrichaceae, albeit with low support. Bell et al. (2020) showed that the Tetraphidaceae is sister to the Polytrichaceae plus all other peristomate mosses in the analysis of plastid data, whereas mitochondrial data places these two families in a clade sister to all other peristomate mosses. In summary, the interrelationships of nematodontous mosses and especially the transition to arthrodontous mosses remains insufficiently understood, being inconsistent in studies of different authors.

The peristome types are characterized by a peristomial formula established by Edwards (1979), that consists of numbers of cells in three (excepting rare exotic cases) peristome-forming layers, so called the outer, primary and inner peristomial layers (OPL, PPL, and IPL), as defined by Blomquist and Robertson (1941). The numbers in the peristomial formulae correspond to the numbers of cells in one-eighth of each layer. Importance and variants of peristomial formulae are only briefly illustrated here in **Figure 1**, as they are discussed in detail in numerous publications (Edwards, 1979; Edwards, 1984; Shaw et al., 1987; Shaw et al., 1989a; Shaw et al., 1989b; Shaw et al., 2011; Ignatov et al., 2015; Ignatov et al., 2018a).

Despite three main peristomial types, the diplolepideous opposite, haplolepideous, and diplolepideous alternate, were found remarkably strictly confined to three main phylogenetic lineages of mosses, Funariidae, Dicranidae and Bryidae (**Figure 1**), the question about the ancestral state remains open. The limited number of peristome development studies precludes the complete understanding of its evolution (Liu et al., 2019). Specifically, the opposite versus alternate arrangement of exostome against endostome elements remains enigmatic (Budke et al., 2007). One of the intriguing points is how *Diphyscium* and Dicranidae, separated by the "diplolepideous opposite" groups in the grade from Takakiales to Timmiales (**Figure 1**) share strongly unequal cell divisions, appearing in a 4:2:3 pattern at the early stage of peristome development.

Searching for an answer to this question, we addressed the Gigaspermaceae which has an intermediate phylogenetic position between *Diphyscium* and all other arthrodontous mosses. The Gigaspermaceae includes five genera, and none of them has a peristome (Goffinet et al., 2009). The numerous peristome development studies done in eperistomate mosses of the genera *Aphanorrhegma, Ephemerum*, and *Physcomitrium* have brought interesting results (Schwartz, 1994). One species of Gigaspermaceae, *Lorentziella imbricata* (Mitt.) Broth., was also studied for sporophyte development by Rushing and Snider (1980), however the strong peristome reduction in that cleistocarpous genus, as well as an undeveloped methodology prior the peristomial formula approach appeared, did not shed enough light on the problem.

Liu et al. (2012) phylogenetic analysis revealed two lineages in the Gigaspermaceae, one with strongly reduced sessile sporophytes including two genera, *Gigaspermum* and *Lorentziella*, and a second clade including *Oedipodiella*, *Chamaebryum*, and *Costesia*, the two latter having elongate setae, i.e., with less reduced sporophytes. *Costesia* was found in the basal position to the two other genera of the second clade, and therefore is the most promising for peristome studies. Two alternative hypotheses were tested: the first one implies that *Costesia* possesses a 4:2:3 pattern, inherited from the Diphysciales, which is the closest ancestral group in the phylogenetic tree (**Figure 1**); an alternative hypothesis was that *Costesia* has opposite cell arrangement in peristomial layers similar to that in diplolepideous opposite groups, the Funariidae, where Gigapermaceae were placed earlier, and which is the closest descendant group in the grade from Takakiales to Timmiales.

MATERIALS AND METHODS

Samples

Plants of *Costesia macrocarpa* (Schimp.) Cuvertino, Miserere & Buffa (= *C. spongiosa* Thér.) with young sporophytes were collected from Lago Peñuelas National Reserve, in Valparaíso Region, central Chile (Chile, Valparaíso Region, Valparaíso

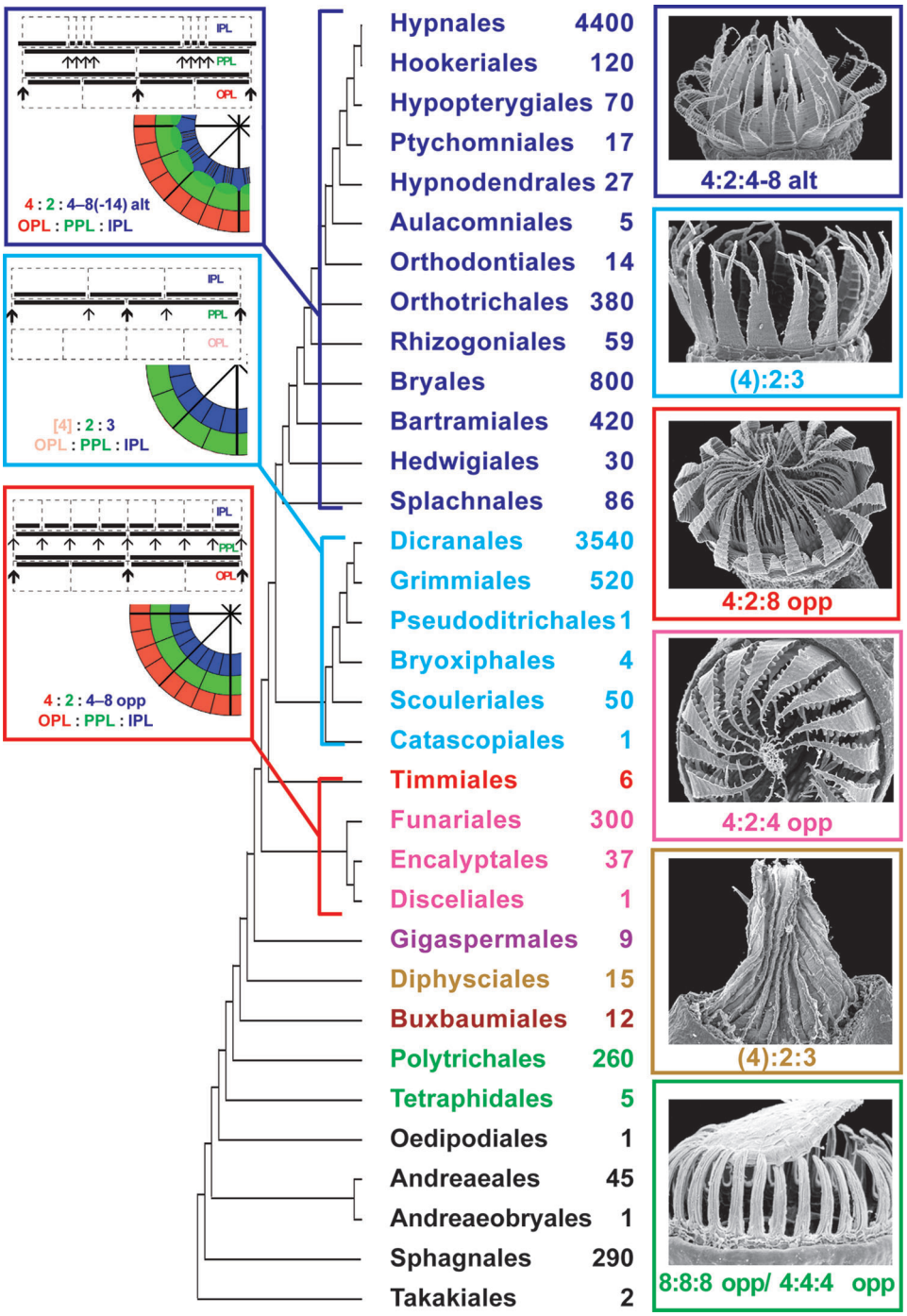


FIGURE 1 | Phylogeny of mosses, following in general Liu et al., 2019, showing the approximate number of species in orders calculated from literature data (e.g., Flora of North America North of Mexico, 2007, 2014) and Tropicos Database (http://legacy.tropicos.org/namesearch.aspx, accesses 15 January 2020). SEM images of one representative with most common structure of completely developed peristome and most common peristomial formula are also given for each lineage of peristome diversification (nematodontous, diplolepideous opposite, haplolepideous, and diplolepideous alternate). Orders of eperistomate mosses are given in black, while orders of peristomate mosses of one subclass (or class) are given in one color: Tetraphidopsida (light green), Polytrichopsida (green), Buxbaumiidae (dark brown), Diphysciidae (brown), Gigaspermidae (purple), Funariidae (magenta), Timmiidae (red), Dicranidae (cyan), Bryidae (blue). Schemes illustrating peristomial formulae of the diplolepideous opposite, haplolepideous, and diplolepideous alternate peristomes are also shown. Characteristic representatives are illustrated with SEM images from Ignatov and Ignatova (2003, 2004), based on writing permission).

Province, National Reserve Lago Peñuelas, 33°10'58"S, 71° 29'12"W, 338 m alt., copse of *Quillaja saponaria*, on sandy soil, 24 June 2019, Ignatov, Ignatova & Larraín, s.n., MHA).

Some living plants were subsequently cultivated in the testing chamber MLR32 Sanyo (temperature + 7°C/+ 12°C (night/day), light period 10 h, PPFD - 14μmol • m^{-2} • s^{-1}), thus enabling various stages of sporophyte development to be studied.

Additionally, the sections of young sporophytes of *Atrichum undulatum* (Hedw.) P. Beauv. were included for comparison. Specimens were fixed and observed with the same protocols outlined below (voucher: Moscow, Tsitsin Main Botanical Garden, 11 Sept 2019, Ignatov & Spirina s.n., MHA).

Anatomy Studies

All material collected in the field was fixed shortly after collecting in 2.5% glutaraldehyde in 0.05M PBS. Further steps were performed after several weeks or a few months. Specimens were post-fixed with 1% osmium tetroxide in PBS, pH 6.8, for 6 h. The material was then dehydrated through an ascending ethanol-acetone series to 100% acetone. Next, samples were embedded in Araldite 6005 medium, according to the manufacturer's protocol. Sections were cut 2 μm thick with glass knives, put on glass slides without mounting medium, stained with 0.01% berberine or in combination with DAPI, and scanned under LSCM Olympus FV-1000 based on Olympus BX61, using 473 nm or combination of 405 and 473 nm lasers. Z-stacks of several scans were usually obtained and are presented here.

We made a complete series of sections 2 μm thick, allowing measure the distance from the section where sporophyte apical cell first appears within the ring of calyptra. Sectioning was stopped when patterning of peristomial layers became irregular. The series show different stages of sporophyte development, as can be assumed from the number of cells in the peristomial layers, especially in IPL, which however do not necessarily correspond to the length and width of sporophytes. Altogether twenty series of transverse sections and sixteen series of longitudinal sections were done, of which eleven series of transverse sections represent specific patterns discussed below (**Supplementary Materials 1**). An additional study of *Atrichum undulatum* was done for comparison after a pattern resembling Polytrichaceae was noticed in *Costesia*. It included serial sections for four young capsules; razor blade sectioning of fresh material was also made in order to confirm that the pattern seen in serial sections can readily be found in other populations.

RESULTS

The longitudinal sections represent plants at stages from a whole sporophyte length of 0.4 mm (**Figure 2B**) to a length of 1.7 mm for capsule only (**Figure 3C**), when the sporogeneous layer is apparently differentiated. In longitudinal sections of young sporophytes, the area of enlarged cells is seen right below the apex, corresponding to the endothecium at the level where peristome is formed in most moss species (**Figure 2**). It is formed of large cells 20–30 μm wide (**Figure 2B**), without or with very few cell divisions in them. Cells below this zone are smaller because much more divisions took place. The amphithecium compartments are perpendicular to capsule wall in the lower part of this "peristome zone", and the more distal, the more oblique they are. Cells of the "urn zone" at earlier stages (amphithecium 1–3 cells thick, width of sporophyte at annulus level 110–130 μm, **Figures 2B, C, E**) are arranged in longitudinal rows without apparent differentiation. Later (amphithecium 4–5 cells thick, width of sporophyte at annulus level 160 μm, **Figure 3A**) cells in the middle of the "urn zone" become more variable in shape, while rows connected to lower cells of endothecium in "peristome zone" are more or less apparently differentiated. Their relative regularity and position correspond to their further differentiation into sporogeneous tissue.

At early stages, when the whole sporophyte is ca. 800 μm long, the "peristome zone", "urn zone", "seta zone", and "foot zone" occupy ca. 150, 250, 250, and 150 μm respectively (**Figure 2E**). The "seta zone" is composed of strongly elongate cells, while the foot includes numerous short cells with thin walls, becoming especially wavy due to fixation (**Figure 2E**).

Later on, cells of the 'peristome zone' additionally divide and enlarge, forming parenchymal cells beneath the operculum. The 'urn zone' is five times enlarged both longitudinally and transversally, and the spore sac is separated from the capsule wall, and hangs on filaments (**Figure 3C**). In the spore sac, the sporogeneous layer is differentiated (**Figure 3B**) by darker color of less vacuolated cells, more regular in shape compared to the tissue of the massive columella. The outer spore sac layer continues in the peristomial zone in the IPL layer (**Figure 3A**). At the stage shown in **Figures 3C, D240, D248**, the regular patterns in the "peristome zone" are already lost. Cells above the urn are irregularly arranged, except only at the level of the uppermost 'urn zone', where cells around endothecium form a conspicuous ring of radially elongate cells, at places approaching four per 1/8 sector. At this stage operculum tissue is much expanded (**Figure 3** D150, cf. 3C). However, at the earlier stage (**Figure 3A**), cells are regular enough to see patterns characteristic for peristomial layers in transverse sections (**Figures 4**).

These regular cell patterns are apparent only in a limited interval of 10–20 μm, rarely up to ca. 40 μm, corresponding to levels shortly above the urn top (**Figure 3A**). They are not distinct in all the sectors of the circumference (cf. **Figure 4** and **Supplementary Materials 1**) and not in all series of sections, being still not expressed in the youngest one, where amphithecium is only two-layered, and the oldest, where cells of peristomial layer lose their regular arrangement (**Figure 3D**248). However, the peristomial formula observed in most octants in capsules on the stages where the cell patterns are most evident (**Figure 4**) deserves to be discussed.

