CHROMOSOMAL EVOLUTION IN PLANTS

EDITED BY: Martin A. Lysak and Hanna Weiss-Schneeweiss

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-400-1
DOI 10.3389/978-2-88971-400-1

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

CHROMOSOMAL EVOLUTION IN PLANTS

Topic Editors:
Martin A. Lysak, Masaryk University, Czechia
Hanna Weiss-Schneeweiss, University of Vienna, Austria

Citation: Lysak, M. A., Weiss-Schneeweiss, H., eds. (2021). Chromosomal Evolution in Plants. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-400-1

Table of Contents

frontiers in Plant Science

EDITORIAL
published: 29 July 2021
doi: 10.3389/fpls.2021.726330

Editorial: Chromosomal Evolution in Plants

Martin A. Lysak[1]* *and Hanna Weiss-Schneeweiss*[2]*

[1] Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czechia, [2] Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

Keywords: chromosomal rearrangements, cytogenomics, karyotype evolution, next generation sequencing, paleogenomics, polyploidy, repeatome, repetitive DNA

Editorial on the Research Topic

Chromosomal Evolution in Plants

Chromosomal evolution is a driving force underlying diversification and speciation in plants. Chromosome numbers and morphology, as well as nuclear DNA amount and composition, are very diverse in plants. During the last few decades, a range of new tools has been developed that allow for better insights into chromosomal and genome evolution in plants, far beyond chromosome counting and establishing karyotype structure. One of the most important tools is next generation sequencing (NGS) technology which has recently revolutionized nearly all biological disciplines. Genome skimming and whole-genome sequencing elevated classical cytogenetics to a new level of modern evolutionary cytogenomics (see also Hu et al., 2020). Comparative plant (cyto)genomics allows for physical localization of DNA sequences on chromosomes, improving whole-genome and chromosome-level assemblies, but also provides a wealth of data on nuclear repeatomes, chromosome structures, mechanisms of chromosomal alterations, and interphase chromosome organization. This Research Topic brings together a collection of research, methodological and review articles advancing our understanding of the role of chromosomal and karyotype evolution in land plant diversification and speciation. It also allows for the exchange of data, ideas, and new hypotheses taking full advantage of novel methodological approaches. The Research Topic consists of 13 original papers, three reviews and one perspective. Altogether these studies fall into one of the four broader intersecting topics: (1) Pathways of karyotype and genome evolution and their role in plant speciation (2) Polyploidy and post-polyploid genome diploidization (3) Paleogenomics and paleogenomes (4) Repeatome and genome size evolution.

PATHWAYS OF KARYOTYPE AND GENOME EVOLUTION AND THEIR ROLE IN PLANT SPECIATION

Karyotype changes often accompany plant diversification and speciation. Among flowering plant families, the Brassicaceae (mustards, crucifers) has become a prominent model system for studying plant genome and chromosomal evolution due to the availability of abundant genomic resources, such as the high-quality reference genome of *Arabidopsis thaliana*.

The Brassicaceae tribe Aethionemeae, containing a single genus *Aethionema*, is the sister clade to all remaining crucifer lineages. The evolution of the genomes and phylogenetic relationships of this tribe within the Brassicaceae is addressed by Walden et al. Crucifer genomic blocks (GBs) were used to infer syntenies between the genome sequences of *Ae. arabicum* and other crucifer genomes. In contrast to the largely conserved genomic structure of most non-polyploid Brassicaceae lineages, GBs in *Aethionema* were highly rearranged, similar to genomes in the tribe Arabideae. Thus, Arabideae (e.g., the Alpine rock-cress, *A. alpina*) might also have diverged before the other major crucifer lineages.

OPEN ACCESS

Edited and reviewed by:
Kathleen Pryer,
Duke University, United States

***Correspondence:**
Martin A. Lysak
martin.lysak@ceitec.muni.cz
Hanna Weiss-Schneeweiss
hanna.schneeweiss@univie.ac.at

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

Received: *16 June 2021*
Accepted: *07 July 2021*
Published: *29 July 2021*

Citation:
Lysak MA and Weiss-Schneeweiss H (2021) Editorial: Chromosomal Evolution in Plants. Front. Plant Sci. 12:726330. doi: 10.3389/fpls.2021.726330

The Brassicaceae tribe Boechereae comprises about 130 species in nine genera including a number of apomictic species of hybrid origin. Mandáková et al. report here on analyses of the genome structure of seven genera in this tribe using comparative chromosome painting. The ancestral Boechereae genome of $n = 7$ is inferred to be derived from an $n = 8$ genome by descending dysploidy. Although most of the Boechereae genera have a conserved genomic structure, chromosomal rearrangements (most often inversions) have accompanied the divergence of some genera and species within the tribe.

Elucidating chromosomal rearrangements leading to karyotype restructuring requires the use of appropriate genomic tools, as has been elegantly shown for numerous crucifers. Another family for which such tools have now been developed is the Musaceae (bananas). Here, Šimoníková et al. report on the development of chromosome/chromosome-arm specific oligo painting probes that allow for the identification of all chromosomes in *Musa* spp. and for physically anchoring pseudomolecules of the reference genome sequence to individual chromosomes. This has, for the first time, allowed us to gain deeper insight into the evolution of A, B, and S banana genomes, as well as identify chromosomal translocations that have accompanied genome evolution and speciation in *Musa*.

Phylogenetic hypotheses and molecular clock analyses have become essential components of studies on karyotype evolution, as demonstrated by Costa et al. in their analysis of the Amaryllidaceae subfamily Allioideae (e.g., chives, garlics, onions). The three tribes of this subfamily (Allieae, Gilliesieae, and Tulbaghieae) have evolved toward their current intercontinental disjunctions to the Northern Hemisphere, South America, and Southern Africa, respectively. Using a dated phylogeny, together with models of chromosome number evolution and diversification rate analysis, the authors reconstruct the biogeography of the Allioideae to be the result of vicariance due to the break-up of Gondwana; centric fissions are identified as the most likely mechanism of karyotypic diversification in the tribe Gilliesieae.

Chromosome number variation in eukaryotes results from frequent, sometimes very complex, chromosomal rearrangements, which may eventually lead to the reduction of chromosomes, i.e., genetic linkage groups (descending dysploidy). Although such reductions are usually easily detectable in comparative analyses, reconstructing the mechanism underlying these changes is more challenging. Here, Udall et al. investigate the chromosomal rearrangements mediating $n-1$ descending dysploidy in two genera (*Gossypioides* and *Kokia*), sister to the cotton genus (*Gossypium*), using high-quality genome assemblies and Hi-C data. These genome comparisons show that the two descending dysploidies were not mediated by simple recombination between two ancestral chromosomes, but that multiple steps were required to generate the extant genome structures; the evolution of plant chromosomes can be complex and does not necessarily follow the most parsimonious pathways.

As plant genome assemblies are steadily being improved by the increasing read length, we are obtaining a better view of the incidence, scale and structure of chromosomal rearrangements in plant genomes. Inversions are probably the most common chromosomal rearrangements and are ubiquitous across the plant kingdom. Here timely, Huang and Rieseberg review the latest research on chromosome inversions in plant genomes, focusing on the role of inversions in speciation in the presence of gene flow and models of inversion fixation. The authors also discuss sequential inversions and the purported functional link between suppressed recombination and the differentiation of sex chromosomes in plants, such as in the dioecious *Silene* species (Bačovský et al.).

POLYPLOIDY AND POST-POLYPLOIDY GENOME DIPLOIDIZATION

Whole genome duplications (WGD) have accompanied the evolution of angiosperms since the beginning. Past and ongoing rounds of polyploidization continue to shape the majority of extant plant genomes.

In a perspective article, Levin offers a personal view on the continued impact of polyploidization on plant diversification. Specifically, the author focuses on mechanisms of polyploid diversification, reasons why WGDs are not followed by immediate diversification thrusts, and on polyploidy going forward. The polyploid wave which began roughly 60 million years ago will, as Levin suggests, continue to rise in the coming millennia, in part due to increasing climate and environmental changes and the impact of other anthropogenic drivers. The proportion of polyploid species might also increase as polyploids are likely to gain an advantage over diploids in changing environments. The increase in polyploidy in the coming millennia might also involve the frequent formation of higher level polyploids and post-polyploid dysploids.

Another process that has a significant impact on plant genome evolution is hybridization. The genomic consequences of such genome mergers alone and in combination with WGD are addressed here in a review paper by Glombik et al. The genomes of newly established interspecific hybrids often undergo dramatic changes, including chromosomal rearrangements, changes in the amount and localization of repetitive DNAs, and gene expression modifications. Successful hybridizations and WGD events are followed by genomic and cytological diploidization events. One of the most significant aspects of this process is establishment of diploid-like chromosome pairing. The authors provide an overview of the current knowledge of genomic changes in interspecific homoploid and alloploid hybrids, focusing on chromosome pairing, and discuss parental genome dominance at various levels of organization in relation to the stability of hybrid genomes.

Domesticated plant species often have polyploid ancestry. Here, Hardigan et al. analyse the genome structure and diversity of the allo-octoploid cultivated strawberry (*Fragaria* × *ananassa*). Comparative genomic analyses show that geographically diverse wild octoploids were diploidized, nearly completely collinear, and retained strong macro-synteny with their diploid progenitor taxa. The conserved genome structure of octoploid *Fragaria* species allows for

unimpeded gene flow during repeated cycles of homoploid hybridization without the formation of reproductive barriers or loss of fertility.

SPECIALIZED CHROMOSOMES AND CHROMOSOMAL SYSTEMS

Dioecy is found in 5–6% of angiosperms genera (Vyskot and Hobza, 2004). Some of these taxa have developed differentiated sex chromosomes that have evolved from regular autosomal chromosome pairs. The genus *Silene* (campions, Caryophyllaceae) includes both dioecious and gynodioecious species. Two dioecious species, *Silene latifolia* and *S. dioica*, possess heteromorphic sex chromosomes and an XX/XY sex determination system. Here, Bačovský et al. addressed the origin of X/Y sex chromosomes in *Silene* by mapping an oligo painting probe enriched for X-linked scaffolds to chromosomes of two dioecious and two gynodioecious species. The results suggest shared ancestry of the sex chromosomes in the two dioecious species, accompanied by subsequent extensive chromosomal rearrangements.