Four series of transverse sections are shown in **Figure 4**. Here we show different sporophyte developmental stages, which however do not necessarily correspond to the length and width of the sporophytes, and therefore we arrange them by the number of cells, especially in IPL

The series of the youngest sporophytes have too irregular cell arrangements, thus not relevant for the discussion of peristomial

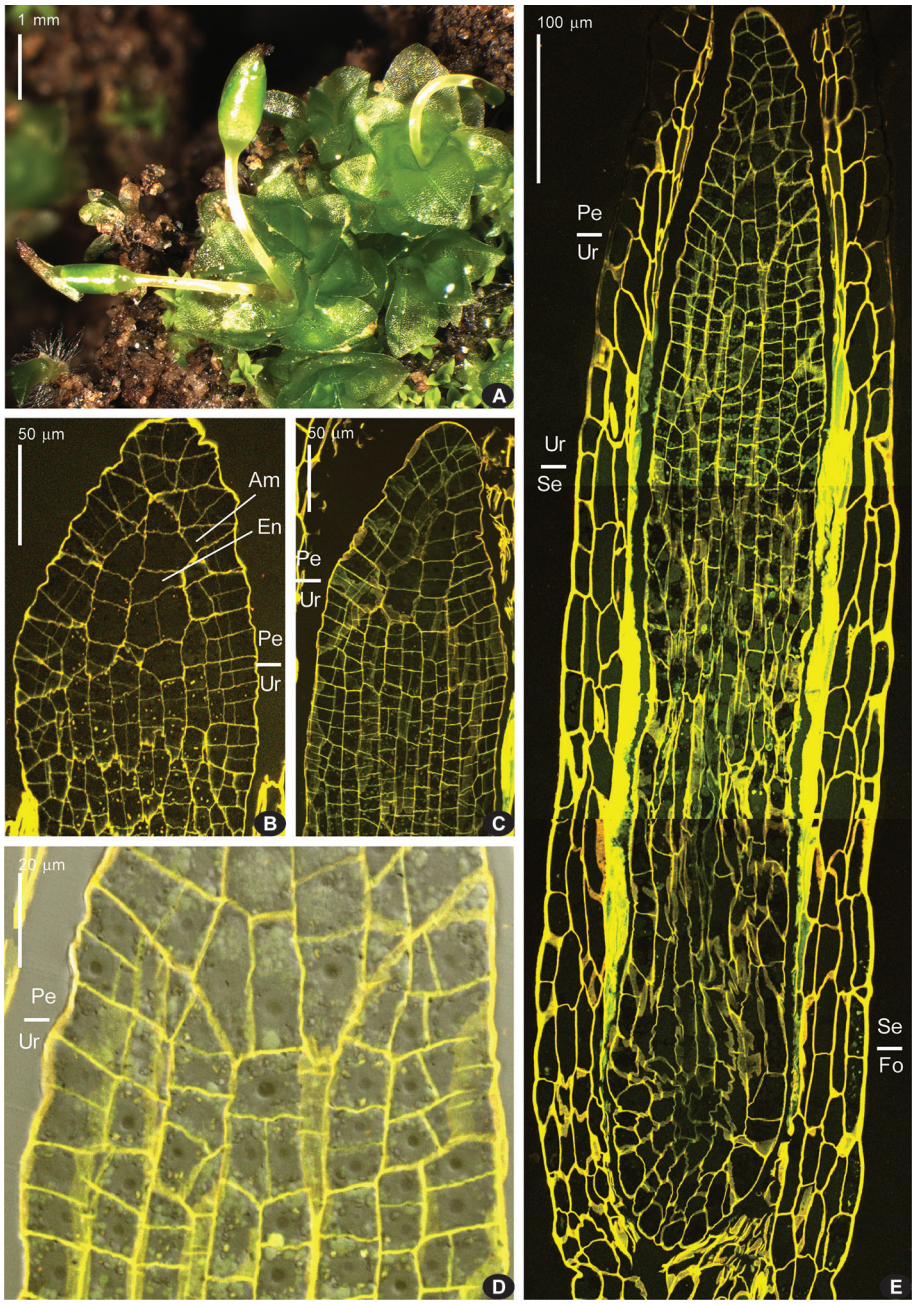


FIGURE 2 | *Costesia macrocarpa* habit (under stereomicroscope) and longitudinal sections (LSCM) of young sporophytes. **(A)** habit of plants with immature capsules, note sporophyte on the far right with strongly curved seta, a not rare case, and two sporophytes in one perichaetium in the center; **(B)** upper part of sporophyte, with peristome zone ca. 80 µm long; (Am), amphithecium, (En), endothecium; **(C)** next stage of sporophyte development, cells in urn zone are arranged in rather regular rows, but still not differentiated into sporogeneous tissue and columella; **(D)** close up of the transition from urn zone (Ur) to peristome zone (Pe), where the direction of cell compartments in amphithecium turns to oblique; **(E)** epigonial stage of sporophyte development, with "peristome zone" (Pe), "urn zone" (Ur), "seta zone" (Se), and "foot zone" (Fo). Bright fluorescence of mucilage is especially conspicuous in the seta zone.

layers. The sections with apparent patterning were usually quite a few in *Costesia* compared to other arthrodontous mosses, and the younger the capsule was the fewer were the sections useful for the analysis.

A relatively young sporophyte, putatively comparable with that shown in **Figure 2B** is presented in **Figure 4A**. Peristomial formulae in **Figure 4** A150 are 2:2:1, 4:2:1, and 4:2:2, with anticlinal cell walls in IPL strongly offset (small arrows) against

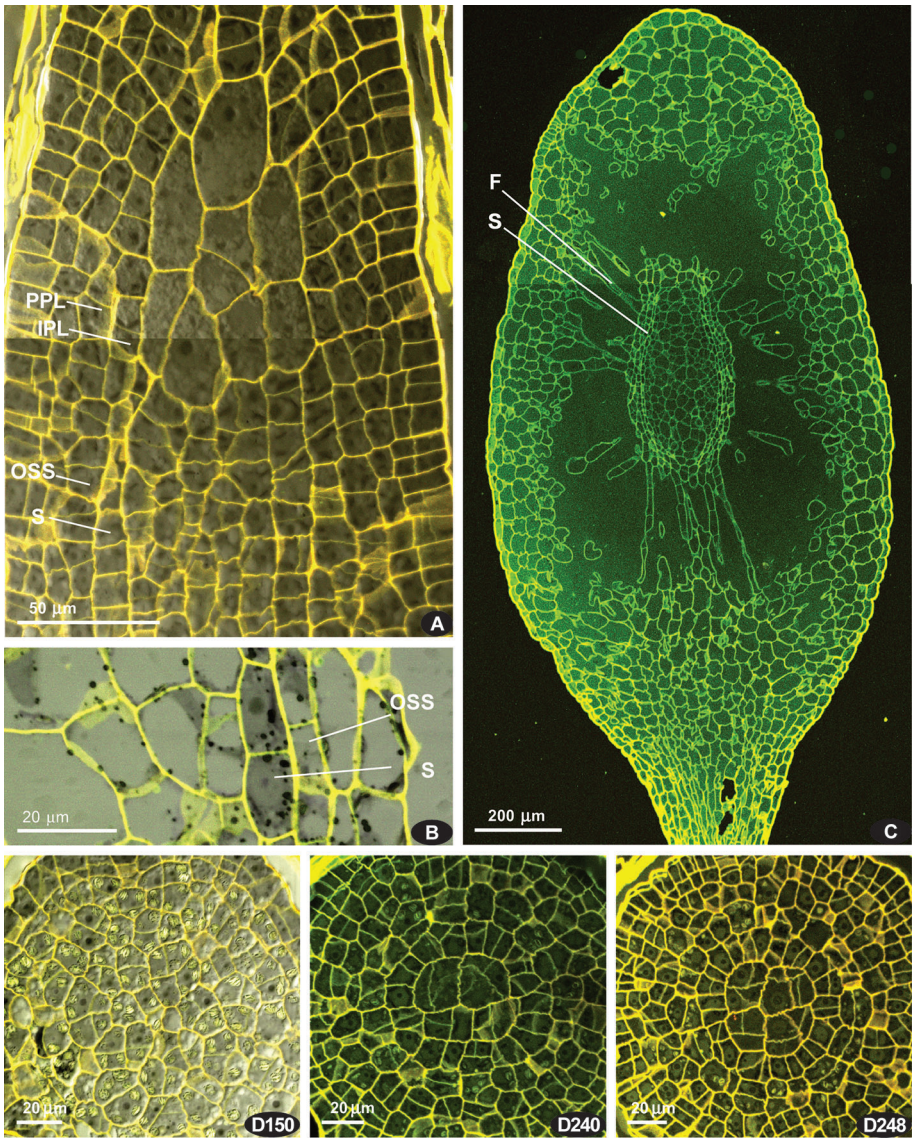


FIGURE 3 | *Costesia macrocarpa*, longitudinal sections, LSCM. **(A)** beginning of differentiation of the sporogeneous layer (S), discernible by quadrate cell shape; note also a relatively regular arrangement of cells in lower part of the peristomial zone; **(B)** spore sac part, with sporogreneous layer contrasted by adding of pseudo-transmission channel of the LSCM; **(C)** capsule, showing the position of spore sac; **(D)** series of transverse sections of capsule of moderately late stage, though earlier than shown in **(C)**; peristomial layers lost patternization at this stage except the innermost ring of longitudinally transverse cells (D248); numbers indicate the distance from the sporophyte apex in µm.

anticlinal cell walls in PPL, thus representing a 'haplolepideous pattern'. The same pattern is seen in the sections at the next 10 µm. Noteworthy is the presence of anticlinal-curved divisions that cut off PPL cells, and sometimes also in IPL cells, in an unusual way (cf. **Figure 4** A156).

The series B comprises a {4}:4:2 pattern, where two cells of the octant belong to IPL, and four cell to the PPL, while OPL is still undifferentiated. Therefore the cells outwards the PPL, which are outermost amphithecial and epidermal cells at the same time at this stage are shown in braces. The {4}:4:2 pattern appeared in this series since the beginning of the PPL differentiation, cells outwards IPL remain undivided in some octants (i.e., PPL *sensu stricto* is still lacking, **Figure 4** B96), to the level where this {4}:4:2 formula is seen in the majority of the 1/16 sectors (**Figure 4** B140). Thus, the {4}:4:2 formula occurs throughout 44 µm interval, being clearly expressed in between 96 µm and 140 µm from capsule apex in all sections (e.g., **Figure 4** B108).

In the other series were {4}:4:2 pattern occurs, it is expressed usually in 10 to 20 µm intervals at the transition from peristomial zone to urn zone, which is apparent as cells in the middle of the transverse section are smaller and lacking quadrant arrangement (cf. **Figure 4** B96 and C102 with B140 and C130).

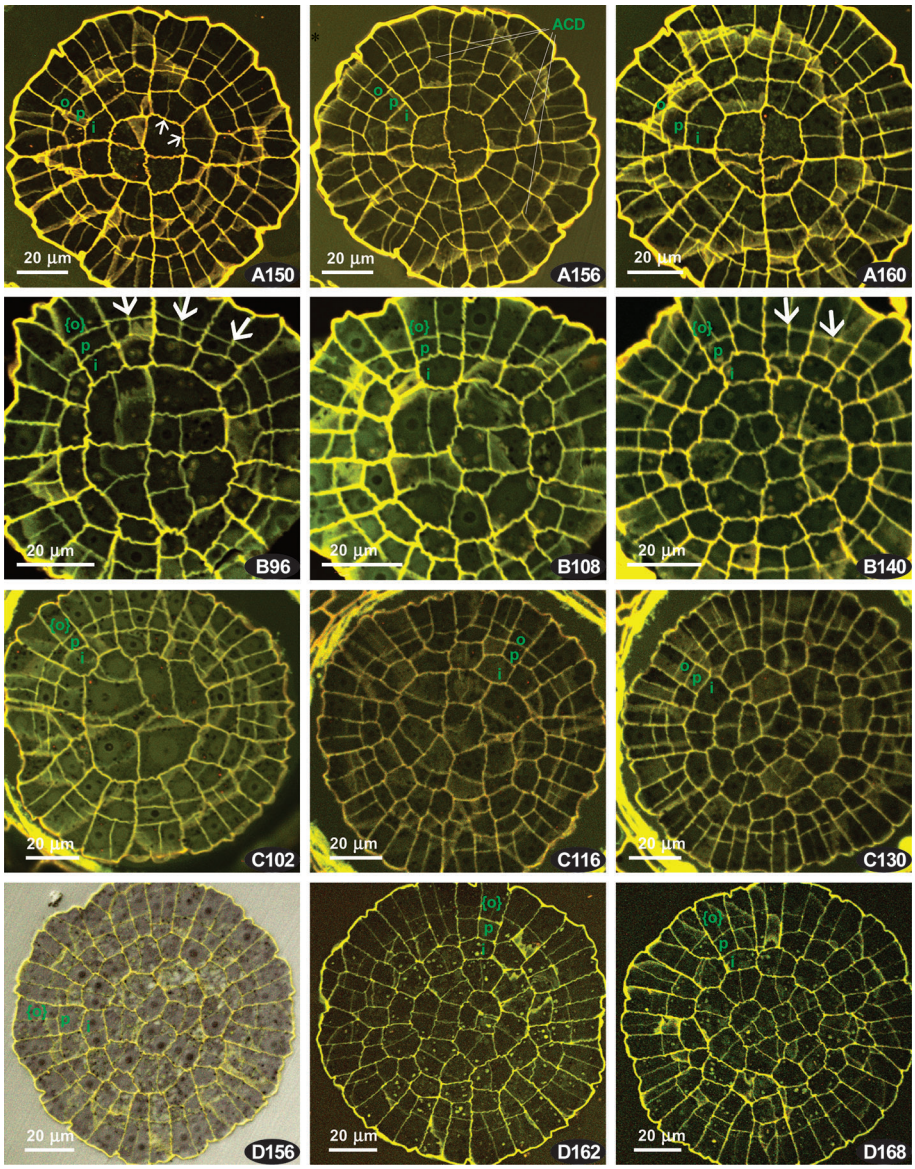


FIGURE 4 | *Costesia macrocarpa*, transverse sections of young capsules, LSCM. **(A–D)** four series of transverse sections; numbers indicate the distance from the sporophyte apex in µm. Small arrows point to anticlinal cell walls resulted in offset divisions in IPL; large arrows point to additional divisions in PPL, resulting in {4}:4:2 formula; ACD, anticlinal curved divisions. The OPL, PPL, and IPL layers are denoted by o, p. i, and in the case of three cells thick amphithecium the outermost layer is marked as {o}.

The presence of the {4}:4:2 pattern characteristic for the Polytrichaceae (Wenderoth, 1931) forced us to compare *Costesia* with the Polytrichaceae, so we studied sporophytes at early stages of development in *Atrichum undulatum*. There is also a variation between capsules, and two variants (out of six studied) are shown in **Figure 5**. Unexpectedly, in some series at the early stages, but where the amphithecium is already three-layered, a number of 2:2 patterns with slight but distinct offset was observed (**Figure 5** A62, A70), although in other series this pattern is lacking and all the divisions were fairly regular, despite sporophytes superficially looked almost identical (**Figure 5B**). Later the formula {4}:4:2 appeared, and then cells in IPL divided more or less precisely aligned to PPL cells, resulting in a {4}:4:4 and later a 4:4:4 (cf. **Figure 5** A170, A188, A202), and 8:8:8 (Wenderoth, 1931). However, in all four series of sections and additional studies of fresh material of *Atrichum* we observed the {4}:4:2 formula along a considerable distance at early stages of peristome development (**Figure 5** and **Supplementary Materials 1**).