Centromeres are chromosomal regions that facilitate faithful chromosome segregation during cell division. The vast majority of plants possess monocentric chromosomes with localized centromeres. Several plant families, however, harbor species or genera with holocentric chromosomes. Cyperaceae (sedges) is an example of one of the families in which holocentricity has been demonstrated and is shown to be associated with frequent dysploid and polyploidization events. Here, Burchardt et al. studied karyotype morphology and genome size evolution in 35 *Rhynchospora* taxa. The reported 22-fold genome size variation and chromosome numbers ranging from $2n = 4$ to 61 indicate much higher levels of karyotypic diversity than previously believed. Base chromosome number changes were inferred to have occurred in all *Rhynchospora* lineages via dysploidy and polyploidy, driving chromosomal evolution in the genus.

Another genus that harbors species with both monocentric and holocentric chromosomes is parasitic *Cuscuta* (Convolvulaceae, bindweeds). In most organisms, centromeres are determined epigenetically by the presence of the centromere-specific histone variant CENH3. In holocentrics, CENH3 is typically distributed along the entire chromosome length. Here, Oliveira et al. identified CENH3 genes in the holocentric plant *Cuscuta europaea* and found a unique pattern of CENH3 distribution along mitotic chromosomes in this species. Two major CENH3 variants were expressed and co-localized in one to three discrete heterochromatic regions per chromosome that contained the satellite repeat CUS-TR24, whereas the rest of the chromatin appeared to be devoid of CENH3. In contrast, the mitotic spindle microtubules attached at uniform density along the entire chromosome length. The data suggest that CENH3 either lost its function or acts in parallel to an additional CENH3-free mechanism of kinetochore assembly.

REPEATOME AND GENOME SIZE EVOLUTION

The tremendous amount of NGS data available allows for detailed insights into plant genome composition and the dynamics of different types of repetitive DNA during diversification and speciation. Apart from a general characterization of the repeat types, NGS data analyses provide the means to conduct comprehensive comparative analyses on the genomic repeat composition of any plant species and their contribution to genome size change, as well as to analyse the evolution of individual repeat families.

Studies involving repeatome dynamics oftentimes focus on model organisms, whereas few provide comprehensive investigations across the genomes of related taxa. Here, McCann et al. analyse the evolution of repeats in a group of 13 closely related diploid species of the genus *Melampodium* (Asteraceae, sunflower family) differing ca. 4.5-fold in genome size despite possessing the same base chromosome number of $x = 10$. Analyses of genome skimming data using the RepeatExplorer pipeline (Novák et al., 2020) determined that patterns of repeat evolution were found to be highly correlated with the phylogenetic position of the species. Evidence was found for strong phylogenetic signal and differential evolutionary rates of major lineages of repeats in the diploid genomes of *Melampodium*.

Several angiosperm families have evolved biomodal karyotypes that consist of chromosomes in two contrasting size classes. One such system is found in the neotropical genus *Eleutherine* (Iridaceae, irises). Both species possess karyotypes with one large and five small chromosome pairs. The large chromosome pair of *E. bulbosa* additionally carries pericentric inversion in permanent heterozygosity in one of its chromosomes. Here, Báez et al. investigate the repeatomes of both species to test whether the permanent chromosomal inversion in the large chromosome influenced the dynamics of repetitive DNA sequences. While the accumulation of repeats differed between large and small chromosomes, the most abundant repeats showed a similar chromosomal distribution in both homologs of the large pair, regardless of the presence of the permanent inversion.

Long terminal repeat (LTR) retrotransposons constitute the most significant part of most plant genomes and the periodical bursts of their activity play an important role in genome size changes, among others. LTR regions are used to estimate the age of retrotransposons, and thus, to date periods of their activity. Here, Jedlicka et al. measure the LTR divergence of thousands of LTR retrotransposons to determine their age and evolutionary dynamics in 15 plant species representing major lineages across Viridiplantae and ranging nearly 10-fold in genome size. The authors hypothesize that gene conversion might have contributed to the higher observed DNA sequence divergence of LTR regions of nested retrotransposon copies compared to the divergence of LTRs of pre-existing retrotransposons

into which they had been inserted. The negative correlation between the frequency of gene conversion and the abundance of solo LTRs suggests interference between gene conversion and ectopic recombination.

Another type of tandem repeat that is ubiquitous in nearly all known eukaryotic genomes is found in the telomeres, structures that protect and maintain the ends of linear chromosomes. Telomeric DNA sequences are usually tandemly arranged minisatellites, typically following the formula (TxAyGz)*n*. The review article by Peska and Garcia provides an up-to-date overview of the diversity of plant telomeres following recent discoveries of novel telomeric variants in various plant groups. The authors also provide an overview of methods used for identification of telomeric motifs.

Tandemly repeated ribosomal genes (rRNA genes) are widely used in molecular and cytogenetic studies, including analyses of parental (sub)genomes in homoploid and allopolyploid hybrids. Here, Garcia et al. analyzed 5S rDNA arrays in over 80 homoploid and alloploid hybrid species of different evolutionary ages using graph clustering implemented in the RepeatExplorer pipeline. Comparative analysis of 5S rDNAs in hybrids and their progenitor species allowed for identification of homoeologous 5S rRNA gene families in both evolutionarily young and older taxa, thereby facilitating inferences of their origin. The shapes of cluster graphs robustly reflect the organization and homogeneity of 5S rDNA repeats within each of the parental subgenomes. The authors proposed that this approach, together with cytogenetic analyses, might assist with inferring parental origin of the hybrid taxa.

CONCLUSIONS AND PERSPECTIVES

This Research Topic presents most recent advances in the field of chromosomal plant evolution, broadly covering most important topics centered around plant genome evolution on all levels of its organization as well as its temporal and spatial aspects. This collection of papers clearly demonstrates that the rapid development of genomic tools and approaches, such as NGS, high throughput data analyses and novel cytogenomic techniques will continue to facilitate the analyses of more wild plant groups and will allow for in-depth analyses of various aspects of chromosomal and genome evolution. Thus, over the coming years we will continue to test old hypotheses in light of novel methodologies and data, and unravel new patterns and phenomena that contribute to genome evolution in plants.

AUTHOR CONTRIBUTIONS

MAL and HW-S drafted the manuscript. Both authors revised and approved the final version.

FUNDING

MAL was supported by the CEITEC 2020 project (LQ1601) and research in the laboratory of HW-S was supported by grant AI 2119 (FWF).

ACKNOWLEDGMENTS

We greatly appreciate the contributions of all authors and reviewers, as well as the editorial help of Steven Dodsworth.

REFERENCES

Hu, Q., Maurais, E. G., and Ly, P. (2020). Cellular and genomic approaches for exploring structural chromosomal rearrangements. *Chromosome Res.* 28, 19–30. doi: 10.1007/s10577-020-09626-1

Novák, P., Neumann, P., and Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* 15, 3745–3776. doi: 10.1038/s41596-020-0400-y

Vyskot, B., and Hobza, R. (2004). Gender in plants: sex chromosomes are emerging from the fog. *Trends Genet.* 20, 432–438. doi: 10.1016/j.tig.2004.06.006

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 October 2019
doi: 10.3389/fpls.2019.01170

Together But Different: The Subgenomes of the Bimodal *Eleutherine* Karyotypes Are Differentially Organized

Mariana Báez[1], Magdalena Vaio[2], Steven Dreissig[3], Veit Schubert[3], Andreas Houben[3] and Andrea Pedrosa-Harand[1]*

[1] Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Federal University of Pernambuco, Recife, Brazil,
[2] Laboratory of Genetics, Department of Plant Biology, College of Agronomy, University of the Republic, Montevideo, Uruguay,
[3] Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

OPEN ACCESS

Edited by:
Martin A. Lysak,
Masaryk University, Czechia

Reviewed by:
Kui Lin,
Beijing Normal University, China
Hanna Weiss-Schneeweiss,
University of Vienna, Austria

***Correspondence:**
Mariana Báez
marian_jaris@hotmail.com

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

Received: *07 June 2019*
Accepted: *27 August 2019*
Published: *07 October 2019*

Citation:
Báez M, Vaio M, Dreissig S, Schubert V, Houben A and Pedrosa-Harand A (2019) Together But Different: The Subgenomes of the Bimodal Eleutherine Karyotypes Are Differentially Organized. Front. Plant Sci. 10:1170. doi: 10.3389/fpls.2019.01170

Bimodal karyotypes are characterized by the presence of two sets of chromosomes of contrasting size. *Eleutherine bulbosa* (2*n* = 12) presents a bimodal karyotype with a large chromosome pair, which has a pericentric inversion in permanent heterozygosity with suppressed recombination, and five pairs of three to four times smaller chromosomes. Aiming to understand whether high copy number sequence composition differs between both chromosome sets, we investigated the repetitive DNA fraction of *E. bulbosa* and compared it to the chromosomal organization of the related *Eleutherine latifolia* species, not containing the pericentric inversion. We also compared the repetitive sequence proportions between the heteromorphic large chromosomes of *E. bulbosa* and between *E. bulbosa* and *E. latifolia* to understand the influence of the chromosome inversion on the dynamics of repetitive sequences. The most abundant repetitive families of the genome showed a similar chromosomal distribution in both homologs of the large pair and in both species, apparently not influenced by the species-specific inversions. The repeat families Ebusat1 and Ebusat4 are localized interstitially only on the large chromosome pair, while Ebusat2 is located in the centromeric region of all chromosomes. The four most abundant retrotransposon lineages are accumulated in the large chromosome pair. Replication timing and distribution of epigenetic and transcriptional marks differ between large and small chromosomes. The differential distribution of retroelements appears to be related to the bimodal condition and is not influenced by the nonrecombining chromosome inversions in these species. Thus, the large and small chromosome subgenomes of the bimodal *Eleutherine* karyotype are differentially organized and probably evolved by repetitive sequences accumulation on the large chromosome set.

Keywords: retrotransposons, satellite DNA, repetitive sequences accumulation, DNA replication, histone modification, inversion

INTRODUCTION

Bimodal karyotypes are characterized by the presence of two sets of chromosomes of contrasting size. The origin of bimodal karyotypes is usually associated with one of the following processes: (i) chromosomal rearrangements involving fusion–fission events generate large chromosomes as fusion products of small chromosomes, or small chromosomes result from the fission of large chromosomes (Burt, 2002; Schubert and Lysak, 2011; Yin et al., 2014). (ii) The combination of different parental species (allopolyploidization) may combine species with different chromosome sizes (McKain et al., 2012; Shirakawa et al., 2012). (iii) The differential accumulation of repetitive sequences may increase the size of a subset of chromosomes (de la Herrán et al., 2001). Bimodal karyotypes are common within several animals groups, such as birds, reptiles, and amphibians (Stock and Mengden, 1975; Masabanda et al., 2004; Noronha et al., 2016). Also, many plant genera, such as *Agave*, *Yucca*, *Hosta* (Akemine, 1935; Watkins, 1936; Palomino et al., 2012), *Aloe* (Brandham and Doherty, 1998; Fentaw et al., 2013), and *Hypochaeris* (Fiorin et al., 2013), show bimodal karyotypes.

In animal bimodal karyotypes, gene content, the abundance of heterochromatic repetitive sequences, and the replication behavior differ between both chromosome sets (McQueen et al., 1998; Smith et al., 2000). For instance, chicken microchromosomes are early replicating, harbor twice as many genes as macrochromosomes, and are associated with an increased gene transcriptional activity (McQueen et al., 1998). In contrast, in most bimodal plant groups, the chromosome organization is largely unknown. The bimodal karyotypes of some Orchidaceae species contain large chromosomes with a higher proportion of C-banding–positive heterochromatin (D'Emerico et al., 1999). In *Ornithogalum longibracteatum* (Hyacinthaceae), one satellite DNA sequence (satDNA) is the major constituent of the heterochromatin of the large chromosomes (Pedrosa et al., 2001). A specific satDNA, found in *Muscari comosum* (Hyacinthaceae), is related to the heterochromatic bands of the large chromosomes, and it has been suggested to cause the increase of asymmetry of the karyotypes within this genus (de la Herrán et al., 2001). Independent of composition and origin, in both animal and plant bimodal species, it was suggested that the maintenance of these chromosome size differences could be related to the genome structure and function (Coullin et al., 2005; Vosa, 2005; Griffin et al., 2015).

Eleutherine (Iridaceae) is a neotropical genus of the subfamily Iridoideae and comprises two species, both with bimodal karyotypes (Goldblatt and Snow, 1991). *Eleutherine bulbosa* ($2n = 12$) has one chromosome pair (chromosome I), which is three to four times larger than the other pairs. The large chromosome pair is heteromorphic due to an asymmetric pericentric inversion in heterozygosity, encompassing about 70% of the chromosome and resulting in one acrocentric and one metacentric homolog (Guerra, 1988). This pair contains two DAPI-positive heterochromatic bands. They are located interstitially in the long arm of the acrocentric and terminally in the short arm of the metacentric homolog. CMA-positive bands are located in the pericentromeric region of both homologs. The presence of rDNA sites is limited to chromosome pair I. While the 35S rDNA sites are located inside of the chromosomal inversion, the 5S rDNA sites are duplicated in the terminal region of the long arm of both chromosomes, outside of the inversion (Feitoza and Guerra, 2011). The second species of the genus, *Eleutherine latifolia*, has also a $2n = 12$ bimodal karyotype with a pair of large acrocentric chromosomes, but without an inversion (Goldblatt and Snow, 1991).

All small chromosomes of *E. bulbosa* are enriched in euchromatin marks, like acetylated histone H4K5 and dimethylated H3K4. In contrast, the large chromosome pair is 5-mC hypermethylated (Feitoza and Guerra, 2011), showing a chromatin differentiation between both chromosome sets. Meiotic analysis showed that the inverted region of the large chromosome pair was devoid of recombination, with chiasmata observed only outside the inversion loop (Guerra, 1991). All analyzed individuals and populations of *E. bulbosa* were heterozygous, and this heterozygosity is supposed to be fixed preferentially by asexual reproduction (Guerra, 1988; Guerra, 1991).

The process of recombination is linked to the evolution of repetitive sequences, as observed for satellite DNA homogenization *via* gene conversion (Feliner and Rosselló, 2012). Furthermore, unequal recombination between homologous chromatids or illegitimate recombination was proposed as powerful mechanisms for removing repetitive sequences (Tenaillon et al., 2010) and decreasing genome size (Renny-Byfield et al., 2011). Thus, chromosomal regions devoid of recombination, such as inverted regions, could tend to accumulate different types of repetitive sequences, which may evolve differentially from the rest of the genome. Within this context, the 5-mC hypermethylation of *E. bulbosa* chromosome pair I, which contrasts to the small chromosome pairs, and the lack of recombination between the homologs of chromosome pair I in a large segment led to the following questions. Does the distribution of repetitive sequences, epigenetics histone marks and timing of DNA replication differ between large and small chromosomes within a bimodal karyotype? Does the repetitive composition differ in the nonrecombining inverted region between the homologs of chromosome pair I? Is the distribution of repeats conserved between large and small chromosomes within the *Eleutherine* species?

Therefore, we describe the repeat composition and chromosome organization of *E. bulbosa*. The findings were compared to the sister species *E. latifolia*, also showing a bimodal karyotype but lacking the large chromosome inversion.

MATERIALS AND METHODS

Materials

Plants of *E. bulbosa* were collected in Aracuara, Bahia State, Brazil (voucher number UFP 82763), and cultivated in the experimental garden of the Laboratory of Plant Cytogenetic and Evolution from the Federal University of Pernambuco, Recife, Brazil. Seeds of *E. latifolia* were kindly provided by Dr Guadalupe Munguía Linno from Guadalajara University, Mexico. Seeds were germinated in a wet chamber (3–4 months), the seedlings were transferred into soil and cultivated in a germination room at 24°C.

Genome Size Estimation

Samples were prepared from 40 to 50 mg of young leaves of *E. bulbosa* (Miller) Urban or *E. latifolia* (Standl and L. O. Williams) Ravenna in 1 mL of LB nuclear isolation buffer and filtered through a 30 μm nylon filter (Doležel et al., 2007). *Solanum lycopersicum* L. (2C = 1.96 pg) served as standard. Nuclei were stained with propidium iodide (50 μg mL^{-1}), and RNase (50 mg mL^{-1}) was added to prevent staining of double-stranded RNA. The nuclear DNA content was determined with a Partec CyFlow SL (Partec) flow cytometer, and results were analyzed with Flomax program. For genome size estimations, three replicates were analyzed, and the nuclear DNA content for each species was calculated according to the formula:

$$2C \text{ nuclear DNA content of the sample (pg)} = \frac{\text{sample G0/G1}}{\text{reference standard G0/G1}} \times 2\text{C nuclear DNA content of the reference standard}$$

Extraction of Genomic DNA and DNA Isolation From Microdissected Chromosomes

Total genomic DNA was extracted from young leaves of one individual of *E. bulbosa* using the DNAeasy Plant Mini Kit (Qiagen) according to manufacturer's instruction.

For microdissection of the two large homologous chromosomes, root tips were collected from bulbs, pretreated in 8-hydroxyquinoline at 10°C for 24 h and fixed in 2% formaldehyde for 15 min under vacuum. Fixed root tips were chopped in a nuclei isolation buffer (Doležel et al., 2007) and filtered through a 30-μm nylon membrane. The cell solution was centrifuged onto a microscopic slide at 2,000 revolutions/min (rpm) for 10 min (Shandon, CytoSpin3). Slides, mounted in 4′, 6-diamidino-2-phenylindole (DAPI), were hit with a metal needle to physically separate chromosomes from broken cells. Ten chromosomes of each large chromosome homolog were isolated by microdissection with a glass needle, using a Zeiss Axio Zoom. V16 microscope coupled to an AxioCam 289 MRc5 digital camera (Zeiss) and the aureka® microsampling platform (aura-optik). The chromosomal DNA was amplified by multiple displacement amplification according to the protocol described in Dreissig et al. (2015) with minor modifications. Briefly, chromosomes were collected in 0.5 μL H_2O and 1 μL sample buffer (GE Healthcare, Genomiphi V2) and then were incubated in alkaline lysis buffer (Gole et al., 2013) and 0.1 μg/μL of proteinase K (Sigma) at 37°C for 1 h, followed by heat inactivation at 65°C for 10 min. After incubation, 0.5 μL of neutralizing buffer (Gole et al., 2013) was added, and the samples were left on ice, while a master mix was prepared (3.5 μL of sample buffer, 4.5 μL reaction buffer and 0.5 μL of enzyme mix; Genomiphi V2; GE Healthcare). The samples were incubated at 30°C for 8 h followed by heat inactivation at 65°C for 10 min and then cooled down to 4°C and kept at −20°C. A primer pair specific for CL29 of *E. bulbosa*, an LTR Ty3/Gypsy-Tat repeat (Online Resource 1), was used to check whether the generation of chromosome-derived DNA was successful.

Next-Generation Sequencing and Sequences Analysis

Genomic DNA and chromosome microdissection-derived DNA were used for paired-end, 100-bp reads, and single-end, 250-bp reads, Illumina sequencing, respectively (Genbank Bioproject PRJNA549830). The repetitive fraction analysis was performed with 400 Mbp of reads of the genomic DNA (0.32× genome coverage) and 2,160 Mbp for each homolog of the large chromosome pair (~8× genome coverage). Sequenced reads were analyzed with the similarity-based read clustering method, implemented in the RepeatExplorer pipeline (Novak et al., 2013). Reads were filtered by quality with the default sets (quality cutoff value = 10, within a 95% of the bases in the sequence), and genomic paired-end reads were joined with the interlaced tool. For single-end reads, datasets from both chromosome types were assigned a unique identifier and joined into a single dataset with the concatenate tool. For both datasets (genomic and chromosomes), clustering was performed with a minimum overlap of 55% and a similarity of 90%. For sequences of microdissected chromosomes, three independent analyses were performed, using a different dataset of reads of the same sequencing, to confirm the proportions of each cluster on the two different homolog chromosomes. Repeat annotation and classification were performed for those clusters with an abundance >0.01%. For basic repeat classification, protein domains were identified using the tool "Find RT Domains" in RepeatExplorer (Novak et al., 2013). Searches for sequence similarity, using different databases (GenBank and TIGR), were performed, and graph layouts of individual clusters were examined using the SeqGrapheR program (Novak et al., 2013). Satellite DNAs were identified based on the graph layout and further examined using DOTTER (Sonnhammer and Durbin, 1995).