DISCUSSION

The comparison of complete series of the transverse sections with longitudinal sections (**Figures 2** and **3**) ensures that the

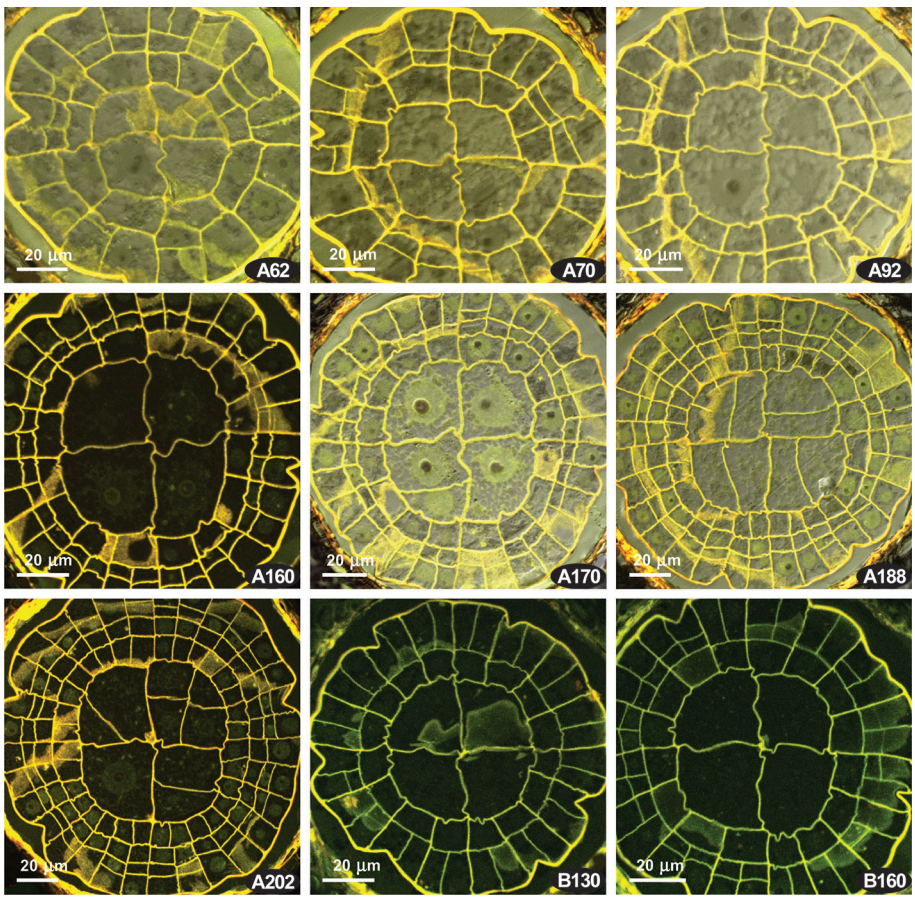


FIGURE 5 | *Atrichum undulatum*, transverse sections of young capsules, LSCM. **(A, B)** two series of transverse sections; numbers indicate the distance from the sporophyte apex in µm.

regular patterns illustrated in **Figure 4** correspond to the level of the proximal part of the "peristomial zone", thus the discussion on the peristomial formula for *Costesia* should be as important as it is for other eperistomate mosses, like *Physcomitrium* and *Aphanorhegma* (Schwartz, 1994).

The obtained observation revealed two patterns in *Costesia* worth discussing: (1) asymmetric divisions occur in IPL in most distal part of the amphithecium (cf. **Figure 4**, Table in **Supplementary Materials 1**), and (2) number of cells in PPL twice as many as in IPL.

Asymmetric cell divisions in IPL are typically associated with a haplolepideous pattern type. Some strongly asymmetric divisions were found in *Costesia*, but they are irregular and occur in a limited distal part of the sporophyte (**Supplementary Materials 1**). Such irregular asymmetric divisions were also observed in *Atrichum*, but only in a short interval below the sporophyte apex (**Supplementary Materials 1**).

Budke et al. (2007) suggested to separate (and differently score) slightly offset and strongly offset divisions in peristomial layers. These authors came to the conclusion that strong offset anticlinal divisions in the IPL occur in *Diphyscium* and in the majority of the studied haplolepideous mosses, whereas in other groups these divisions are less prominent. The published illustration shows that the slightly offset divisions occur in such basal lineages as *Tetraphis* (Shaw and Anderson, 1988) and *Oedipodium* (Shimamura and Deguchi, 2008), and at early stages of sporophyte development in some representatives of the "diplolepideous opposite" group, e.g., *Physcomitrium* (Schwartz, 1994), and in "diplolepideous alternate" mosses (Shaw et al., 1989a). Taking these points into consideration, the presence of offset divisions in few octants at the earlier stage of development in *Costesia* (**Figure 4**) and even *Atrichum* (**Figure 5**) may not be a significant phylogenetic marker.

In **Figure 6**, the distribution of various cell divisions is summarized, allowing for the conclusion that the least definite case of slightly offset divisions occurs in most groups in certain taxa, or at certain stages of development (based on published data, see **Supplementary Materials 2**). The developmental patterns based on strictly determined unequal cell divisions occur in Dicranidae and Bryidae, two groups that include no less than 85% of species of the world moss flora. Distinctly aligned cell divisions are characteristic of the arthrodontous

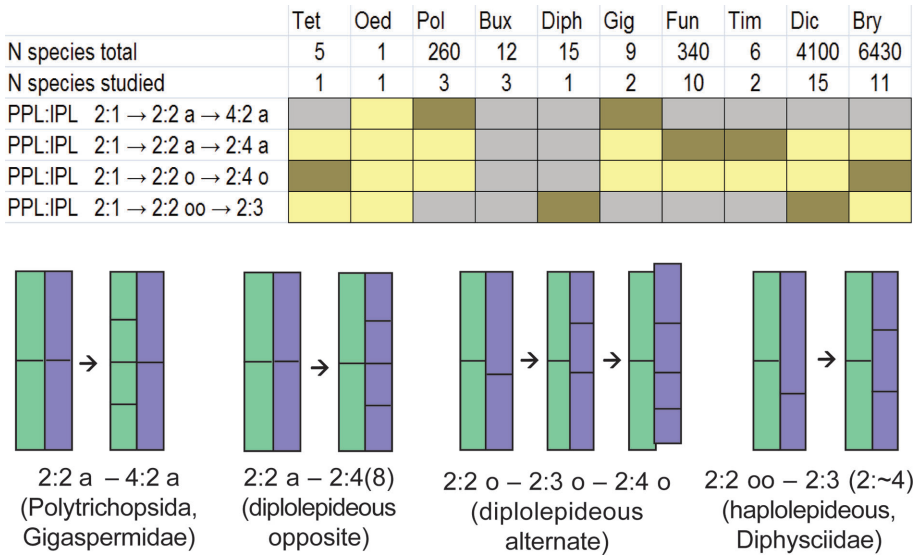


FIGURE 6 | Distribution of aligned, slightly offset and strongly offset positions of the anticlinal cell wall in IPL against corresponding cell wall in PPL at the critical stage of transition from peristomial formula x:2:1 to x:2:2 among classes, subclasses and some orders of peristomate mosses based on literature data combined in **Supplementary Materials 2** and on our observations. Letter 'a' indicates aligned position of anticlinal cell walls in IPL against anticlinal cell walls in PPL, letter 'o' denotes a slightly offset position (5%–32%), while "oo" means a strongly offset position (>33%). Oed – Oedipodiopsida; Tet – Tetraphidopsida; Pol – Polytrichopsida; Bux – Buxbaumiidae; Diph – Diphysciidae; Gig – Gigaspermidae; Fun – Funariidae; Tim – Timmiidae; Dic – Dicranidae; Bry – Bryidae. Number of species in these taxa and number of studied species is provided according to **Supplementary Materials 2**. Colors indicate the frequency of occurrence of these characters in a particular group of species: brown – frequent; yellow – rare; grey – not observed; Buxbaumiidae are not referred to any of these cases as their formulae are indefinite (Ignatov et al., 2018b). Schemes below show the position of anticlinal cell walls in PPL (green) and IPL (blue) at the stage of transition from 2:1 to 2:2 (left) which is critical for the subsequent peristome development, and the resulting arrangement of cells in these two peristomial layers (right) in main lineages of peristome diversification. Frequencies are integral from the variation within transverse section and among species, the latter is shown in **Supplementary Materials 2**.

Funariales, Encalyptales, Disceliales, and Timmiales, which include less than 3% of species of the known moss diversity. The Funariales represent the largest of the latter orders, and the majority of its species have no peristome or a strongly reduced one. Aligned cell divisions are also a characteristic of nematodontous Polytrichales, to some of which *Costesia* appeared to be most similar in having {4}:4:2 pattern at early stages of development.

The presence of the {4}:4:2 pattern in *Costesia* was noteworthy. No other arthrodontous mosses have such regular presence of additional cell divisions in PPL. Species with endostome adherent to the exostome, e.g., *Catoscopium* (Ignatov et al., 2015), *Encalypta* (Ignatov et al., 2018a), and *Splachnum* (Schwartz, 1994) may have such divisions in PPL, but they are rather rare and commence only at the latest stages of development. Therefore *Costesia* is similar in the presence of this pattern to the family Polytrichaceae, which might look unexpected. However there is another moss where this pattern can also be discerned: this is *Lorentziella*, another genus of Gigaspermaceae. Its sporophyte development was carefully studied by Rushing and Snider (1980), and despite their description did not specifically address this pattern, it is clear from the detailed and carefully prepared illustrations. The early stages of sporophyte development in *Lorentziella* are not specific, being common for all bryophytes (Shaw and Anderson, 1988), whereas at the later stages the pattern {4}:4:2 in "peristomial" layers (four cells in PPL, two in IPL) are seen in several octants. Interestingly, this pattern appears simultaneously with the differentiated layer called by Rushing and Snider (1980) as "endothecium 1". It differs in a somewhat darker color and indicates the stage (in time)/or level (in space) where the sporogeneous tissue starts differentiation. It can be assumed that this level and stage correspond to what we see in *Costesia* at the level of annulus.

The presence of a {4}:4:2 pattern in *Lorentziella* additionally ensures us that the same pattern in *Costesia* is not occasional and may be discovered in other genera of Gigaspermaceae. How much it can be interpreted in favor of a relationship to Polytrichaceae is a question without answer: no attempts were made to compare these groups previously, because they are drastically different morphologically. The only similarity that is worth mentioning, is the robust capability to form underground rhizomes in both of these families. Although, this type of growth occurs in a number of groups of subclass Bryidae as well. If this coincidence is not simply occasional, it could be understood only after further studies.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

General plan (MI), field work organization (JL). Field work conducting (JL, MI, EI), plant fixation and embedding (US), anatomy study (MK, US), microscopy (US, MK, MI), manuscript preparation (MI, EI, JL).

FUNDING

Russia Foundation for Basic Research, 19-04-00976 and MBG Institutional research project 18-1180214901111-5.

ACKNOWLEDGMENTS

We are grateful to CONAF for the collecting permits given within Reserva Nacional Lago Peñuelas. We thank Kavita Elliot for improving the English in the original manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.536862/full#supplementary-material

REFERENCES

Bell, D., Lin, Q.-S., Gerelle, W. K., Joya, S., Chang, Y., Taylor, Z. N., et al. (2020). Organellomic data sets confirm a cryptic consensus on (unrooted) land-plant relationships and provide new insights into bryophyte molecular evolution. *Amer. J. Bot.* 107 (1), 91–115. doi: 10.1002/ajb2.1397

Blomquist, H. L., and Robertson, L. L. (1941). The development of the peristome in *Aulacomnium hetersostichum*. *Bull. Torrey Bot. Club* 68, 569–584. doi: 10.2307/2481457

Brotherus, V. F. (1925). "Musci (Laubmoose) 2. Hälfte 11," in *Die natürlichen Pflanzenfamilien, Zweite Auflage*. Eds. H. G. A. Engler and K. Prantl (Berlin: Duncker & Humblot), 1–542.

Budke, J. M., Jones, C. S., and Goffinet, B. (2007). Development of the enigmatic peristome of *Timmia megapolitana* (Timmiaceae; Bryophyta). *Amer. J. Bot.* 94, 460–467. doi: 10.3732/ajb.94.3.460

Dixon, H. N. (1932). "Classification of mosses," in *Manual of Bryology*. Ed. F. Verdoorn (The Hague: Martinus Nijhoff), 397–412.

Edwards, S. R. (1979). "Taxonomic implications of cell patterns in Haplolepidous moss peristomes," in *Bryophyte Systematics*. Eds. G. C. S. Clarke and J. G. Duckett (New York: Academic Press), 315–346.

Edwards, S. R. (1984). "Homologies and inter-relationships of moss peristomes," in *New Manual of Bryology*. Ed. R. M. Schuster (Nichinan: Hattori Botanical Laboratory), 658–695.

Fleischer, M. (1904). *Die Musci der Flora von Buitenzorg, Vol. 1* (Leiden: Brill), 1–379.

Frey, W., and Stech, M. (2009). "Bryophyta (Musci, mosses)," in *Syllabus of plant families A. Engler"s Syllabus der Pflanzenfamilien. Part 3. Bryophytes and seedless vascular plants, 13th edn*. Ed. W. Frey (Stuttgart: Gebr. Borntraeger Verlagsbuchhandlung), 116–257.

Goffinet, B., Buck, W. R., and Shaw, A. J. (2009). "Morphology, anatomy, and classification of the Bryophyta," in *Bryophyte biology, 2nd edn*. Eds. B. Goffinet and A. J. Shaw (Cambridge: Cambridge University Press), 55–138.

Ignatov, M. S., and Ignatova, E. A. (2003). Flora mkhov srednei chasti evropeiskoi Rossii. Tom 1. Sphagnaceae—Hedwigiaceae. Moss flora of the Middle European Russia. Volume 1: Sphagnaceae—Hedwigiaceae. *Arctoa* 11 (Supplement 1), 1–608.

Ignatov, M. S., and Ignatova, E. A. (2004). Flora mkhov srednei chasti evropeiskoi Rossii. Tom 2. Fontinalaceae—Amblystegiaceae. Moss flora of the Middle European Russia. Volume 2: Fontinalaceae—Amblystegiaceae. *Arctoa* 11 (Supplement 2), 609–960.

Ignatov, M. S., Spirina, U. N., Ignatova, E. A., Krug, M., and Quandt, D. (2015). On the systematic position of the moss genus *Catoscopium* with a new approach to the peristome reduction study. *Arctoa* 24, 389–415. doi: 10.15298/arctoa.24.32

Ignatov, M. S., Spirina, U. N., Kolesnikova, M. A., Ashikhmina, D. A., Ignatova, E. A., and Polevova, S. V. (2018a). Peristome development pattern in *Encalypta* poses a problem: what is the primary peristomial layer in mosses? *Arctoa* 27 (1), 1–17. doi: 10.15298/arctoa.27.01

Ignatov, M. S., Spirina, U. N., Kolesnikova, M. A., Ashikhmina, D. A., Ignatova, E. A., and Polevova, S. V. (2018b). *Buxbaumia*: a moss peristome without a peristomial formula. *Arctoa* 27 (2), 172–202. doi: 10.15298/arctoa.27.17

Ligrone, R., and Duckett, J. (2011). Morphology versus molecules in moss phylogeny: new insights (or controversies) from placental and vascular anatomy in *Oedipodium griffithianum*. *Pl. Syst. Evol.* 296 (3–4), 275–282. doi: 10.1007/s00606-011-0496-1

Liu, Y., Budke, J. M., and Goffinet, B. (2012). Phylogenetic inference rejects sporophyte based classification of the Funariaceae (Bryophyta): rapid radiation suggest rampant homoplasy in sporophyte evolution. *Molec. Phylogen. Evol.* 62, 130–145. doi: 10.1016/j.ympev.2011.09.010

Liu, Y., Johnson, M. G., Cox, C. J., Medina, R., Devos, N., Vanderpoorten, A., et al. (2019). Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nat. Commun.* 10, 1–11. doi: 10.1038/s41467-019-09454-w

Mitten, W. (1859). Musci Indiae Orientalis, an enumeration of the mosses of the East Indies. *J. Proc. Linn. Soc. Bot.* 1, 1–96. doi: 10.5962/bhl.title.156377

Newton, A. E., Cox, C. J., Duckett, J. G., Wheeler, J. A., Goffinet, B., Hedderson, T. A. J., et al. (2000). Evolution of the major moss lineages: phylogenetic analyses based on multiple gene sequences and morphology. *Bryologist* 103, 187–211. doi: 10.1639/0007-2745(2000)103[0187:EOTMML]2.0.CO;2

Philibert, H. (1884). De l'importance du péristome pour les affinités naturelles des mousses. *Rev. Bryol.* 11, 49–52.

Rushing, A. E., and Snider, J. A. (1980). Observations on sporophyte development in *Lorentziella imbricata* (Mitt.) Broth. *J. Hattori Bot. Lab.* 47, 35–44.

Schwartz, O. M. (1994). The development of the peristome-forming layers in the Funariaceae. *Int. J. Pl. Sci.* 155, 640–657. doi: 10.1086/297204

Shaw, A. J., and Anderson, L. E. (1988). Peristome development in mosses in relation to systematics and evolution. II. *Tetraphis pellucida* (Tetraphidaceae). *Amer. J. Bot.* 75, 1019–1032. doi: 10.1002/j.1537-2197.1988.tb08809.x

Shaw, A. J., Anderson, L. E., and Mishler, B. D. (1987). Peristome development in mosses in relation to systematics and evolution. I. *Diphyscium foliosum* (Buxbaumiaceae). *Mem. New York Bot. Gard.* 45, 55–70.

Shaw, A. J., Anderson, L. E., and Mishler, B. D. (1989a). Peristome development in mosses in relation to systematics and evolution. III. *Funaria hygrometrica, Bryum pseudocapillare*, and. *B. bicolor. Syst. Bot.* 14, 24–36. doi: 10.2307/2419049

Shaw, A. J., Mishler, B. D., and Anderson, L. E. (1989b). Peristome development in mosses in relation to systematics and evolution. IV. Haplolepideae: Ditrichaceae and Dicranaceae. *Bryologist* 92, 314–325. doi: 10.2307/3243400

Shaw, A. J., Szövényi, P., and Shaw, B. (2011). Bryophyte diversity and evolution: windows into the early evolution of land plants. *Amer. J. Bot.* 98 (3), 352–369. doi: 10.3732/ajb.1000316

Shimamura, M., and Deguchi, H. (2008). "Sporophyte anatomy of *Oedipodium griffithianum (Oedipodiaceae)*", in *Bryology in the New Millennium* Eds. H. Mohamed, B. B. Baki, and A. Nasrulhaq-Boyce (Kuala Lumpur, Malaysia: Institute of Biological Sciences, University of Malaya), 319–325. doi: 10.3732/ajb.1000316

Taylor, E. C. (1962). The Philibert peristome articles. An abridged translation. *Bryologist* 65, 175–212. doi: 10.2307/3241044

Wenderoth, H. (1931). Beiträge zur Kentnis des Sporophyten von *Polytrichum juniperinum* Willdenow. *Planta* 14, 244–385.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 07 April 2021
doi: 10.3389/fpls.2021.517547

Genetic Diversity and Population Structure in Bryophyte With Facultative Nannandry

*Annick S. Lang[1], Thies Gehrmann[2] and Nils Cronberg[1]**

[1] Department of Biology, Lund University, Lund, Sweden, [2] Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands

Among plants, gender dimorphism occurs in about 10% of all angiosperms and more than 50% of all moss taxa, with dwarf males (DM) found exclusively in some unisexual mosses. In this study, we explore the role of male dwarfism as a reproductive strategy in the widespread acrocarpous moss *Dicranum scoparium*, which has facultative male dwarfism, having both dwarf males (DMs) and normal-sized males (NMs). We retrieved 119 SNP markers from transcriptomes which were used to genotype 403 samples from 11 sites at seven localities in southern Sweden. Our aims were to compare the genetic variability and genetic structure of sexually reproducing populations at different geographic levels (cushion, site, and locality) and compare in particular the relative contribution of females, dwarf males and normal-sized males to the observed genetic diversity. The numbers of DMs differed strongly between sites, but when present, they usually outnumbered both females and NMs. Low genetic differentiation was found at locality level. Genetic differentiation was strongest between cushions for females and NMs and within cushions for DMs indicating small scale structuring and sometimes inbreeding. NMs were more clonal than either DMs or females. Genetic diversity was similar between females and DMs, but lower for NMs. Two haplotypes were shared between females and DMs and one haplotype was shared between a DM and a NM. In conclusion, our results show that DMs and NMs play different roles in reproduction, inbreeding may occur at cushion level, but gene flow is high enough to prevent substantial genetic drift.

Keywords: SNP, sexual reproduction, genetic variation, spore dispersal, nannandry, moss (Musci)

OPEN ACCESS

Edited by:
Bernard Goffinet, University of Connecticut, United States

Reviewed by:
Alejandra Moreno-Letelier, National Autonomous University of Mexico, Mexico
Olaf Werner, University of Murcia, Spain

****Correspondence:***
Nils Cronberg
nils.cronberg@biol.lu.se

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *04 December 2019*
Accepted: *08 March 2021*
Published: *07 April 2021*

Citation:
Lang AS, Gehrmann T and Cronberg N (2021) Genetic Diversity and Population Structure in Bryophyte With Facultative Nannandry. Front. Plant Sci. 12:517547. doi: 10.3389/fpls.2021.517547

INTRODUCTION

Unisexuality is usually associated with a certain degree of sexual dimorphism (Badyaev, 2002; Poissant et al., 2010). Evolution of size dimorphism can occur when males and females have distinct roles in reproduction and potentially different requirements, so that they are exposed to contrasting natural and/or sexual selection pressures (Badyaev, 2002; Isaac, 2005). For example, sexual size dimorphism with larger males is generally attributed to combat ability, dominance establishment and access to females for mating (e.g., McElligott et al., 2001; Isaac, 2005) and is

Abbreviations: NMs, normal-sized males; DMs, dwarf males; Sampled localities in Southern Sweden: Lu, Snogeröd Lunnerna; Ro, Röan; Mu, Munkarp; KH, Klöva Haller; SL, Skanörs Ljung; Da, Dalby; Bj, Bjärsjölagård.

mostly found in mammals. Reversed sexual size dimorphism, with larger females, is found in many animals, particularly among invertebrates. It is explained as (1) a response of reduced male-male competition, (2) extreme niche differentiation between genders, or (3) may be the result of the "Ghiselin-Reiss small-male hypothesis" (Blanckenhorn et al., 1995), in which selection favor small males because they require less energy and hence can invest more time in sexual activities. Male dwarfism, a form of extreme size dimorphism where male measure at most half of the size of females, is most common in marine taxa (Fairbairn et al., 2007) and in small organisms, such as invertebrates, fishes, algae, and bryophytes. It might occur when (4) the probability to encounter a sexual partner is low and the need to keep sexual partners nearby is high, as for example in the bone-eating marine worm *Osedax* or anglerfishes (Vollrath, 1998; Pietsch, 2005; Vrijenhoek et al., 2008; Rouse et al., 2015): in these cases males evolved into rudimentary creatures, which are permanently attached to a normal-sized female. In contrast to well-known examples from the animal kingdom, the evolution of sexual dimorphism in plants has been much less studied (Geber et al., 1999; Puixeu et al., 2019), although gender dimorphism occur in approximately 6–10% of all angiosperms (Geber et al., 1999; Vamosi et al., 2003) and in more than half of all moss taxa (Vanderpoorten and Goffinet, 2009; Frey and Kürschner, 2011). Male dwarfism, also called nannandry, has evolved several times independently in unrelated moss families (Vanderpoorten and Goffinet, 2009; Hedenäs and Bisang, 2011) and can either be obligate or facultative, i.e., males growing only as dwarfs or occurring as normal-sized males as well.

Mosses and other bryophytes (liverworts and hornworts) share the trait of having a lifecycle dominated by the haploid gametophyte, the diploid sporophyte generation is initiated after sexual fertilization and remain connected to the maternal shoot throughout its short existence. Most mosses are capable of both asexual and sexual reproduction. Asexual reproduction is common in mosses, allowing clonal regeneration from gametophytic fragments or specialized vegetative diaspores (gemmae). Sexual reproduction depends on water for antherozoids (sperm cells) to swim to the archegonia (Cronberg et al., 2006a; Rydgren et al., 2006). Due to the water dependency, fertilization in unisexual moss species is hampered by drought and physical separation of males and females (Vanderpoorten and Goffinet, 2009; Hedenäs and Bisang, 2011). Asexual reproduction is significant for short-range colonization and maintenance of existing populations and genotypes (Laaka-Lindberg et al., 2003; Glime, 2007). Sexual reproduction is important for recombination during meiosis which takes place in the diploid sporophyte, for adaptability and maintenance of genetic diversity as well as for long-range dispersal through mostly wind-dispersed spores (Cronberg, 2000). A high realized gene flow is often assumed from molecular markers (Patiño and Vanderpoorten, 2018), but attempts to sow spores in natural environment have often failed (Mishler, 1988) and only few studies have been designed to actually reveal genetic variability and genetic population structures at local scales.

Nannandry is unique to mosses among land plants and appears to have evolved as a means to increase the fertilization efficiency by decreasing the separation distance between genders and thus increasing the number of available mating partners for a female (Moya-Laraño et al., 2002; Hedenäs and Bisang, 2011; Rosengren and Cronberg, 2014). Minute males, originating from wind-dispersed spores, grow epiphytically on a much larger female and remain attached during their (bi)annual life-cycle. The growth is reduced to a few leaves, but nevertheless they produce fertile sexual branches (perigonia). Several dwarf males can be found on a single female, usually clustering close to the female sexual branches (perichaetia). Whereas the gender is controlled by sex chromosomes, empirical studies have demonstrated that the growth form of males (dwarfed or normal-sized) could either be genetically determined (in species with obligate dwarf males) or controlled by the supporting female (Glime and Bisang, 2017 and reference therein).

If females in a nannandric species are predominantly fertilized by dwarf-males we can expect that the system is driven toward higher specialization for nannandry (e.g., smaller male spore sizes, with higher dispersal capacity at the expense of lower potential to germinate in the normal substrate) and eventually obligate nannandry. In a species with facultative nannandry, a dwarf male transmits its genes during one season, whereas a normal-sized male can transmit its genes over many years in the same patch and hence could slow down adaptation. However, if normal-sized males and dwarf-males are equally frequently fertilizing the females, the system may stay in a facultatively nannandric condition, without drift toward obligate nannandry. The relative frequency of dwarf males and normal-sized males and their associated haplotypic networks is therefore a key for understanding the development of a nannandric species.

Recent ecological and genetical studies of the nannandric species *Homalothecium lutescens* (Hedw.) H. Rob. have greatly extended our knowledge about nannandry in bryophytes (Rosengren et al., 2013, 2015, 2016). This species is pleurocarpous, having sporophytes on short lateral branches and forming loose and patchy colonies on calcareous ground. Recruitment of dwarf males takes place within colonies, sometimes leading to extreme inbreeding, but the studies also revealed that gene flow occurs between colonies within a population and even from external populations, frequently enough to maintain genetic variability and to prevent substantial genetic drift at population level. The fact that *H. lutescens* is almost completely obligately nannandrous has important implications for its population dynamics; dwarf males are sensitive to environmental factors such as drought, which may cause them to go extinct from time to time in local patches, while the females are more strongly clonal and persistent. This means that females follow a repeated recruitment model (sensu Eriksson, 1989, 1993) with perennial, highly clonal patches, whereas the DMs follow a metapopulation recruitment model (sensu Levins, 1969), having annual, non-persistent individuals with strong fluctuation in numbers between years.

The genus *Dicranum* is unrelated to *Homalothecium*, belonging to the acrocarpous moss groups, characterized by apical sporophytes on erect sparsely branched shoots and

growth in dense tufts, radially expanding through vegetative proliferation. About half of the 131 *Dicranum* species reproduce only asexually, amongst the sexually reproducing species, 21 lack dwarf males and 25 are reported as nannandric, only two of which, *Dicranum scoparium* Hedw. and *Dicranum bonjeanii* De Not., have facultative dwarfism (Pichonet and Gradstein, 2012). In this study we focus on the widespread species *D. scoparium* sampled from seven localities, using 90 single-nucleotide polymorphic markers (SNPs). Our overall aims are to employ the molecular markers in a moss with facultative nannandry in order to (1) reveal the hierarchical genetic population structure (2) determine if the levels of genetic diversity and clonality compare between females and males and (3) assess the relative shares of dwarf males and normal-sized males to the total standing genetic diversity.

MATERIALS AND METHODS

Study Species

Dicranum scoparium is widely distributed in the Holarctic and is growing on various substrates, preferably on acidic soil, rotten logs, tree stems and sometimes in sympatry with other *Dicranum* species (Crum and Anderson, 1981; Hedenäs and Bisang, 2004; Ireland, 2007; Lang et al., 2014). Both normal-sized and dwarf males can occur simultaneously in a same cushion. Dwarf males are annual and are easily found in the late summer on the tomentum of a female stem that carries a young sporophyte. They are supposedly non-clonal, whereas the perennial normal-sized males are clonal, resembling the females and, if present, found in the external part of a cushion.

Sample Collection

Fresh bryophyte material was collected in August 2015 across Skåne, South Sweden. In total 54 fertile cushions (including two cushions of *D. bonjeanii* and five of *Dicranum majus*) from 11 sites were collected from seven different localities, namely Snogeröd Lunnerna (Lu), Röan (Ro), Munkarp (Mu), Klöva Haller (KH), Skanörs Ljung (SL), Dalby (Da), Bjärsjölagård (Bj; **Figure 1**). In each of these localities, between one and three sites (if possible) were sampled. Two sites were considered distinct if they were separated from at least 2 km and by a river, road or ravine. Sexually reproducing populations are easily recognized in the field due to the presence of sporophytes and are generally associated with the presence of dwarf males and/or normal-sized males (Pichonet and Gradstein, 2012, pers. observ.). To perform genetic diversity analyses, we collected about 3 cm^2 of *Dicranum* moss from sexually reproducing cushion. From each of these cushions, five fertile female stems were sampled. A stem can have several branches. Each of these branches were extracted separately as they also carried distinct sporophytes. Therefore, the number of female samples genotyped and then analyzed is bigger than five. All dwarf males found of on theses female stems were separated for *in vitro* growth and all normal-sized males within each of these cushions were sampled. The collected material was kept alive in the experimental garden of Lund University for morphological identification, resampling (if necessary), and further observation of vegetative growth.

In vitro Culture of Dwarf Males

In order to get sufficient amounts of DNA or RNA from a single dwarf male individual we induced vegetative growth by separating them from the female and placing them onto a nutrient agar medium. Here, we used a medium as described in Rudolph et al. (1988). Each male was carefully removed from the female tomentum using a dissecting microscope and thoroughly inspected for contaminants before plating. The number of males found in each locality varied greatly, from only three males found in one cushion from Lunnerna to 384 found among the cushions from Bjärsjölagård. In order to remove bacterial, fungal or algal contamination, clean parts of the newly grown gametophyte were replated in new petri dishes regularly during a year after cleaning treatment. This treatment involved soaking the gametophyte in a solution of 0.05% NADCC at constant shaking for 1 min followed by two rinses, each 1 min in demineralized water. Removal of contamination remained the most difficult problem because males did not always survive the replating process. For example, 92 out of 526 dwarf males from Bj and only 2 out of 51 DM from Da survived. Hence, out of more than 500 dwarf males initially cultured, only 140 dwarf males were subject to DNA extraction (**Supplementary Appendix 1**).