Amplification, Cloning, and Sequencing

The seven most abundant repeats of the total genome, three satellite DNAs (satDNA: Ebusat1, Ebusat2, Ebusat3) and four LTR-retrotransposons (LTR-RT) (Ty1/Copia-Maximus and -Tork and Ty3/Gypsy-Tat and -Chromovirus), were polymerase chain reaction (PCR) amplified. In addition, one satDNA (Ebusat4) from microdissected acrocentric chromosome DNA was also PCR amplified. For satellite DNAs, primers were designed, facing outward of the repeat unit, for the consensus sequences and from the region where most of the reads were conserved. LTR-RT–specific primers were designed to amplify the Integrase domain, commonly used in chromosome analyses with repetitive sequences, for being suggested as the most conserved domain within the domains of the retrotransposons (**Table S1**). The conserved region of the integrase domain was identified using the SeqGrapheR program (Novak et al., 2013). Forty nanograms of genomic DNA was used for all PCR reactions with 1× PCR buffer (Invitrogen), 2 mM $MgCl_2$, 0.1 mM of each dNTP, 0.4 μM each primer, 0.025 U *Taq* polymerase (Platinum *Taq* DNA polymerase; Invitrogen), and water. Polymerase chain reaction conditions were as follows 94°C 3 min, 30× (94°C 1 min, 55°C 1 min, 72°C 1 min), and 72°C 10 min. Polymerase chain reaction fragments were purified from a 1% agarose gel using

the AxyPrep DNA gel extraction kit (Axygen Biosciences) and cloned with the pGEM®-T Vector cloning system (Promega) using JM109 *Escherichia coli* high-efficiency competent cells (Promega), following manufacturer's instructions. One positive clone of each repetitive element was sequenced with a 3500 Genetic Analyzer Sanger sequencing platform at the Biosciences Center of the Federal University of Pernambuco for confirming its identity. Sequences were deposited in the GenBank database as MK228130-MK228135.

Chromosome Preparation and Fluorescence *In Situ* Hybridization

Cloned satellite DNAs, rDNAs, and retrotransposons were labeled with either Cy3-dUTP, Cy5-dUTP, or digoxigenin-11-dUTP by nick translation using a nick translation mix (Roche, Brazil) or with DNase I (0.002 U) and DNA polymerase (4 U) enzymes following Kato et al. (2004). 35S rDNA sites were detected with the pTa71 clone from *Triticum aestivum* (Gerlach and Bedbrook, 1979). Clone D2 from *Lotus japonicus* (Pedrosa et al., 2002) was used to detect the 5S rDNA.

Chromosomes were prepared from root tips collected from bulbs, pretreated in 0.02 M 8-hydroxyquinoline at 10°C for 24 h and fixed in ethanol: acetic acid (3:1 v/v) for 2 to 24 h at room temperature and stored at −20°C. Fixed root tips were digested with 2% cellulase-20% pectinase for 90 min at 37°C, and squashed in a drop of 45% acetic acid. Fluorescent *in situ* hybridization was performed as described by Pedrosa et al. (2002). The hybridization mix contained 50% (v/v) formamide, 10% (w/v) dextran sulfate, 2× SSC, and 5 ng/μL of each probe. Slides were denatured at 75°C for 5 min, and the final stringency of hybridization was 76%.

Images were captured using a Leica DM5500 B microscope with a Leica DFC345 FX coupled camera and the LAS AF software. Images were edited with Adobe Photoshop CS5.

Immunodetection of Histone Modifications and Active RNA Polymerase II

Antibodies for three different histone modifications were used: one euchromatic mark, rabbit anti-histone H3K4me3 (Abcam1012, diluted 1:300), and two pericentromeric chromatin marks: mouse anti-H3S10ph (Abcam 14955, diluted 1:2,000) and rabbit anti-H2AT120ph (Demidov et al., 2014, diluted 1:500). The latter antibody was developed for the same peptide as described in Dong and Han (2012). A mark for transcriptional activity was also applied: rat anti-RNAPIISer2ph (Millipore 04-1571, diluted 1:100). For immunostaining, root tips were pretreated with 2 mM 8-hydroxyquinoline for 24 h at 10°C. For RNAPIISer2 detection, nuclei were isolated from leaves. Both were fixed in freshly prepared 4% paraformaldehyde (dissolved in 1× PBS) for 30 min on ice and then washed three times for 15 min in 1× PBS on ice. Fixed root tips and leaves were chopped in a nuclei isolation buffer (Doležel et al., 2007) and filtered through a 30-μm nylon membrane. The cell suspension was used to prepare slides by centrifugation onto a microscopic slide at 2,000 rpm for 3 min (Shandon, CytoSpin3). Slides were incubated in 3% bovine serum albumin (BSA) for 30 min at 37°C. Primary antibodies, diluted in 1% BSA, were incubated overnight at 4°C and detected with Alexa 488–conjugated anti-rabbit (Dianova 711-545-152, diluted 1:200), Alexa 488–conjugated anti-mouse (Molecular probes A11001, diluted 1:200), Alexa 488–conjugated anti-rat (Dianova 112-545-167, diluted 1:200), or goat Cy3-conjugated anti-mouse (Dianova 115-165-062, diluted 1:300) antibodies in 1% BSA and incubated for 1 h at 37°C.

After immunostaining with H2AThr120ph and RNAPIISer2ph, fluorescence *in situ* hybridization (FISH) was performed subsequently to analyze the colocalization with the Ebusat2 satellite and the Ty3/Gypsy-Tat LTR-retrotransposon, respectively. Therefore, the slides were washed twice in 1× PBS, fixed in ethanol:acetic acid (3:1 v/v) for at least 24 h at room temperature in the dark, dehydrated, and prehybridized in 15 μL of DS20 (50% formamide, 10% dextran sulfate, 2× SSC) overnight at 37°C. Slides were washed in 2× SSC, dehydrated, and denatured in 0.2 N NaOH in 70% ethanol for 10 min at room temperature. Afterward, additional dehydration was performed, and the slides were hybridized with 50 ng of the probe in DS20 overnight at 37°C.

Images for histone modifications were captured using an epifluorescence microscope BX61 (Olympus) equipped with a cooled CCD camera (Orca ER, Hamamatsu). To achieve super-resolution for RNAPIISer2ph imaging, spatial structured illumination microscopy (3D-SIM) was applied using a 63/1.4NA Oil Plan-Apochromat objective of an Elyra PS.1 microscope system and the software ZEN (Carl Zeiss GmbH, Germany) (Weisshart et al., 2016). For histone modification marks, the fluorescence intensity was estimated along the chromosomes using the ImageJ software (Abràmoff et al., 2004). Intensity measurements were done at 10 or eight consecutive circles of ~50 to 60 pixels each, along with both homologs of the large chromosomes pair and one small chromosome pair, respectively. Five metaphases per mark were measured, and a mean of the measurements of each position along the chromosomes was calculated. We defined a ratio between the intensity of DAPI and the histone modification fluorescence along the chromosomes.

DNA Replication Analysis

DNA replication analysis was performed with the EdU kit (BCK-EdU 594-1, baseclick GmbH, Germany) following the manufacturer's protocol. Root tips were collected and incubated in a humid chamber with a filter paper embedded in an EdU solution for 3 h at room temperature for the incorporation of the dNTP analog. After incorporation, root tips recovered in water for 30 min, were pretreated in 8-hydroxyquinoline for 24 h at 10°C, and were fixed in ethanol:acetic acid (3:1 v/v). The cell walls were digested by treating the root tips with an enzyme mix of 0.7% cellulase R10 (Duchefa C8001), 1% pectolyase (Sigma P3026), and 1% cytohelicase (Sigma C8274) for 90 min at 37°C. The squashing of chromosomes was performed in a drop of 45% acetic acid. Then, the slides were immersed into liquid nitrogen to remove the coverslips. The slides were incubated with 3% BSA for 20 min at room temperature and then with the detection mixture for 30 min at room temperature in the dark. Sequential FISH with the Ebusat 1 and Ty3/Gypsy-Tat LTR-RT probes was performed as described above. Images were captured as described above.

RESULTS

Bimodal Karyotype of *E. bulbosa* Is Characterized by a Differential Chromosome-Type Specific Repeat Distribution

To characterize the repetitive DNA fraction of *E. bulbosa* (1C = 1.25 Gbp), its genome was sequenced at 0.32× genome coverage. Reads, comprising in total 400 Mbp, were grouped into 118,468 clusters containing from 2 to 90,876 reads. Clusters included 49.7% of all reads, with the major 257 clusters representing at least 0.01% of the genome each. The analysis revealed three major satellite DNA families (satDNAs), 11 transposable element families (LTR-retrotransposons and LINE), DNA transposons and ribosomal DNA sequences (**Table 1**). The largest clusters were identified as satDNAs: CL1 (Ebusat1) representing 4.99% of the genome with a complex repeat unit with variable length; CL2 (Ebusat2) representing 2.16% of the genome with a 261-bp repeat unit; and CL3 (Ebusat3) representing 1.4% of the genome. The LTR-like retrotransposons constituted approximately 28% of the genome, with the Ty3/gypsy superfamily exceeding 2.15-fold the genome proportion of the Ty1/copia superfamily. Within the former, Tat and Chromovirus were the only highly abundant lineages. Within Ty1/copia retrotransposons, eight lineages were identified, with Maximus and Tork being the most abundant.

The most abundant tandem repeat Ebusat1 localized at the large chromosome pair, mostly forming two interstitial blocks within the chromosome inversion: in the long arm of the acrocentric and in the short arm of the metacentric chromosome. A third weaker signal was observed outside of the inversion, in a distal position on the long arm of both chromosomes (**Figure 1**). Ebusat2 was located in the centromeric region of all chromosomes, showing similar hybridization intensities among all chromosomes (**Figure 1**). Both satDNAs colocalized with DAPI$^+$ bands, except for the third smaller band of Ebusat1. For Ebusat3, no hybridization signals were detected, although the tandem repeat nature was confirmed by Dotter, and the amplified PCR fragments showed a ladder-like pattern.

All four most abundant LTR-retrotransposon lineages showed a high accumulation on the large chromosome pair, with dispersed labeling along the entire chromosome, except for the proximal region in chromosome I corresponding to the 35S ribosomal DNA (**Figures 1** and **S1**). Three of them (Tat, Chromovirus, and Tork) showed pericentromeric labeling on the small chromosomes, while Maximus showed a more scattered distribution, enriched proximally on small chromosomes (**Figures 1** and **S1**). These results demonstrate that both large and small chromosome types share the same repetitive DNA sequences, except for Ebusat1, which is present only in the large pair. However, except for Ebusat2, each chromosome type displays a specific chromosomal distribution for these repetitive sequences, with a higher abundance of repeats in the large chromosome pair.

In order to compare the DNA composition of the homologs of the large chromosome pair, microdissection was performed, and the isolated chromosome-specific DNA was sequenced. About 2,160 Mbp of the sequence reads from each homolog (within and outside the inversion) were used for comparative *in silico* analyses, representing ~8× coverage for each chromosome. The reads were grouped into 13,279 clusters containing from 2 to 80,108 reads. The major 267 clusters, with a minimum of 0.01% of the chromosome DNA proportion, represented ~27% of the chromosome DNA. Further, two different clustering analyses were run using two different sets of sequencing data each, showing similar results. The analysis revealed four satDNA families, Ebusat1, Ebusat2, and Ebusat3, as well as one previously undetected one, Ebusat4, seven LTR-RT lineages previously characterized in the genomic analysis, LINEs, DNA transposons, and ribosomal DNA sequences. Except for Ebusat2, the amount of all satDNAs seems to be higher in the larger chromosomes, in agreement with the distribution of Ebusat1 in this chromosome pair. In contrast, the LTR-RT lineages, except TAR in the metacentric chromosome, showed a similar or lower proportion within chromosome I than the total genome proportions. This may indicate a preferential amplification of the most abundant tandem repeats during the process of whole-genome amplification after microdissection (**Table S2**). Ebusat4 was mostly localized within the chromosome inversion, associated with the interstitial bands of the large chromosome pair as Ebusat1, in the long arm of the acrocentric and in the short arm of the metacentric chromosome (**Figure 1**). The hybridization signals were similar in both chromosomes of the large pair in disagreement with the differential proportion of Ebusat4 observed after analyzing the DNA composition of microdissected chromosomes. This difference could be explained by the uneven amplification of microdissected DNA (**Table S2**).

These results confirmed the presence of the major repetitive families in the large chromosome pair, as well as the presence of one new satDNA, Ebusat4, clearly enriched in both homologs of this pair. This distribution of repeats was not influenced by the occurrence of the nonrecombining inverted region of the large chromosome pair.

TABLE 1 | Proportion (%) of repetitive elements present in the total genome of *E. bulbosa*.

Repetitive element		Total genome (%)
Satellite	Ebusat 1	4.99
	Ebusat 2	2.16
	Ebusat 3	1.40
	Ebusat 4	—*
LTR-Ty3/Gypsy	Tat	14.26
	Chromovirus	1.76
LTR-Ty1/Copia	Maximus	1.78
	Tork	1.46
	TAR	1.23
	AleII	1.19
	Angela	0.98
	Ivana/Oryco	0.31
	Bianca	0.48
	AleI.Retrofit	0.031
Unclassified LTR		6.39
LINE		0.19
DNA Transposons		2.08
rDNA		0.53
Microsatellite		0.33
Unclassified		1.88
Total		43.04

**Found only in the acrocentric chromosome via in silico analysis.*

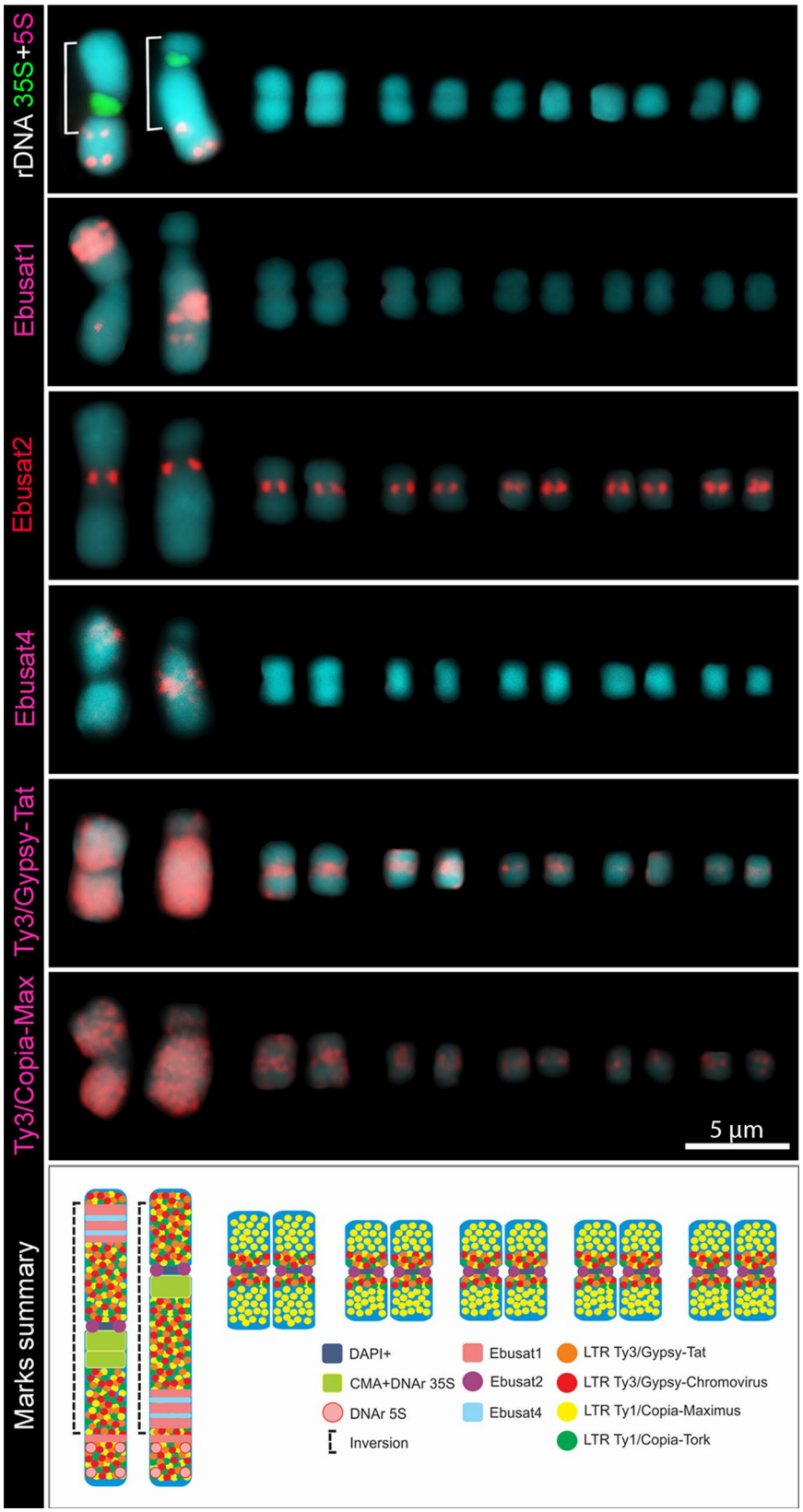

FIGURE 1 | Comparative karyograms showing the distribution of repetitive elements in the *E. bulbosa* chromosomes. The brackets indicate the inverted chromosome region. The scheme below summarizes the repeat distribution.

The Chromatin Composition Differs Between Large and Small Chromosomes

As the identified repetitive DNA families showed a distinct chromosome type–specific distribution, we investigated the distribution of a subset of posttranslational histone modifications to understand whether the observed DNA composition differences are associated with a different chromatin organization.

The small chromosomes of *E. bulbosa* showed strong labeling along the chromosome arms by the euchromatin histone mark H3K4me3, with weaker labeling in the pericentromeric regions. In contrast, both large homologs showed weaker labeling along the entire chromosomes (**Figure 2**). Histone modification marks for pericentromeric regions displayed contrasting results. While all chromosomes showed a pericentromeric distribution for phosphorylated H3 at serine 10, the large chromosome pair was strongly phosphorylated with H2AThr120ph at the proximal chromosome regions, flanking the pericentromeres and including the rDNA site. Toward distal arm regions, the phosphorylation of H2AThr120 gradually decreased. All small chromosomes showed a weak phosphorylation at the pericentromeres and weaker signals on the proximal and distal regions (**Figure 2** and **Movies S1**, **S2**). The different distribution of both histone modifications and between the small and large chromosomes was confirmed by fluorescence intensity measurements. Subsequent FISH with the centromeric Ebusat2 satellite repeats confirmed its localization between both proximal H2AThr120ph positive regions (**Figure 2**).

We also analyzed whether different transcriptional activities between both chromosome types exist. Therefore, we applied specific antibodies against RNA polymerase II phosphorylated at serine 2 (RNAPIISer2ph) as a mark for transcriptional activity. Antibodies specific for the phosphorylation state of a peptide allow the discrimination between active and inactive RNAPII (Bourdon et al., 2012). For the elongation step of transcription, phosphorylation at serine 2 is required (Ni et al., 2004). RNAPIISer2 displayed a dispersed distribution in interphase nuclei, with a certain accumulation within the nucleus interior (**Figure 3**). Subsequent FISH with LTR-RT Ty3/Gypsy-Tat showed that the regions enriched with these repetitive elements, which are more abundant in the large chromosome pair and mostly present at the periphery of interphase nuclei, were only