DNA Extraction

Total DNA from 769 samples including females, sporophytes, dwarf, and normal-sized males was extracted. Prior to extraction, stems of females, and normal-sized males were cleaned in demineralized water. After removing the surplus water with a tissue, all stems, sporophytes were placed in separate microtubes and kept in a −80°C freezer until extraction. After cultivation, dwarf males and associated protonemas were removed from the agar plates and dried in silica gel until extraction. DNA extractions were done using DNeasy Plant mini kit (Qiagen, Hilden, Germany) according to the protocol of the manufacturer and eluted in 100 µl AE buffer. All DNA extracts were stored at −20°C.

RNA Extraction

We applied a strategy involving transcriptome analyses for selection of SNP markers, which requires RNA extracts of high quality and quantity. Fresh plant material was used for RNA extraction and one female stem per locality was randomly sampled in order to represent the genetic variability of the population. Three additional extraction was done for pooled normal-sized males, sporophytes (diploid generation) and one *Dicranum majus* Turner female gametophyte. Each stem was first cleaned in deionized water and then flash frozen in liquid nitrogen before being stored at −80°C. Consequently, RNA from ten samples were obtained using the RNeasy Plant mini kit following the protocol of the manufacturer (Qiagen, Hilden, Germany). The quality, quantity and integrity of the RNA extractions were measured on agarose gels, with Qubit® 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA,

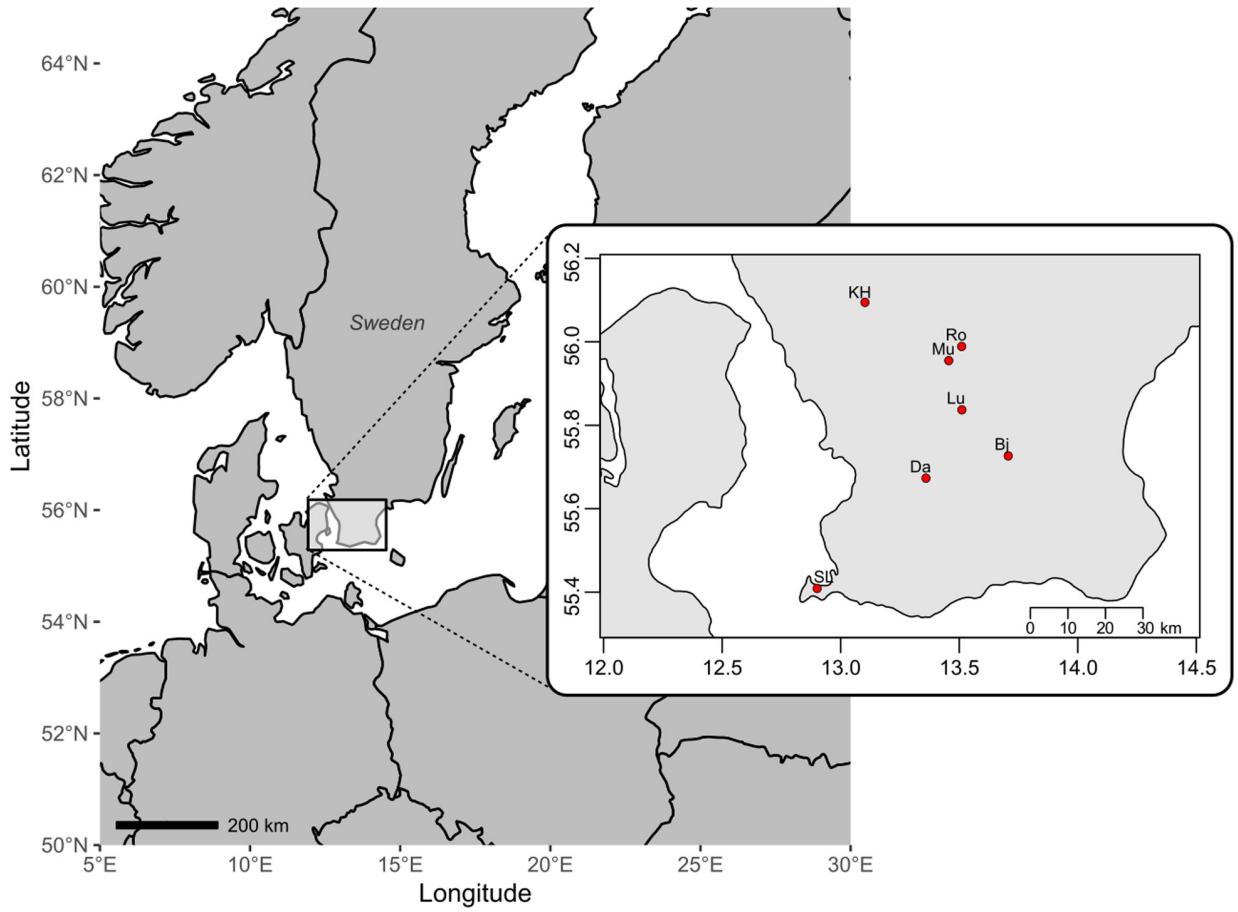


FIGURE 1 | Map of southern Sweden with seven localities where populations of the moss *Dicranum scoparium* have been sampled: Bj, Bjärsjölagård; Da, Dalby; KH, Klöva Haller; Lu, Snogeröd Lunnerna; Mu, Munkarp; Ro, Röan; SL, Skanörs Ljung.

United States) and BioAnalyzer (Agilent Technologies, Palo Alto, CA, United States), respectively.

SNP Localization, Typification and Genotyping

The SNP library was produced with the MiSeq system (Illumina, United States) using the 300 bp paired-end settings. The library construction was carried out at Lund University sequencing facility[1] using NEXTflex RNA-Seq kit (Bioo Scientific, Austin, TX, United States).

The sequencing of the moss transcriptomes resulted in about 1 million reads/sample (**Supplementary Appendix 2**). The sample DaF was used to create a *De Novo* reference assembly using Trinity (Grabherr et al., 2011; Haas et al., 2013). Reads from other samples were realigned using BOWTIE (Langmead et al., 2009) with –m option, allowing no multialigned reads in order to reduce as much as possible false positive SNP calling. We use the samtools (Li et al., 2009) mpileup tool to produce VCF files with all single nucleotide variants in the aligned BAM files. In total, nearly 25,000 variants were found. SNP filtering was done using scripts specifically designed for this purpose[2]. To remove variants called due to indel alignments, we removed SNPs within 100 bp of each other. To further support the authenticity of the variants, we selected only variants which were identified in at least three individuals. This also removes rare variants which are not useful in a genetic diversity screening (most variants are rare variants). Finally, to ensure the variants were still adequate markers of genetic diversity, we removed variants which were present in more than 8 (out of 10) samples. This resulted in a final selection of 119 SNPs.

The genotyping of 738 DNA samples was carried out at the SNP&SEQ Technology Platform, Department of Medical Sciences of Uppsala University. It was performed using a multiplexed primer extension chemistry of the iPLEX assay (Ross et al., 1998; Storm et al., 2003; Gabriel et al., 2009). The allele was then detected by mass spectrometry with a MassARRAY analyzer (Agena Bioscience, Hamburg, Germany). Raw data was converted to genotype data using Typer software (Agena Bioscience).

[1]https://www.biology.lu.se/services/dna-sequencing-facility/next-generation-sequencing-on-illumina-miseq

[2]https://github.com/thiesgehrmann/annickMossVarCall

SNP Data Analyses

The analyses were performed on *D. scoparium* individuals only, excluding sporophytes data. The final number of individuals analyzed in this study is found in **Table 1**. In order to analyze female and male genotypic diversity, or haplotypic diversity in this case, as well as levels of clonality within populations, we selected all gametophytic data from the raw SNP dataset (GAM). Then, because genet recruitment of female, normal-sized males, and nannandrous males was expected to follow different recruitment models, two subsets were created according to the sex of the samples, i.e., one subset containing female data (FEM) and one with male data (males). Additionally, to estimate the genetic differentiation within the two male types, dwarf males (DMs) and normal-sized males (NMs) were also analyzed separately. All analyses were performed in R (R Development Core Team, 2013) and genetic diversity metrics as well as levels of clonality within each population were characterized using the packages *adegenet* version 2.1.2 (Jombart, 2008; Jombart and Ahmed, 2011), *poppr* v. 2.8.5 (Kamvar et al., 2014, 2015), and *ade4 v. 1.7-15* (Chessel et al., 2004; Dray and Dufour, 2007; Dray et al., 2007; Bougeard and Dray, 2018). The analyses were performed with population hierarchy that included two levels, i.e., locality and cushion. Implementation and interpretation of *poppr* outputs were adapted from http://grunwaldlab.github.io/Population_Genetics_ in_R/ (Grünwald et al., 2017).

Firstly, failed genotyped individuals and SNPs were excluded from the raw data set. Then, the different subsets were created. Individuals and SNPs containing more than 10% missing information, invariable and uninformative loci (minimum allele frequency of 0.01) independently for each subset were excluded. Finally, we controlled again that the data set would contain no more than 10% missing information and eventually remove the individual or SNP that would not meet the criteria.

To remove the effect of genetic linkage due to clones and adjust for missing data and genotyping errors that may occur during sequencing large datasets (c.f. Kamvar et al., 2015), multilocus lineages (MLL) were calculated with the *mlg.filter* function by collapsing unique MLH (multilocus haplotypes) utilizing Nei genetic distance given the farthest neighbor clustering algorithm. This method allows to incorporate genotypes that have missing data or genotyping error into a parent clusters. The clustering threshold was set for each dataset individually and missing data were replaced with the mean of the alleles for the entire data set.

Multilocus lineages were further used for the estimation of the haplotypic diversity (G; Stoddart and Taylor, 1988) and Nei's gene diversity, also called unbiased gene diversity ($\hat{h}$; cf. Grünwald and Hoheisel, 2006). Furthermore, the relative contribution of clonality within a population was assessed by using the index of association (IA) and its standardized form, $\bar{r}_d$, which accounts for the number of loci sampled (Agapow and Burt, 2001). Considering that IA is based on the variance of pairwise distances between MLLs, IA should be high in clonal populations and low in populations undergoing sexual reproduction (Grünwald and Hoheisel, 2006). A significance test with 1000 permutations was run to assess for significance. Hierarchical structure of genetic variation was explored by running an analysis of molecular variance [AMOVA (Excoffier et al., 1992)] considering cushion as subpopulations within localities. The analysis was conducted using the *poppr.amova* function and the *ade4* implementation. Fixation indices (Φ statistics) were interpreted as a measure of subpopulation differentiation. The probability of retrieving the observed Φ under the null hypothesis of no differentiation between subpopulations was calculated using the *randtest* function of the *ade4* package with 1000 iterations of the permutation test.

Assuming that the diversity of DM haplotypes residing on a female is determined by the number male spores that germinated on the female, the proportion of DMs should covary with haplotypic diversity in a population. Therefore, the proportion of each sex type at locality was plotted against the haplotypic diversity G. Kruskal Wallis non-parametric test was used to examine the differences in G between sex types. Furthermore, an ANOVA was used to test for the differences of $\hat{h}$ between sex types as residuals followed a normal distribution. We also compared Nei's gene diversity $\hat{h}$ against G in order to display the contribution of each haplotype to the total variability at cushion level.

Finally, population structure was analyzed in two ways. First, a principal coordinate analysis (PCoA) was performed on female and male individuals. The analysis was based on a pairwise genetic distance matrix, with Prevosti's distance, equivalent to dissimilarity distance. Then, a minimum-length spanning network (msn) of all unique MLL was constructed based on the Prevosti's distance in order to assess how female and male haplotypes were related across different populations. Resulting graphs were edited with Inkscape v.1.0.0.

TABLE 1 | Summary of the *D. scoparium* dataset showing sample size, number of haplotypes and codominant loci for each data subset after passing all filtering criteria.

Dataset	*n*	Loci	MLH	MLL
GAM	403	68	258	148
FEM	211	64	107	61
males	192	74	148	90
DM	134	74	113	72
NM	57	61	25	14

Ten percentage or more missing data as well as invariable and uninformative loci were excluded for each data set independently. n, number of individuals; Loci, number loci; MLH, number of multilocus haplotype; MLL, number of multilocus lineages obtained after collapsing of unique MLH. Numbers are given for the total dataset (GAM) and the female (FEM), male (males) dwarf males (DM), and normal sized (NM) subsets.

RESULTS

SNP Detection in RNA and DNA Genotypification

From the 10 initial samples, 119 SNPs were identified for genotype screening (see section "Materials and Methods"). Out of 769 extracted DNA samples, 738 were suitable for genotyping, whereas 31 lacked enough DNA for genotyping. Of the 119

SNPs, 90 SNP markers were identified in all 738 genotyped DNA samples. The fraction of SNP markers with call rate >0 reached 75.63% and the average sample call rate per SNP reached 88.49%. The fraction of individuals with call rate >0 reached 95.80% while the average SNP call rate per sample reached 69.86%. Five out of 90 SNPs were invariable (5.56%). The 29 deficient SNPs had no allele signal cluster separation (sample call rate of 0%) which resulted in inaccurate genotype calls.

Higher Genetic Diversity Among Females and Dwarf Males Than Among Normal-Sized Males

After exclusion of individuals and SNPs with 10% or more missing data and invariable loci, the total *D. scoparium* haploid dataset contained 403 individuals and 68 loci (**Table 1**). The proportions of males (NMs and DMs) and females reflect the frequencies at each site. Five sites had both NMs and DMs, three had only NMs and two only DMs (**Table 2**). The number of DMs differed strongly between sites, and when present, they usually outnumbered both females and NMs, and at one site (Bj), they were particularly abundant (**Table 2**). The female dataset was composed of 211 individuals and 64 loci, representing 61 MLLs. The male dataset contained 192 individuals and 74 loci, representing 90 MLLs (**Table 1**).

Multilocus haplotypic diversity (G) differed between localities, from G = 5.45 in Dalby to G = 18.39 in Bjärsjölagård and the underlying proportions of female, NM and DM haplotypes were profoundly disparate. The NMs had significantly lower haplotypic diversity (mean G = 1.8) than both females (mean G = 6.9) and DMs (mean G = 8.2) (Kruskal–Wallis; H = 6.8; df = 2; P = 0.03). Furthermore, the haplotypic diversity for DMs varied a lot dependent on the number of individuals that were possible to genotype (mean G = 8.2; **Table 3**). Although the different sample sizes explain some of the variation in G, **Figure 2A** illustrates that the proportion of females or NMs in a population had low influence on haplotypic diversity and that only the proportion of DMs correlated with haplotypic diversity (R^2 = 0.98), although this is largely due to two localities with abundant population of dwarf males (Bj and Ro).

TABLE 2 | Number of *D. scoparium* individuals in each data subset used for the analysis after passing all filtering criteria.