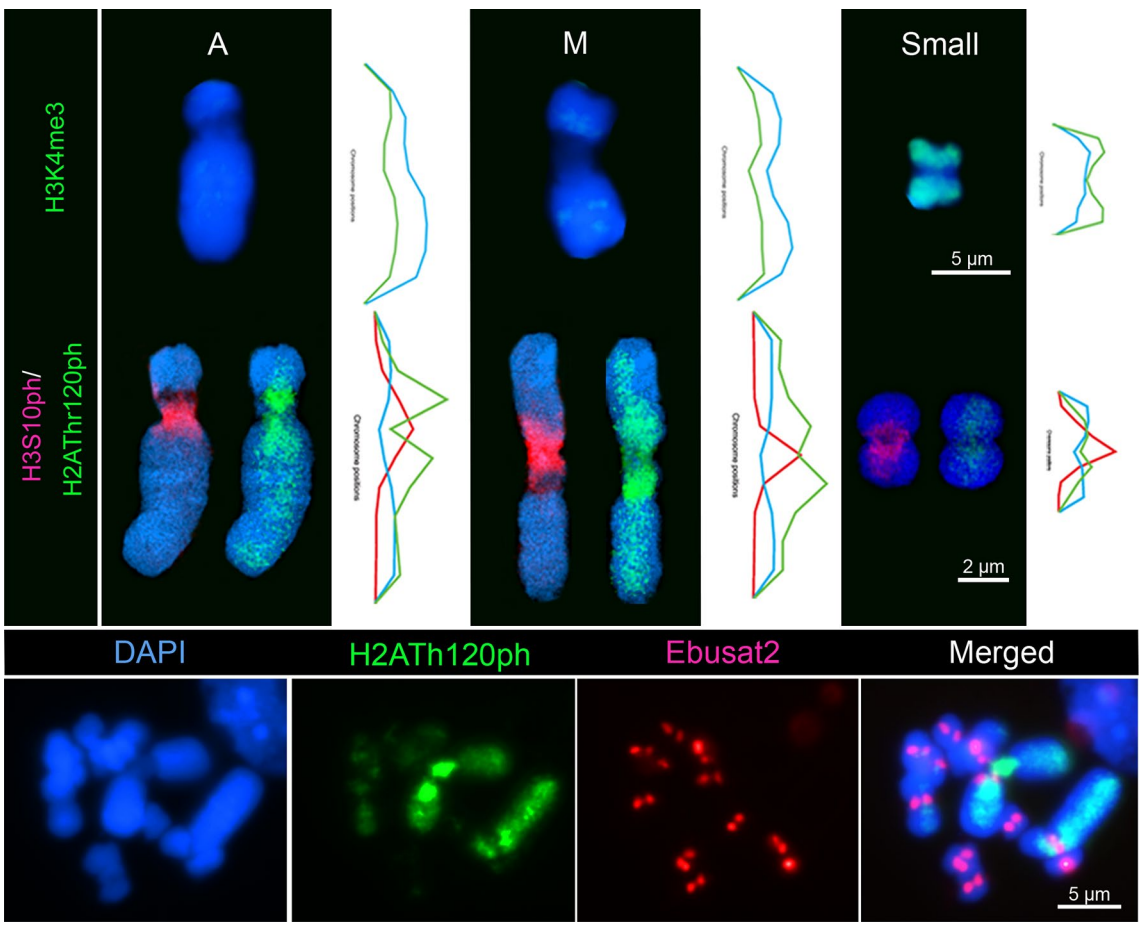

FIGURE 2 | Chromatin organization differs between the large and small chromosomes of *E. bulbosa*. Top: Distribution of the euchromatin mark H3K4me3 on the large acrocentric (A), metacentric (M), and a small chromosome, each with a fluorescence intensity histogram of whole chromatin (DAPI) and H3K4me3 staining. Middle: Distribution of the pericentromeric histone modification marks H3S10ph and H2AThr120ph on the three chromosome types, with their respective fluorescence intensity histograms. Bottom: H2AThr120ph distribution along the chromosome arms and the Ebusat2 satDNA localization at the centromeres.

seldom associated with active RNAPII. In contrast, regions free of these repetitive elements, mostly in the central nuclear region, were enriched with RNAPIISer2ph (**Figure 3**).

Altogether, these results demonstrate that chromatin composition and transcriptional activity differ between large and small chromosomes. These differences are possibly a consequence of the different repetitive sequence composition in both chromosome types.

Replication Dynamics Differ Between Large and Small Chromosomes

As large and small chromosomes present a different repeat and chromatin composition, replication analysis was performed to uncover a potentially different replication behavior of both chromosome types. Compared to the small chromosomes, the large chromosome pair incorporated EdU at a different time. Furthermore, incorporation into the large chromosomes was observed as bands along the chromosome arms (**Figure 4A**). Some cells showed this banding pattern together with the incorporation at the pericentromeric regions on all small chromosomes. Subsequent FISH with Ebusat1 and the Ty3/Gypsy-Tat element showed a partial colocalization between the EdU incorporation and these repeats in metaphase chromosomes as well as in interphase nuclei, mostly at the nucleus periphery (**Figures 4B**, **C**). These data indicate that the different chromosome structures also influence the replication timing, likely as a consequence of the presence of distinct repetitive sequences and chromatin compositions in the large and small chromosomes.

Chromosomal Distribution of Repeats Is Similar in *E. latifolia*

To analyze whether the subgenome repeat distribution is a specific feature of *E. bulbosa*, or a bimodal karyotype feature in the genus, we hybridized the major repeat sequences of *E. bulbosa* to the chromosomes of the sister species *E. latifolia* (2C = 1.4 pg). Both 35S and 5S rDNA sites were located on both of the large acrocentric homologs, with the 35S sites at the pericentromeric regions and the 5S sites at the interstitial regions of the long arms (**Figure 5**). The position of the 35S rDNA is similar to that in the acrocentric chromosome of *E. bulbosa*. The 5S rDNA sites are also duplicated in closer proximity as in *E. bulbosa*. However, in one of these homologs, both signals are located more distally, suggesting a paracentric inversion present in heterozygosity. The most abundant repetitive elements of *E. bulbosa*, both satDNAs (Ebusat1 and Ebusat2), the Ty3/Gypsy-Tat and Ty1/Copia-Maximus LTR-RT lineages, showed a similar distribution in *E. latifolia* as it was found in *E. bulbosa* chromosomes. Ebusat1 showed two distal bands at the long arms of the large acrocentric pair, one stronger and one weaker band, with inverted orientation between the homologs, confirming the paracentric inversion (brackets in **Figure 5**). Ebusat2 displayed a centromeric distribution in all chromosomes of the complement. Ty3/Gypsy-Tat and Ty1/Copia-Maximus LTR-RT exhibited an accumulation on the large acrocentric pair, similar as observed in *E. bulbosa*, irrespective of the position of the inversion. Ty1/Copia-Maximus was uniformly dispersed along the largest acrocentric chromosome pair. It was proximally distributed

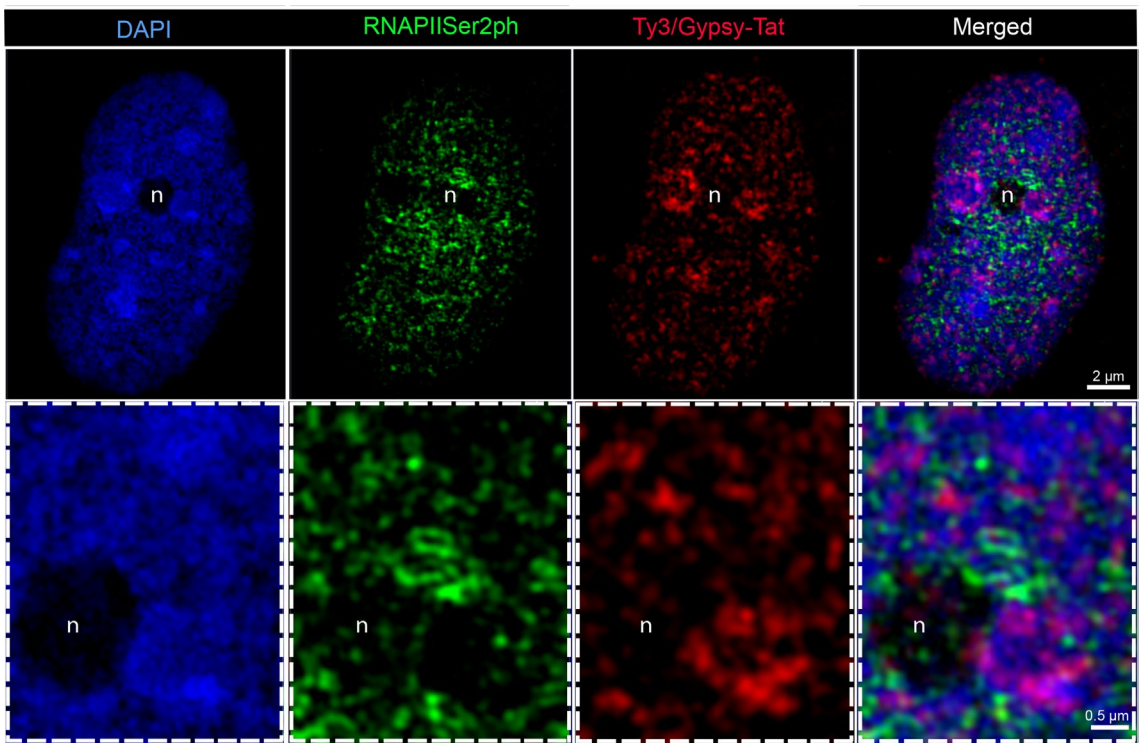

FIGURE 3 | Distribution of RNAPIISer2ph and Ty3/Gypsy-Tat repetitive elements in *E. bulbosa* interphase nuclei, analyzed by super-resolution microscopy (SIM). Active RNAPII phosphorylated at serine 2 (RNAPIISer2ph) accumulates especially at LTR Ty3/Gypsy-Tat elements localized around the nucleolus (n).The enlarged region below is indicated by dashed rectangle.

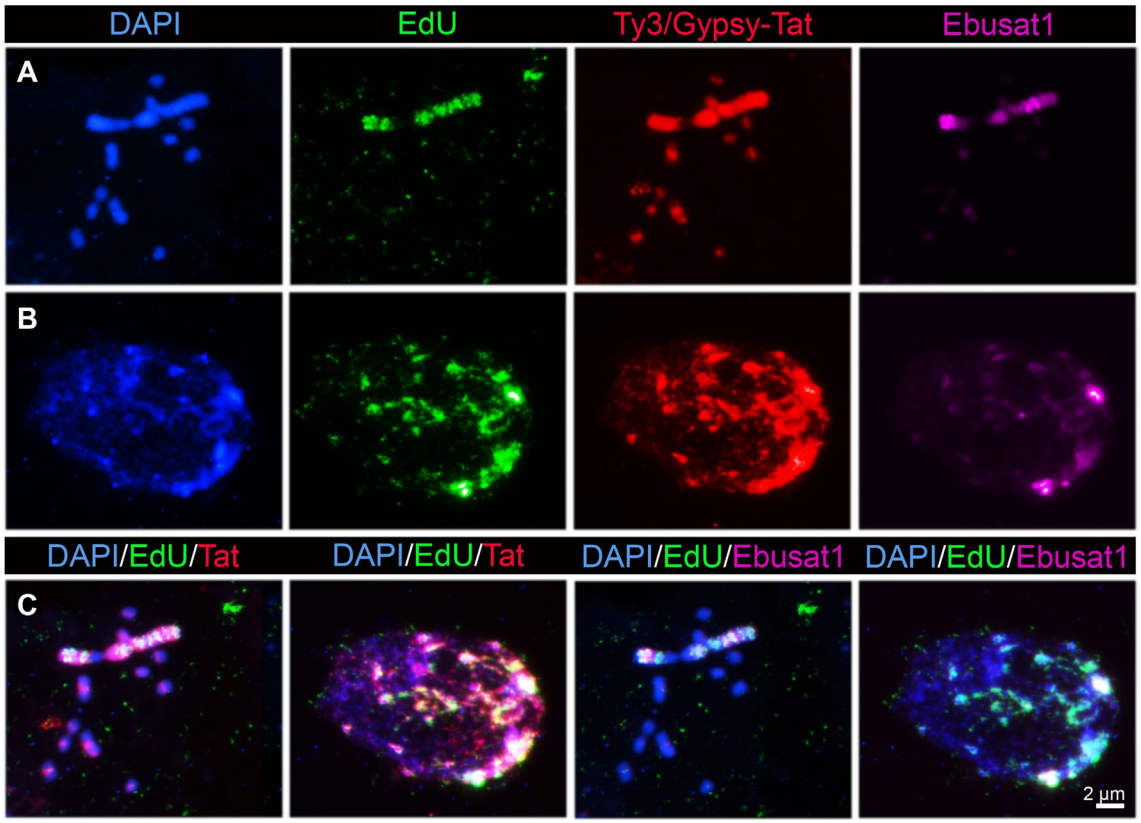

FIGURE 4 | Replication dynamics differs between the large and small chromosomes of *E. bulbosa*. Metaphase chromosomes **(A)** and interphase nuclei **(B)** after EdU incorporation followed by LTR Ty3/Gypsy-Tat and Ebusat1 DNA localization. Colocalization is shown in **(C)**.

on all small chromosomes, but denser at the pericentromeric regions. Similarly, Ty3/Gypsy-Tat was uniformly dispersed along the large pair and showed a pericentromeric distribution in the small chromosomes (**Figure 5**). Chromosome signals from both satDNAs and LTR-RT were weaker in *E. latifolia* than in *E. bulbosa*. This suggests differences of these repeats families between both species, likely due to differences in sequence similarity and/or abundance. However, both species have the same distribution of repeat sequences, indicating that the chromosome set–specific repeat distributions are conserved in the genus and was not influenced by species-specific chromosome rearrangements. We suggest that this type of chromosome set–repeat distribution is a characteristic of bimodal karyotypes.