Locality	Site	Cushion	FEM	DM	NM
Bj	1	3	21 (2)	88 (43)	0
Da	1	4	19 (5)	2 (2)	2 (1)
KH	1	4	18 (8)	0	10 (2)
KH	2	6	15 (6)	0	1 (1)
Lu	1	3	7 (4)	3 (1)	2 (1)
Lu	2	1	5 (1)	0	0
Lu	3	2	9 (6)	0	2 (2)
Mu	1	8	30 (9)	4 (3)	8 (3)
Ro	1	5	23 (6)	25 (17)	8 (2)
Ro	2	5	18 (6)	6 (2)	0
SL	1	6	46 (9)	6 (5)	24 (3)
Total	16	47	211 (61)	134 (72)	57 (15)

The localities are Bj, Bjärsjölagård; Da, Dalby; KH, Klöva Haller; Lu, Snogeröd Lunnerna; Mu, Munkarp; Ro, Röan; SL, Skanörs Ljung. Numbers in parenthesis are the number of unique multilocus lineage, corresponding to the number of individuals with unique haplotype. Site, site number within the locality; Cushion, number of cushions within each site; FEM, number of females; DM, number of dwarf males; NM, number of normal-sized males.

TABLE 3 | Genetic diversity at locality level for the combined dataset (GAM), female (FEM), dwarf male (DM), normal-sized male (NM), and all male (males) data subsets.

Locality		*n*	MLH	MLL	G	$\hat{h}$	IA	$\bar{r}_d$
Bj	GAM	110	80	49	18.39	0.19	8.44	0.18
	FEM	21	8	2	1.57	0.16	26.19	0.97
	males	89	74	49	32.07	0.19	6.8	0.13
	DM	88	76	43	26.34	0.2	7.61	0.14
	NM	0						
Da	GAM	23	15	8	5.45	0.31	6.78	0.14
	FEM	19	9	5	3.97	0.29	8.88	0.21
	males	4	4	3	2.67	0.31	8.77	0.22
	DM	2	2	2	2	0.4	NA	NA
	NM	2	2	1	1	0	NA	NA
KH	GAM	44	30	17	11.13	0.33	2.71	0.04
	FEM	33	18	14	10.37	0.33	2.46	0.04
	males	11	7	3	1.75	0.17	20.14	0.52
	DM	0						
	NM	11	5	3	1.75	0.19	19.4	0.53
Lu	GAM	28	17	13	7.84	0.34	7.03	0.13
	FEM	21	11	10	6.04	0.34	7.59	0.15
	males	7	5	3	1.81	0.21	22.65	0.6
	DM	3	2	1	1	0	NA	NA
	NM	4	4	3	2.67	0.35	11.57	0.33
Mu	GAM	42	30	14	10.63	0.32	3.4	0.06
	FEM	30	21	9	8.18	0.31	3.45	0.07
	males	12	9	5	2.57	0.2	14.12	0.29
	DM	4	4	3	2.67	0.28	17.47	0.47
	NM	8	3	3	1.68	0.09	19.16	0.96
Ro	GAM	80	51	32	19.05	0.36	2.68	0.05
	FEM	41	24	12	10.44	0.35	3.1	0.06
	males	39	28	21	12.17	0.32	4.27	0.07
	DM	31	23	19	12.81	0.32	5.21	0.09
	NM	8	3	2	1.28	0.08	18.82	0.99
SL	GAM	76	35	16	9.66	0.35	4.56	0.07
	FEM	46	16	9	7.45	0.37	5.43	0.09
	males	30	21	7	2.87	0.24	14.06	0.3
	DM	6	6	5	4.5	0.24	9.38	0.24
	NM	24	8	3	2.27	0.18	21.13	0.92
Average	GAM				11.74	0.31		
	FEM				6.86	0.31		
	males				7.99	0.23		
	DM				7.05	0.24		
	NM				1.68	0.15		

n, sample size; MLH, Multilocus haplotype; MLL, Multilocus lineages; G, Haplotypic diversity based on Stoddard and Taylor's index; $\hat{h}$, Nei's unbiased gene diversity; IA, index of association; $\bar{r}_d$, Standardized IA. All $\hat{h}$ and IA values were significant ($P < 0.001$), based on 1,000 permutations.

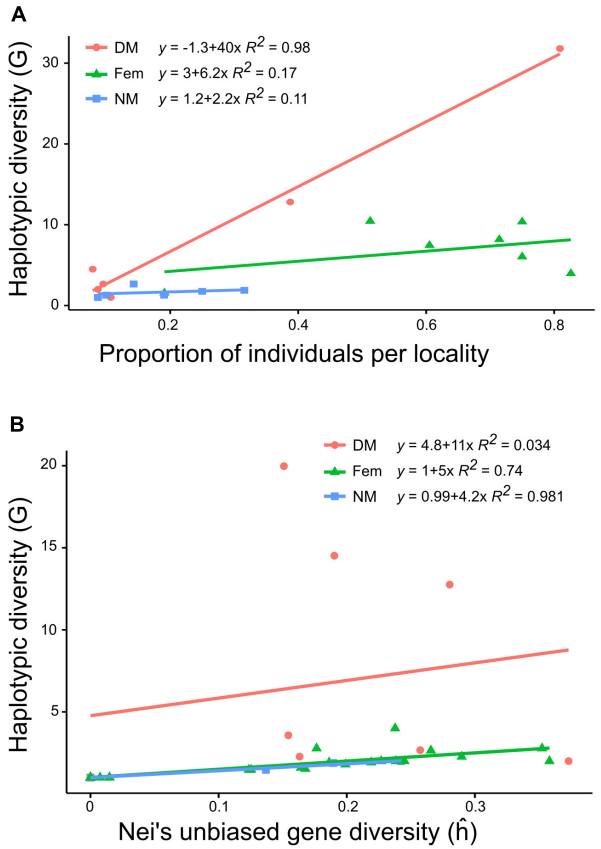


FIGURE 2 | Scatter plot with regression lines and R^2 showing **(A)** the proportion of individuals at locality level against the genotypic diversity (G) and **(B)** the expected gene diversity ($\hat{h}$) against G for each sex type (dwarf males, females, and normal-sized males). Colors of regression line and point represent sex types (red dot, dwarf males; Green triangle, females; Blue square, normal-sized males).

Overall, levels of gene diversity $\hat{h}$ were slightly higher for females (mean $\hat{h}$ = 0.31) than males (mean $\hat{h}$ = 0.23) (ANOVA; $F = 4.83$; $P = 0.048$). The DMs (mean $\hat{h}$ = 0.24) were usually more variable than the NMs (mean $\hat{h}$ = 0.15) and NMs had significantly lower gene diversity than both females and DMs (ANOVA; $F = 4.22$; $P = 0.03$). In most localities NMs had $\hat{h}$ values ranging between 0 and 0.19, but one exception was found at Lu, with $\hat{h}$ = 0.34 for both females and NMs (**Table 3**). At cushion level, which is the scale relevant for fertilization, gene diversity was strongly positively correlated with haplotypic diversity, except at patches with numerous DMs (**Figure 2B**). This means that each normal-sized haplotype contributes about equally much to the total genetic variability at cushion level. The cushion level estimate of $\hat{h}$ is equal or higher for DMs as compared to NMs, but the variation is potentially distributed among many more haplotypes for DMs.

The degree of clonality was measured with the index of association (IA). Clonality was strongest for NMs at all localities, with high IA values and $\bar{r}_d$ values ranging between 0.32 and 1, suggesting high clonality. As expected, DMs had $\bar{r}_d$ values close to 0, which indicate low clonality but somewhat surprisingly, female IA and $\bar{r}_d$ values were similar to those obtained for DMs. The populations in Bj showed an opposite pattern: levels of haplotypic diversity were very low for females and very high for males, although levels of gene diversity were comparable. Furthermore, IA and $\bar{r}_d$ values indicated high clonality of females at this locality.

Population Structure

Almost 90% of the total variation was partitioned nearly equally within cushions and among cushions within localities, while only 11% of the variation occurred among localities (**Table 4**: GAM). 25% of the molecular variance was associated with a differentiation within female cushions (haplotypes among localities) and 68% with differentiation among cushions within localities (haplotypes among cushions within localities). Males also displayed a high relative differentiation at cushion level (DM = 47.1% and NM = 30%, respectively). The highest fixation index was within cushions in both females and NMs (FEM: $\Phi = 0.75$; NM: $\Phi = 0.7$).

A PCoA was produced to visualize the genetic distance of female and male individuals (**Figure 3**). Overall, the low differentiation among the female samples at locality levels observed in the AMOVA analyses was reflected in the PCoAs. The first three axes explained 16.25% of cumulative variation (PC1: 7.12%, PC2: 4.76%, and PC3: 4.37%). No clear subclustering was displayed, only few individuals from Bj were forming a separate cluster, best separated by PC1 and PC3. However, looking at locality with more than one site individually (Lu, Ro), some fine structure could be seen. In population Lu (black shapes), site 1 and 3 were forming two clusters with certain degree of overlapping and site 2 was more distant from the two former. Ro had two sites (red empty shapes) which were separated by PC1 and PC2. Although SL had only one site, the individuals were forming two clusters separated by PC2.

The first three axes of the male PCoA explained 28.13% cumulative variance (PC1: 15.13%, PC2: 7.32%, and PC3: 5.68%). Similar to the females, the male PCoA displayed little structure at locality level (**Figure 4**), except for Bj. Most individuals (all dwarf males) of this locality were separated by PC1 from all other individuals, all localities included. Whereas no differentiation could be revealed at site level, more structure was observed at the sex level, i.e., between normal-sized males (NM, big shapes) and dwarf males (DM, small shapes). NMs from a certain site tended to cluster more closely, being more similar to each other than DMs. PC2 segregated strongly the individuals in SL 1 into three clusters, two containing only NMs and one containing only DMs (purple triangles). Furthermore, individuals from Lu 1 (black triangles) were strongly separated from the individuals from Lu 3. Finally, PC3 tended to segregate individuals from KH 1 and 2, Lu 1 and 3 as well as DMs and NMs from Ro 1.

To visualize relationships between female and male MLLs, a minimum-spanning network was constructed (**Figure 5**). The number of individuals per MLL (or haplotype) ranged between 1 and 16. Sixty-one female MLLs, 13 NM MLLs, and 77 DM MLLs were having more than one individual of which 7, 1, and 2 MLLs for female, NM and DM, respectively, where shared between

TABLE 4 | Analysis of molecular variance (AMOVA).

Dataset	Hierarchical level	Df	Sum of squares	Mean squares	Variation (%)	Φ	*P*
GAM	Φ within cushion	356	3886.05	10.92	42.59	0.57	*0.001*
	Φ among cushion within locality	40	3837.14	95.93	45.92	0.52	*0.001*
	Φ among locality	6	2110.06	351.68	11.48	0.11	*0.001*
FEM	Φ within cushion	164	1022.13	6.23	25.3	0.75	*0.001*
	Φ among cushion within locality	40	3130.33	78.26	68.27	0.73	*0.001*
	Φ among locality	6	890.44	148.41	6.43	0.06	*0.001*
DM	Φ within cushion	123	1637.63	13.31	47.14	0.53	*0.001*
	Φ among cushion within locality	5	386.62	77.32	19.8	0.3	*0.001*
	Φ among locality	5	950.75	190.15	33.06	0.33	*0.040*
NM	Φ within cushion	46	338.72	7.36	30.01	0.7	*0.001*
	Φ among cushion within locality	5	145.02	29	50.01	0.62	*0.003*
	Φ among locality	5	667.47	133.49	19.98	0.2	*0.004*

AMOVA results for combined data (GAM), female (FEM), dwarf males (DM), and normal-sized males (NM). P values were obtained for the Φ statistic with 1000 permutations.

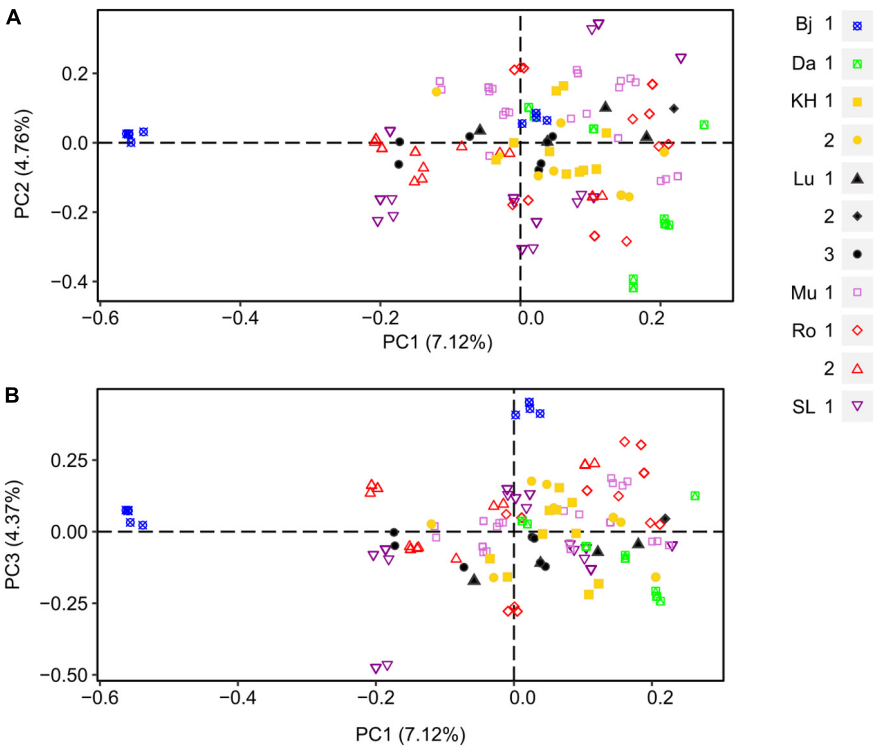


FIGURE 3 | Principal Coordinate Analysis (PCoA) of female populations of *D. scoparium* from southern Sweden based on a genetic distance matrix. Individuals are represented by a symbol according to the locality (Bj, blue; Da, green; KH, yellow; Lu, black; Mu, light purple; Ro, red; SL, purple) and the collection site (number 1.2 or 3). The axis 1 and 2 are shown in scatter plot **(A)** and axis 1 and 3 in scatter plot **(B)**.

individual of different cushions. Additionally, two genotypes were shared between females and DMs, one genotype was shared between DM and NM and one genotype was shared between DM across localities (arrows in **Figure 5**). This last genotype was found in two individuals of population Bj and one of Mu. Overall, there was no clear differentiation of the populations, apart from population Bj. In this populations, DM genotypes were much closer to each other than to genotypes of other populations. Furthermore, one of the two female genotypes was shared among 20 individuals, sampled in two different cushions and both sexes.

DISCUSSION

In this study, we examined the genetic variability and structure of female and male populations of the facultatively nannandric moss *Dicranum scoparium* at different scales. Dwarf males and

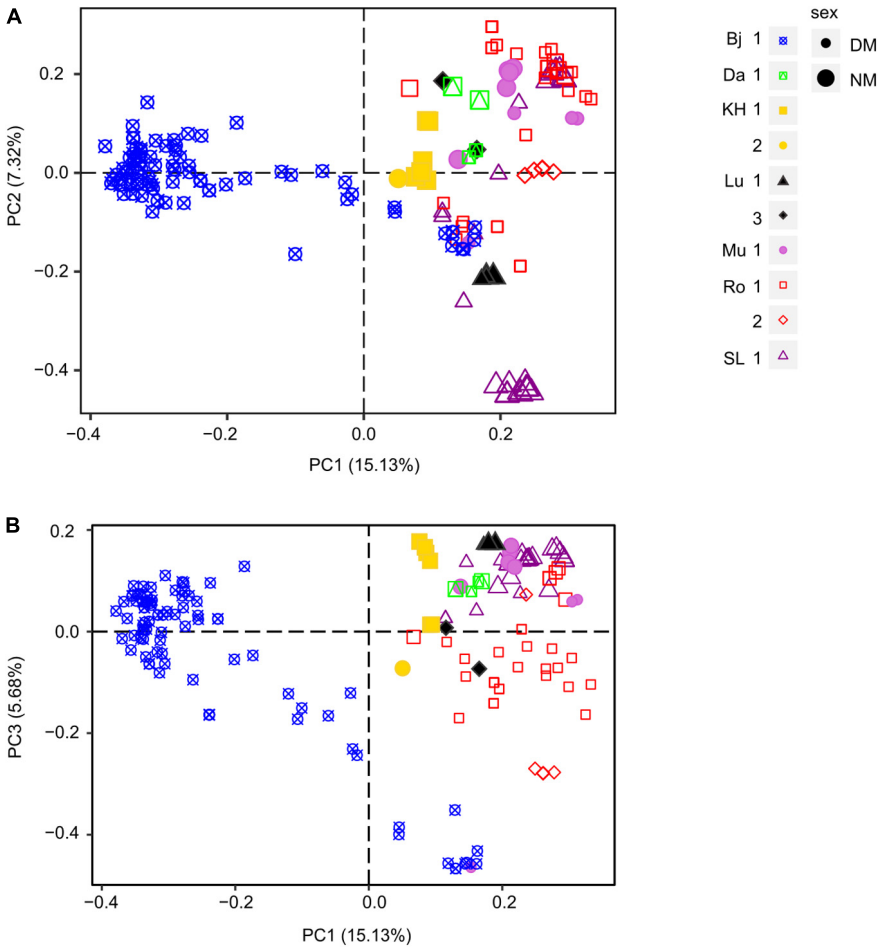


FIGURE 4 | Principal Coordinate Analysis (PCoA) of male populations of *D. scoparium* from southern Sweden based on a genetic distance matrix. Individuals are represented by a symbol (large, NM; small, DM) according to the locality (Bj, blue; Da, green; KH, yellow; Lu, black; Mu, light purple; Ro, red; SL, purple) and the collection site (number 1.2 or 3). The axis 1 and 2 are shown in scatter plot **(A)** and axis 1 and 3 in scatter plot **(B)**.

normal-sized males often occurred together in a moss cushion. We showed that the overall levels of genetic diversity within females and dwarf males were relatively similar and much higher than among normal-sized males. The lower genetic diversity in NMs may therefore be a consequence of fewer haplotypes in the populations. Furthermore, clonality occurred in all three categories, yet at different degrees: level of clonality was high among NMs whereas it was much less so in females and DMs. Most of the genetic differentiation was partitioned within and among cushions and only a low fraction among localities. These results suggest that effective gene flow occurs regularly over long distances.

Experimental Design

Before we interpret the genetic diversity data, it is worth mentioning that (1) we have actively searched for loci that are variable and (2) that the regions are transcribed and therefore the SNP alleles are possibly under selection. This implies that absolute values of genetic diversity measures are not directly comparable with other studies, although our markers in several ways (representing expressed variation, point mutations) resemble allozyme markers, for which much of the population level studies were done in the past.

No reference genome for *D. scoparium* exists. Thus, designing a genetic diversity panel is challenging. Our approach, using RNA-seq and transcriptome assembly enables an affordable construction of transcribed portions of the genome that are likely evolutionarily relevant. Yet, the quality of such an assembly is highly dependent on the depth of sequencing, which is highly variable per transcript, and especially near the end of transcript. Furthermore, SNPs were identified using RNA-Seq reads aligned to the assembled transcriptome. This too is restricting and resulted in only 119 reliable SNPs. Then, due to the cost of RNA-Sequencing, we used only 10 samples (one per locality) that would represent the total genetic diversity of each population. Hence, we selected variants that were relatively common and not entirely recurrent (in at least three and at most eight of the 10 samples). Furthermore, the extraction and genotyping steps themselves also lead to some losses. For example, DM for the populations in Lu were extracted but the genotyping was not successful. Finally,

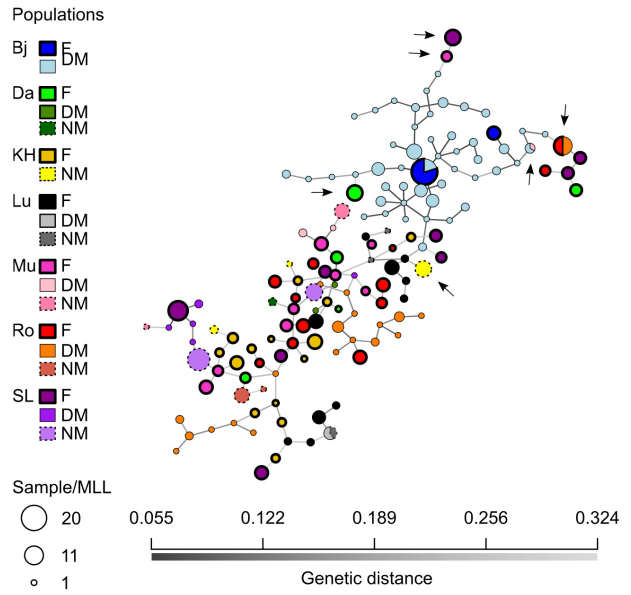


FIGURE 5 | Minimum-spanning network (MSN) of the *Dicranum scoparium* SNPs. Each node represent a unique multilocus lineage (or haplotype in our case). Colors represent sexes per localities and edges represent absolute genetic distance (Bj, blue; Da, green; KH, yellow; Lu, black; Mu, pink; Ro, red; SL, purple; F, female; DM, dwarf-male and NM, normal-sized male).

SNP data pose new challenges when compared to other markers, because they potentially contain more missing data, error in SNP calling due to genotyping errors or other sources of inaccurate allele calls (Mastretta-Yanes et al., 2015). Having this in mind, we note that our data set produced clear results with respect to clonal diversity, gene diversity and genetic population structures for the whole material and when separated into NMs, DMs and females.

Genetic Diversity and Population Structure

Genetic diversity of females and all males taken together were similar, but NMs tended to be genetically less diverse relative to DMs, while females and DMs had comparable levels of genetic diversity. The PCoA, MSN, and AMOVA analyses showed weak genetic differentiation at larger scale, except for the locality Björsjölagård (Bj). At locality levels genetic differentiation for females was particularly low (Φ = 0.06), which is indicative of a high gene flow. The corresponding values for *H. lutescens* at locality level was around 0.10 (Rosengren et al., 2013) and 0.3 (Rosengren et al., 2015) from different regions in S Sweden (Öland and Scania, respectively), suggesting a somewhat higher differentiation among those populations even though the sampling of *D. scoparium* spanned a larger geographic area. The difference between these two nannandric species could potentially be explained by the degree of isolation between populations and/or the fact that perennial, normal-sized males are much rarer in *H. lutescens*, increasing the risk for inbreeding due to son-mother fertilization with DMs.

A large fraction of the total variation in our study partitioned within cushions and among cushions within sites, suggesting that cushions in general have several genets which are notably dissimilar. This is particularly valid among DMs, where the fraction of the total variation residing within cushions reached 47%. Our findings indicate that gene flow occurs between localities and that the relatively high levels of genetic diversity, both within females and males, are maintained through spore dispersal rather than vegetative fragment dispersal. The structure found in *Dicranum* populations are consistent with other moss studies that indicated limited population structure at larger scales but differentiation at cushion level (Cronberg, 2002; McDaniel and Shaw, 2005; Hutsemékers et al., 2010; Paasch et al., 2015; Désamoré et al., 2016). DMs and NMs often occurred simultaneously but the two types of males are obviously under different constraints and thereby contribute differently to the population dynamics. When present, dwarf males are often numerous and display a level of genetic diversity similar to the one of females. The proportion of DMs covaries positively with the haplotypic diversity of a population (at locality level), which is not the case for females and NMs. The vicinity of DMs to the female perichaetia is likely to be a great advantage in competition with NMs for fertilization, but NMs may be important for fertilization when conditions are unfavorable for DMs, as they are perennial and probably less sensitive to temporarily adverse climatic conditions. In this context is worth to notice that, if anything, the haplotypic and genetic diversity of DMs may be underestimated since a high number of the DMs were lost before DNA extraction.

The level of gene diversity found in *D. scoparium* is comparable to the dwarf males and females of the nannandric species *H. lutescens* (Rosengren et al., 2013, 2016). Although males were most likely dispersed locally, the authors showed that a fraction of males originated from more distant populations, hence maintaining high levels of genotypic diversity within cushions and across localities. In our study, we found NMs growing on soil, at the border of cushions, with generally much lower haplotypic diversity than DMs and females. This compares with results obtained for the non-nannandric species *Syntrichia caninervis* (Baughman et al., 2017), whose males displayed a lower genetic diversity than females. In both *H. lutescens* and *S. caninervis*, the differences in diversity within the male and female populations can be attributed to local environmental conditions such as moisture and temperature, which may affect spore dispersal and germination success (McDaniel et al., 2007; Rosengren and Cronberg, 2014; Rosengren et al., 2016; Baughman et al., 2017). Whereas several studies have demonstrated the positive association between moisture and presence of males (Tamm, 1964; Bowker et al., 2000; Sagmo Solli et al., 2000; Rosengren and Cronberg, 2014), the environmental constraints on NMs are largely unknown.

Finally, the haplotype networks provided some evidence of inbreeding events occurring in populations, especially at Ro and Bj. Whereas extreme inbreeding indicative of repeated son-mother fertilizations was fairly frequently found in *H. lutescens* (Rosengren et al., 2016), only two haplotypes were shared between females and DMs in our study (**Figure 5**). However,

the DM haplotypes were sometimes clustering tightly together with the female haplotype they were growing on as seen in the network analysis (**Figure 4**). Therefore, repeated son-mother fertilizations may occur in *D. scoparium* as well, but in general, external DM recruitment seems common enough to prevent extensive inbreeding.

Clonal Diversity

Although the scale of the sampling was not designed to elucidate the clonal structure of populations, we observed that the overall levels of clonality was low and that there was not much difference between the sexes, i.e., females and both types of males taken together. When separating the data set into DMs and NMs, it appears that the male clonality estimate is strongly influenced by DMs, lowering the index of association $\bar{r}_d$. Indeed, we found a higher degree of clonality in NMs as compared to DMs and females, with $\bar{r}_d$ values close to 1 (**Table 3**), suggesting more extensive clonal propagation for NMs and values closer to zero for DMs, indicating recombination. Previous studies on the perennial pleurocarps *Hylocomium splendens* and *H. lutescens* have brought out that populations often contains multiple, but few, genets in patches, suggesting that clonal expansion predominates over sexual recruitment (Cronberg et al., 2006b; Rosengren et al., 2013). Our results indicate that sexual reproduction in the studied populations more often results in recruitment of new female genets, also within female colonies. A high genotypic female diversity at patch level could possibly be explained by imperfect female control of spore germination, so that new normal-sized females sometimes are recruited along with the DMs. Rarer occurrence of NMs and their peripheral distribution in patches, indicate higher male mortality rates at recruitment and/or that female-dominated colonies forces male spores to become DMs and prevent them from escaping dwarfism by falling off the female, at least at the colony center. In order to better explain the control of development into NMs and DMs and associated patterns of genetic variability, further experimental studies are needed.

Bjärsjölagård (Bj): The Rise of a New Population?

The population of Bjärsjölagård (Bj) differs in several aspects; it is genetically divergent despite being geographically relatively close to sites Lu and Da (**Figure 1**). It probably represents a recent colonization event followed by external male recruitment, which may spread some light on genetic consequences of a founder event in a species with nannandry. The site is located in a former lime quarry, which is not a normal substrate for *D. scoparium* to occur, having preference for acidic soils at a pH around 3.2–7.1 (Tyler and Olsson, 2016). The previously open and barren calcareous ground has been partially covered by detritus from a scattered canopy of colonizing *Fagus sylvatica*, *Corylus avellana*, and *Acer platanoides* trees, allowing *D. scoparium* to establish together with species such as *H. splendens* in competition with more calciphilous species. At this locality, we found three isolated female cushions with sporophytes separated by maximum 1 m. No NMs were found, but the population of DMs was the densest of our study, with up to 88 genotyped DMs (**Table 2**; pers. obs.). The haplotypic diversity of the total dataset (GAM) harbored at Bj was fairly similar to the one at Ro. Yet, gene diversity was the lowest of all the localities ($\hat{h}$ = 0.19; **Table 3**). In further comparison to Ro, which had a similar number of females genotyped, we detected only two haplotypes within female populations. Consequently, they retained the lowest haplotype and gene diversity (G = 1.57 and $\hat{h}$ = 0.16, respectively) as well the highest index of association ($\bar{r}_d$ = 0.97) among all female populations. These results indicate that vegetative growth of female cushions is predominant at this locality and are congruent with other moss species (Cronberg et al., 2006b; Rosengren et al., 2013; Baughman et al., 2017). Moreover, five female MLLs from other populations, namely from Mu, Lu, Ro, SL, and Da (arrows in **Figure 5**), were closer genetically to DMs of Bj than individuals of any other localities. Finally, two DMs shared the same MLL with one male from Mu, making it the only MLL shared across localities. These results provide evidence that spores are dispersed over long distances and that male spores germinate more easily on female stems than on soil. Thus, even if the colonizing female population was not diverse, the DM male recruitment via spores results in quite extensive recombination events, which lead to an increase in genotypic diversity.

Conclusion

The genetic influence of dwarfism on population structure has rarely been studied and only few comparable studies exist. Here, we used SNP markers to investigate the genetic structure of seven populations of the facultatively nannandric species *D. scoparium* in Scania (Sweden) in a similar way as earlier done for the obligately nannandric species *H. lutescens*. The genetic differentiation at locality level was low indicating a strong gene flow. In the dense acrocarpous cushions of *D. scoparium*, the haplotypic diversity at patch level was higher for females (and NMs) than in the pleurocarpous and weft-forming *H. lutescens*, suggesting more frequent recruitment.

In *D. scoparium*, DMs and NMs often occurred together but the DMs were genotypically more frequent and divergent. The genotypic networks show that DM individuals at patch level represent both closely related haplotypes (suggesting inbreeding) and remotely related haplotypes (suggesting external gene flow). Nevertheless, NMs may regularly take part in fertilization as they are perennial and probably less sensitive to temporarily adverse climatic conditions. If so, the facultative nannandric system is prevented from transforming to obligate nannandry through sexual selection, contrasting to the nearly obligate or obligate nannandric conditions in *H. lutescens* and most *Dicranum* species.

Nannandry has probably developed to ensure fertilization and increase the number of male gamete donors. We see nannandry in mosses as a fertilization syndrome, a parallel to pollination syndromes in flowering plants, which ensure pollination and increases the available number of pollen donors. More studies on sexual reproduction and diversity in nannandric mosses are needed, especially in regards to facultative dwarfism. It would be particularly useful to determine the environmental condition

for the recruitment and growth of normal-sized males to better understand their contribution to the population dynamics.

DATA AVAILABILITY STATEMENT

The raw dataset generated for this study can be found in the European Nucleotide Archive (https://www.ebi.ac.uk/ena/) under accessions PRJEB37280 and ERP119316 (secondary study accession).