DISCUSSION

Chromosomal Distribution of Repetitive DNA in *E. bulbosa* Reflects Its Bimodal Karyotype

Although repetitive sequences are enriched in the large chromosomes of several bimodal plant species (de la Herrán et al., 2001; Pedrosa et al., 2001), none of them showed a high degree of accumulation of LTR retroelements, together with satDNA, as seen in *Eleutherine*. The repetitive genome fraction of *E. bulbosa* is mainly composed of LTR retroelements, with more abundant Ty3/Gypsy-like than Ty1/Copia-like elements. Within eukaryote genomes, LTR-RT Ty3/Gypsy and Ty1/Copia are most abundant in plants (Kumar and Bennetzen, 1999), with the Ty3/Gypsy elements being the most abundant in the majority of angiosperm families (Weiss-Schneeweiss et al., 2015). All LTR retrotransposons were highly accumulated on the large chromosome pair of *Eleutherine*, showing a uniformly dispersed distribution, similar to the distribution observed for several mobile elements in large genomes (Lamb, 2006; Lamb et al., 2007). In contrast, the small chromosomes of *Eleutherine* accumulated the LTR retroelements only at the pericentromeric regions. This restricted distribution of repetitive sequences is typical for small genome species, e.g. in *Arabidopsis* (Heslop-Harrison et al., 1997), *Phaseolus* (Padeken et al., 2015), and *Brachiaria* (Santos et al., 2015). Thus, the differential distribution of retroelements appears to be related to the bimodal condition and is not influenced by the nonrecombining chromosome inversions in these species. The large and the small chromosome sets of *Eleutherine* constitute two distinct subgenomes with respect to the retroelement distribution.

The chromatin organization is partly similar in *Eleutherine* and bird species. The microchromosomes of birds are enriched in genes and hyperacetylated at histone H4K5, while macrochromosomes are gene-poor and hypoacetylated at histone H4K5 (McQueen et al., 1998). However, unlike the large chromosomes of *E. bulbosa*,

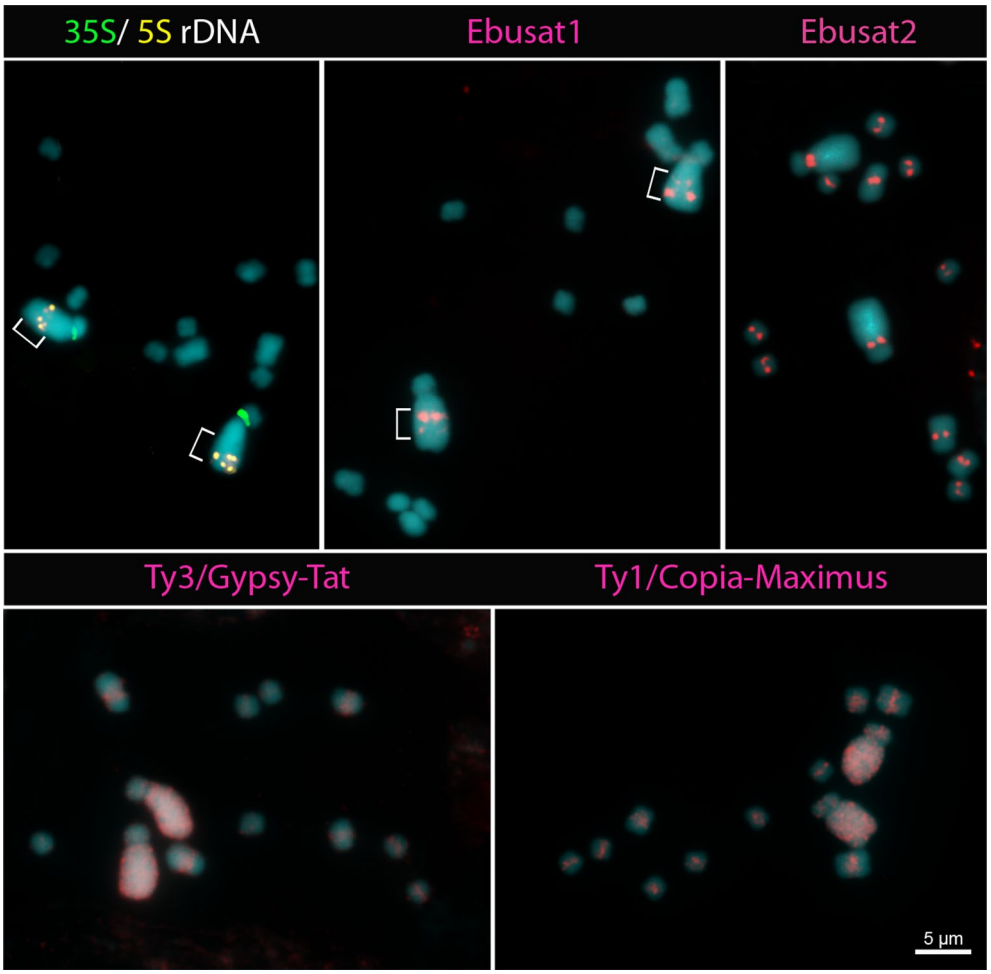

FIGURE 5 | Distribution of repetitive elements in *E. latifolia*. The localization of the 5S, 35S ribosomal DNA (rDNA) sites and the Ebusat1 repeat indicates a paracentric inversion at the long arm of the large chromosome pair (brackets).

bird macrochromosomes are poor in heterochromatin, with less 5-methylcytosine-rich regions than microchromosomes, probably related to CpG islands associated with the high gene content of the microchromosomes (Schmid and Steinlein, 2017). Nevertheless, some transposable elements show the differential chromosomal distribution in bimodal karyotypes of different animal groups. In birds, CR1-like retroelements are spread over nearly all chromosomes but have a higher density on macrochromosomes with a particular banding pattern (Coullin et al., 2005). The *Rex*6 transposable element is also densely distributed in the largest chromosome pairs in several species of *Podocnemis* turtles. It was suggested that *Rex*6 may influence the genomic structure, interfering with gene regulation (Noronha et al., 2016).

Retrotransposons are abundant components of large chromosomes in some plants with bimodal karyotypes, but differentiation of chromosome sets is less evident. The large chromosomes of South American *Hypochaeris* species are enriched by Ty1-Copia LTR-RT along their entire lengths, while small chromosomes lacked those elements in most of the long arms. However, this pattern is not maintained in other species of the genus, mainly not in Old World species (Morocco and Croatia; Ruas et al., 2008). On the other hand, in three species of the genus *Alstroemeria* with asymmetric chromosomes, the distribution of a Ty1-Copia like LTR-RT was equally dispersed over all chromosomes (Kuipers et al., 1998).

Here, two major satellite DNAs represent a large proportion of the genome (~7%). Such a high proportion of one or two satDNA families is unusual in most plant species. Usually, many satDNA families with a low abundance or few satDNA families with slightly larger genome abundance were observed (Hemleben et al., 2007; Steflova et al., 2013). Contrary to what was suggested (Melters et al., 2013), the most abundant satDNA from *E. bulbosa* is not centromeric but accumulated interstitially exclusively in the large chromosome pair. The second most abundant satDNA is present in all centromeric regions. However, whether this location is associated with centromere functions has not yet been clarified. As in *Eleutherine*, large chromosomes of bimodal karyotypes in some plants accumulate satDNA. In *Muscari* species, a large chromosome specific satDNA found in *M. comosum* is conserved within the genus and has been proposed to mediate the increase of karyotype asymmetry (de la Herrán et al., 2001). One of the major components of intercalary heterochromatin on the large

chromosomes of *O. longibracteatum* was also a specific satDNA (Pedrosa et al., 2001). Therefore, as clearly observed in *Eleutherine*, large and small chromosomes within a bimodal karyotype can maintain a differential DNA composition. This can cause structural and functional differences between these subgenomes.

Large and Small Chromosomes Differ in Heterochromatin, Gene Content, and Replication Timing

Chromatin-associated H3K4me3 and H2AThr120ph histone marks differentiate the large and small chromosomes in *E. bulbosa*. The dense distribution of H3K4me3 along all small chromosomes is associated with euchromatin, as it was observed also for other histone modification marks indicating euchromatin in the small chromosomes of this and other species with small genomes (Feitoza and Guerra, 2011). The less intense distribution of H3K4me3 on the large chromosomes is likely associated with a higher proportion of heterochromatin, as revealed by the repeat distribution of in the present work, as well as by 5-methylcytosine and H3K9me2 localization (Feitoza and Guerra, 2011).

The so-called universal histone modification mark for the pericentromeric region, H2AThr120ph (Dong and Han, 2012, Demidov et al., 2014), did not exclusively label the pericentromeric region in *E. bulbosa* as defined by the H3S10ph mark. The large chromosomes also showed intense H2ATh120ph signal along their lengths. This suggests that the proximal chromatin composition may differ between both chromosomes sets and that it is rather related to the chromosome structure than to centromere function. A comparable atypical distribution of this histone mark was found for the large Y chromosome of *Coccinia grandis* (Sousa et al., 2016).

Large and small subgenomes of *Eleutherine* also differ in their replication timing. Large chromosomes showed a banding pattern after EdU incorporation. This suggests that, within these chromosomes, there are regions that perform replication at a different time than those of the small chromosomes. This could be due to the high repeat composition within large chromosomes, as confirmed by the colocalization of some repetitive elements with EdU bands. Differential replication timing was also reported for chicken macrochromosomes displaying late replication and microchromosomes showing early replication, both related to their different gene content (McQueen et al., 1998). This assumption could also be valid for *E. bulbosa*, because the distribution of active RNAPII in interphase nuclei suggests chromosome regions with different transcription activities. Chromatin enriched in active RNAPII did not colocalize to repeat-rich nuclei regions and is possibly associated with small chromosomes. In addition, a weaker RNAPIISer2ph labeling and more repetitive DNA were observed at the nuclear periphery, possibly associated with large chromosomes. Together, these data suggest that the large chromosomes of *Eleutherine* are mainly composed of heterochromatin and heterochromatin-like, early-condensing euchromatin (Guerra, 1988; Feitoza and Guerra, 2011). They have a lower gene density and partially replicate later. In contrast, small chromosomes are composed mainly of euchromatin, are gene-rich, and replicate earlier. Thus, the large and small chromosomes represent structurally and functionally differentiated subgenomes within the same species.

The Bimodal Chromosomal Organization Is Maintained Within the Sister Species *E. latifolia*

Both investigated *Eleutherine* species display a similar repeat distribution, indicating that these repeats originated and underwent a chromosome-type specific accumulation before the separation of these two species. The newly discovered chromosome inversion in *E. latifolia*, a paracentric inversion in the long arm of one homolog of the large chromosome pair, also involved Ebusat1. One breakpoint is apparently close to the breakpoint of the *E. bulbosa* pericentric inversion. Both events may suggest that this chromosome pair is prone to chromosome rearrangements, possibly due to its highly repetitive sequence content. Repetitive sequences are potential sites for chromosome rearrangements through homology-directed recombination repair using ectopic homologous repeats as a template (Charlesworth et al., 2005). Although vegetative reproduction in *Eleutherine* could be responsible for the maintenance of both inversions in heterozygous condition, the recurrent rearrangements in this chromosome pair are possibly associated with a strong purifying selection, with lethality under homozygous conditions. This may lead to permanent heterozygosity of this chromosome pair, even in individuals propagated *via* seeds. First, we hypothesized that the chromosomal inversion in *E. bulbosa* could have led to repetitive sequence accumulation on the large chromosome pair. However, since the repeat distribution is uniform along the entire large chromosome pair, even outside the inverted region, and is similar in both *Eleutherine* species, it is more likely that it is related to the bimodal structuration and function of these karyotypes, rather than the consequence of chromosome inversions.

How could a Bimodal Karyotype Evolve in *Eleutherine*?

Different types of bimodal karyotypes exist, since not in all bimodal species different chromatin and sequence composition between large and small chromosomes exist, as presented for *Eleutherine*. There are different hypotheses for the origin of bimodal karyotypes. Interspecific hybridization appears not to apply to *Eleutherine*, since the bimodal karyotype is characteristic for the whole tribe Tigridieae, with no evidence of allopolyploidy for its origin (Moraes et al., 2015).

Although paleo-alloploidization or ancestral chromosomal rearrangements cannot be completely ruled out, our data indicate an increase of the size of one set of chromosomes due to a differential repetitive sequences accumulation. This phenomenon was suggested for species of the *Muscari* genus. Differential accumulation of one satDNA was associated to the increase in size of a subset of chromosomes (de la Herrán et al., 2001). It is possible that an initial random accumulation of repetitive sequences in one chromosome pair in *Eleutherine* ancestral gradually increased its size and led to its chromatin differentiation, differential replication, and transcription behavior. Consequently, this may lead to a higher repetitive sequence accumulation and divergence of different repetitive families in the large chromosomes, which is less deleterious than in the more gene-rich small chromosomes (Vosa, 2005; see **Figure 6**). In both *Eleutherine* species, this repetitive environment provided a special background for chromosomes inversion.

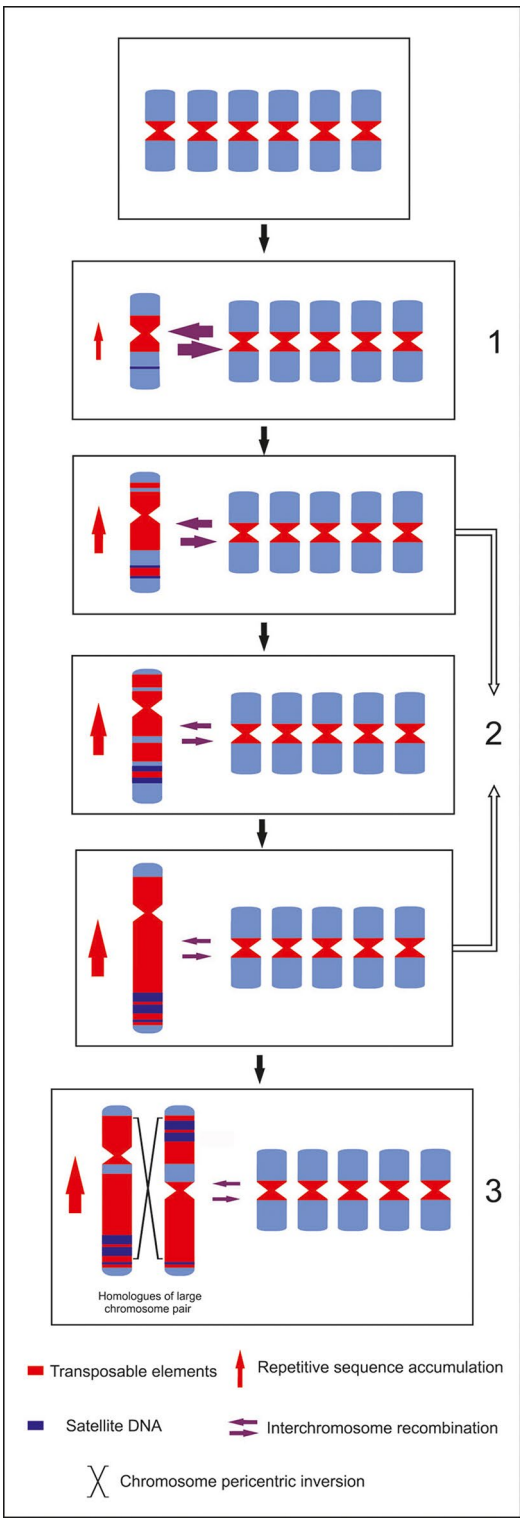

FIGURE 6 | Bimodal karyotype evolution model for *E. bulbosa*. From a symmetric to a bimodal karyotype: (1) random increase of retroelements and satellite DNA; (2) preferential maintenance of retroelements accumulating in the large-chromosome subgenome with increasing chromosome size and decreasing gene density; (3) pericentric inversions lead to the heteromorphic homologs on the large chromosome pair, maintained by recessive lethality with selection advantages in the heterozygote status.

In chicken, the microchromosomes present an increased rate of meiotic recombination compared to the macrochromosomes, perhaps due to a gene composition that favors meiotic recombination. This could lead to evolutionary pressure for an increase of gene density on small chromosomes (Smith et al., 2000; Rodionov et al., 2002). The negative feedback between repeats and recombination may be intensified by intrachromosomal rearrangements, also contributing to suppress recombination and subgenome differentiation in bimodal karyotypes, as our results suggested for *Eleutherine*. The possible gradual increase of meiotic recombination within the small chromosome set and, in consequence, a decrease of recombination between small and large chromosome sets could also favor the differentiation of the structure and function of both chromosome set and the maintenance of the bimodal karyotypes over time.

DATA AVAILABILITY STATEMENT

The raw sequencing data is available in Genbank Bioproject PRJNA549830. The sequences of the repeats found in this work is available in Genbank database as MK228130-MK228135.

AUTHOR CONTRIBUTIONS

MB, AH, and AP-H designed the study. MB and MV carried out the bioinformatics studies. SD carried out the chromosome microdissection with the help of MB. VS carried out the super-resolution microscopy. MB carried out the cytogenetic, molecular and flow cytometry experiments, interpreted the results, and wrote the manuscript with the help of AH and AP-H. All authors read and approved the final manuscript.

FUNDING

We thank the Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq Nº 310804/2017-5), and the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany, for financial support. This study was supported in part by the Coordenacão de Aperfeicoamento de Pessoal de Nıvel Superior–Brasil (CAPES, Finance Code 001) and MB received a scholarship CAPES Nº 99999.003674/2015-00.

ACKNOWLEDGMENTS

We are indebted to Dr Gustavo Souza, from the Federal University of Pernambuco, Brazil for providing *E. bulbosa* plants; to Dr Guadalupe Munguía Linno from Guadalajara University, Mexico, and Dr Eliane Kaltchuk dos Santos from Universidade Federal do Rio Grande do Sul, for providing *E. latifolia* seeds; to Prof Ingo Schubert for discussion; and to Dr Axel Himmelbach (IPK, Gatersleben) for Illumina sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01170/full#supplementary-material

VIDEO S1 | Movie showing the differential chromatin organization within the centromeric region of E. bulbosa chromosomes mitotic complement. Pericentromeric histone modification marks H3S10ph (red) and H2ATh120ph (green), on both chromosome types. Chromosomes are stained with DAPI (blue).

VIDEO S2 | Movie showing the differential chromatin organization within the centromeric region of E. bulbosa metacentric chromosome. Pericentromeric histone modification marks H3S10ph (red) and H2ATh120ph (green). Chromosomes are stained with DAPI (blue).

REFERENCES

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image processing with Image*J. Biophotonics Intern.* 11 (7), 36–42.

Akemine, T. (1935). Chromosome studies on *Hosta* I. The chromosome numbers in various species of *Hosta. J. Fac. Sci. Hokkaido Imp. Univ. Ser. 5 Bot.* 5, 25–32.

Brandham, P. E., and Doherty, M. J. (1998). Genome size variation in the aloaceae, an angiosperm family displaying karyotypic orthoselection. *Ann. Bot.* 82, 67–73. doi: 10.1006/anbo.1998.0742

Bourdon, M., Pirrello, J., Cheniclet, C., Coriton, O., Bourge, M., Brown, S., et al. (2012). Evidence for karyoplasmic homeostasis during endoreduplication and a ploidydependent increase in gene transcription during tomato fruit growth. *Development* 139, 3817–3826. doi: 10.1242/dev.084053

Burt, D. W. (2002). Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* 96, 97–112. doi: 10.1159/000063018

Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95, 118–128. doi: 10.1038/sj.hdy.6800697

Coullin, P., Bed'Hom, B., Candelier, J. J., Vettese, D., Maucolin, S., Moulin, S., et al. (2005). Cytogenetic repartition of