AUTHOR CONTRIBUTIONS

AL and NC wrote the text, conducted the sampling of the specimens, and participated in the interpretation of the results. TG and NC commented and critically reviewed the manuscript. AL conducted the laboratory work and the data analyses. TG designed the scripts specifically designed for SNP filtering. All authors approved the final manuscript.

FUNDING

This work was supported by the Royal Physiographic Society in Lund, Elly Olssons fund, Erik Philip-Sörensens Stiftelse, Kungl. Vetenskapsakademien (BS2015-0084 to AL), and Sven och Lilly Lawskis fond.

ACKNOWLEDGMENTS

We would like to thank the DNA Sequencing Facility of the Department of Biology at Lund University for sequencing and extracting SNP markers and National Bioinformatics Infrastructure Sweden (NBIS) for their valuable help with bioinformatics. We sincerely thank the SNP&SEQ Technology Platform in Uppsala (Science for Life Laboratory, Uppsala University) for genotyping the DNA samples. Finally, we would also like to thank Dr. Maria Mayol Martínez (CREAF Centre for Ecological Research and Forestry, Barcelona) for helpful discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.517547/full#supplementary-material

REFERENCES

Agapow, P. M., and Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Mol. Ecol. Notes* 1, 101–102. doi: 10.1046/j.1471-8278.2000.00014.x

Badyaev, A. V. (2002). Growing apart: an ontogenetic perspective on the evolution of sexual size dimorphism. *Trends Ecol. Evol.* 17, 369–378. doi: 10.1016/S0169-5347(02)02569-7

Baughman, J. T., Payton, A. C., Paasch, A. E., Fisher, K. M., and McDaniel, S. F. (2017). Multiple factors influence population sex ratios in the Mojave Desert moss *Syntrichia caninervis*. *Am. J. Bot.* 104, 733–742. doi: 10.3732/ajb.1700045

Blanckenhorn, W. U., Preziosi, R. F., and Fairbairn, D. J. (1995). Time and energy constraints and the evolution of sexual size dimorphism – to eat or to mate. *Evol. Ecol.* 9, 369–381. doi: 10.1007/BF01237760

Bougeard, S., and Dray, S. (2018). Supervised multiblock analysis in R with the ade4 Package. *J. Stat. Softw.* 86, 1–17. doi: 10.18637/jss.v086.i01

Bowker, M. A., Stark, L. R., McLetchie, D. N., and Mishler, B. D. (2000). Sex expression, skewed sex ratios, and microhabitat distribution in the dioecious desert moss *Syntrichia caninervis* (Pottiaceae). *Am. J. Bot.* 87, 517–526. doi: 10.2307/2656595

Chessel, D., Dufour, A. B., and Thioulouse, J. (2004). The ade4 package-I- One-table methods. *R News* 4, 5–10.

Cronberg, N. (2000). Genetic diversity of the epiphytic bryophyte *Leucodon sciuroides* in formerly glaciated versus nonglaciated parts of Europe. *Heredity (Edinb)* 84, 710–720. doi: 10.1046/j.1365-2540.2000.00719.x

Cronberg, N. (2002). Colonization dynamics of the clonal moss *Hylocomium splendens* on islands in a Baltic land uplift area: reproduction, genet distribution and genetic variation. *J. Ecol.* 90, 925–935. doi: 10.1046/j.1365-2745.2002.00723.x

Cronberg, N., Natcheva, R., and Hedlund, K. (2006a). Microarthopods mediate sperm transfer in mosses. *Science* 313, 1255. doi: 10.1126/science.1128707

Cronberg, N., Rydgren, K., and Okland, R. H. (2006b). Clonal structure and genet-level sex ratios suggest different roles of vegetative and sexual reproduction in the clonal moss *Hylocomium splendens*. *Ecography (Cop.)* 29, 95–103. doi: 10.1111/j.2006.0906-7590.04361.x

Crum, H. A., and Anderson, L. E. (1981). "Dicranum," in *Mosses of eastern North America*, Vol. 1, eds H. Crum, L. E. Anderson, and L. Anderson (New York: Columbia University Press), 151–217.

Désamoré, A., Patiño, J., Mardulyn, P., Mcdaniel, S. F., Zanatta, F., Laenen, B., et al. (2016). High migration rates shape the postglacial history of amphi-Atlantic bryophytes. *Mol. Ecol.* 25, 5568–5584. doi: 10.1111/mec.13839

Dray, S., and Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1–20.

Dray, S., Dufour, A. B., and Chessel, D. (2007). The ade4 package-II: two-table and K-table methods. *R News* 7, 47–52.

Eriksson, O. (1989). Seedling dynamics and life histories of clonal plants. *Oikos* 55, 231–238. doi: 10.2307/3565427

Eriksson, O. (1993). Dynamics of genets in clonal plants. *Trends Ecol. Evol.* 8, 313–316. doi: 10.1016/0169-5347(93)90237-j

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491. doi: 10.1093/genetics/131.2.479

Fairbairn, D. J., Blanckenhorn, W. U., and Székely, T. (2007). *Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism*. New York, NY: Oxford University Press.

Frey, W., and Kürschner, H. (2011). Asexual reproduction, habitat colonization and habitat maintenance in bryophytes. *Flora Morphol. Distrib. Funct. Ecol. Plants* 206, 173–184. doi: 10.1016/j.flora.2010.04.020

Gabriel, S., Ziaugra, L., and Tabbaa, D. (2009). SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr. Protoc. Hum. Genet.* 60, 2.12.1–2.12.18.

Geber, M. A., Dawson, T. E., and Delph, L. F. (1999). in *Gender and Sexual Dimorphism in Flowering Plants*, eds M. A. Geber, T. E. Dawson, and L. F. Delph (Berlin: Springer Verlag). doi: 10.1007/978-3-662-03908-3

Glime, J. M. (2007). *Bryophyte Ecology: Physiological Ecology*, Vol. 1. Houghton, MI: EBook sponsored by Michigan Technological University and the International Association of Bryologists.

Glime, J. M., and Bisang, I. (2017). "Sexuality: size and sex differences," in *Bryophyte Ecology: Physiological Ecology*, Vol. 1, ed. J. M. Glime (Houghton, MI: Ebook sponsored by Michigan Technological University and the International Association of Bryologists).

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. a., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Grünwald, N. J., and Hoheisel, G.-A. (2006). Hierarchical analysis of diversity, selfing, and genetic differentiation in populations of the oomycete *Aphanomyces euteiches*. *Phytopathology* 96, 1134–1141. doi: 10.1094/PHYTO-96-1134

Grünwald, N. J., Kamvar, Z. N., Everhart, S. E., Tabima, J. F., and Knaus, B. J. (2017). *Population Genetics and Genomics in R*. Available online at: http://grunwaldlab.github.io/Population_Genetics_in_R/ (accessed March 17, 2021).

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Hedenäs, L., and Bisang, I. (2004). Key to European *Dicranum* species. *Herzogia* 17, 179–197.

Hedenäs, L., and Bisang, I. (2011). The overlooked dwarf males in mosses—Unique among green land plants. *Perspect. Plant Ecol. Evol. Syst.* 13, 121–135. doi: 10.1016/j.ppees.2011.03.001

Hutsemékers, V., Hardy, O. J., Mardulyn, P., Shaw, A. J., and Vanderpoorten, A. (2010). Macroecological patterns of genetic structure and diversity in the aquatic moss *Platyhypnidium riparioides*. *New Phytol.* 185, 852–864. doi: 10.1111/j.1469-8137.2009.03094.x

Ireland, R. R. (2007). "Dicranum (Family Dicranaceae)," in *Flora of North America*, ed. R. H. Zander (New York, NY: Oxford University Press), 397–420.

Isaac, J. L. (2005). Potential causes and life-historis consequences of sexual size dimorphism in mammals. *Mamm. Rev.* 35, 101–115. doi: 10.1111/j.1365-2907.2005.00045.x

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6:208. doi: 10.3389/fgene.2015.00208

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281. doi: 10.7717/peerj.281

Laaka-Lindberg, S., Korpelainen, H., and Pohjamo, M. (2003). Dispersal of asexual propagules in bryophytes. *J. Hattori Bot. Lab.* 192, 319–330.

Lang, A. S., Kruijer, J. D., and Stech, M. (2014). DNA barcoding of arctic bryophytes- an example from the moss genus *Dicranum* (Dicranaceae, Bryophyta). *Polar Biol.* 37, 1157–1169. doi: 10.1007/s00300-014-1509-7

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* 15, 237–240. doi: 10.1093/besa/15.3.237

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., and Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Mol. Ecol. Resour.* 15, 28–41. doi: 10.1111/1755-0998.12291

McDaniel, S. F., and Shaw, A. J. (2005). Selective sweeps and intercontinental migrations in the cosmopolitan moss *Ceratodon purpureus* (Hedw.) Brid. *Mol. Ecol.* 14, 1121–1132. doi: 10.1111/j.1365-294x.2005.02484.x

McDaniel, S. F., Willis, J. H., and Shaw, A. J. (2007). A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. *Genetics* 176, 2489–2500. doi: 10.1534/genetics.107.075424

McElligott, A. G., Gammell, M. P., Harty, H. C., Paini, D. R., Murphy, D. T., Walsh, J. T., et al. (2001). Sexual size dimorphism in fallow deer (*Dama dama*): do larger, heavier males gain greater mating success? *Behav. Ecol. Sociobiol.* 49, 266–272. doi: 10.1007/s002650000293

Mishler, B. D. (1988). "Reproductive ecology of bryophytes," in *Plant Reproductive Biology. Patterns and Strategies*, eds J. L. Doust and L. L. Doust (Oxford: Oxford University Press), 285–306.

Moya-Laraño, J., Halaj, J., and Wise, D. H. (2002). Climbing to reach females: romeo should be small. *Evolution (N. Y)* 56, 420–425. doi: 10.1554/0014-3820(2002)056[0420:ctrfrs]2.0.co;2

Paasch, A. E., Mishler, B. D., Nosratinia, S., Stark, L. R., and Fisher, K. M. (2015). Decoupling of sexual reproduction and genetic diversity in the femalebiased mojave desert moss *Syntrichia caninervis* (Pottiaceae). *Int. J. Plant Sci.* 176, 751–761. doi: 10.1086/682708

Patiño, J., and Vanderpoorten, A. (2018). Bryophyte biogeography. *CRC Crit. Rev. Plant Sci.* 37, 175–209. doi: 10.1080/07352689.2018.1482444

Pichonet, A., and Gradstein, S. R. (2012). Male dwarfism in the genus *Dicranum* (Dicranaceae)-a review. *Cryptogam. Bryol.* 33, 299–311. doi: 10.7872/cryb.v33.iss3.2012.299

Pietsch, T. W. (2005). Dimorphism, parasitism, and sex revisited: modes of reproduction among deep-sea ceratioid anglerfishes (Teleostei: Lophiiformes). *Ichthyol. Res.* 52, 207–236. doi: 10.1007/s10228-005-0286-2

Poissant, J., Wilson, A. J., and Coltman, D. W. (2010). Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. *Evolution (N. Y)* 64, 97–107. doi: 10.1111/j.1558-5646.2009.00793.x

Puixeu, G., Pickup, M., Field, D. L., and Barrett, S. C. H. (2019). Variation in sexual dimorphism in a wind-pollinated plant: the influence of geographical context and life-cycle dynamics. *New Phytol.* 224, 1108–1120. doi: 10.1111/nph.16050

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Available online at: http://www.r-project.org/ (accessed March 17, 2021).

Rosengren, F., and Cronberg, N. (2014). The adaptive background of nannandry: dwarf male distribution and fertilization in the moss *Homalothecium lutescens*. *Biol. J. Linn. Soc.* 113, 74–84. doi: 10.1111/bij.12332

Rosengren, F., Cronberg, N., and Hansson, B. (2016). Balance between inbreeding and outcrossing in a nannandrous species, the moss *Homalothecium lutescens*. *Heredity (Edinb)* 116, 107–113. doi: 10.1038/hdy.2015.79

Rosengren, F., Cronberg, N., Reitalu, T., and Prentice, H. C. (2013). Genetic variation in the moss *Homalothecium lutescens* in relation to habitat age and structure. *Botany* 91, 431–441. doi: 10.1139/cjb-2012-0258

Rosengren, F., Hansson, B., and Cronberg, N. (2015). Population structure and genetic diversity in the nannandrous moss *Homalothecium lutescens*: does the dwarf male system facilitate gene flow? *BMC Evol. Biol.* 15:12. doi: 10.1186/s12862-015-0545-4

Ross, P., Hall, L., Smirnov, I., and Haff, L. (1998). High level multiplex genotyping by MALDI- TOF mass spectrometry. *Nat. Biotechnol.* 16, 1347–1351. doi: 10.1038/4328

Rouse, G. W., Wilson, N. G., Worsaae, K., and Vrijenhoek, R. C. (2015). Report a dwarf male reversal in bone-eating worms. *Curr. Biol.* 25, 236–241. doi: 10.1016/j.cub.2014.11.032

Rudolph, H., Kirchhoff, M., and Gliesmann, S. (1988). "Sphagnum culture techniques," in *Proceedings of the Bryological Methods Workshop, Mainz*, ed. J. M. Glime (Nichinan: Hattori Botanical Laboratory), 25–34.

Rydgren, K., Cronberg, N., and Økland, R. H. (2006). Factors influencing reproductive success in the clonal moss, *Hylocomium splendens*. *Oecologia* 147, 445–454. doi: 10.1007/s00442-005-0290-2

Sagmo Solli, I. M., Söderström, L., Bakken, S., Flatberg, K. I., and Pedersen, B. (2000). Studies of fertility of *Dicranum majus* in two populations with contrasted sporophyte production. *J. Bryol.* 22, 3–8. doi: 10.1179/jbr.2000.22.1.3

Stoddart, J. A., and Taylor, J. F. (1988). Genotypic diversity: estimation and prediction in samples. *Genetics* 118, 705–711. doi: 10.1093/genetics/118.4.705

Storm, N., Darnhofer-Patel, B., van den Boom, D., and Rodi, C. P. (2003). MALDI-TOF mass spectrometry based SNP genotyping. *Methods Mol. Biol.* 212, 241–262. doi: 10.1385/1-59259-327-5:241

Tamm, C. O. (1964). Growth of *Hylocomium splendens* in relation to tree canopy. *Bryologist* 67, 423–426. doi: 10.2307/3240767

Tyler, T., and Olsson, P. A. (2016). Substrate pH ranges of south Swedish bryophytes-Identifying critical pH values and richness patterns. *Flora Morphol. Distrib. Funct. Ecol. Plants* 223, 74–82. doi: 10.1016/j.flora.2016.05.006

Vamosi, J. C., Otto, S. P., and Barrett, S. C. H. (2003). Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. *J. Evol. Biol.* 16, 1006–1018. doi: 10.1046/j.1420-9101.2003.00559.x

Vanderpoorten, A., and Goffinet, B. (2009). *Introduction to Bryophytes*. Cambridge: Cambridge University Press.

Vollrath, F. (1998). Dwarf males. *Trends Ecol. Evol.* 13, 159–163. doi: 10.1016/s0169-5347(97)01283-4

Vrijenhoek, R. C., Johnson, S. B., and Rouse, G. W. (2008). Bone-eating Osedax females and their "harems" of dwarf males are recruited from a common larval pool. *Mol. Ecol.* 17, 4535–4544. doi: 10.1111/j.1365-294X.2008.03937.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership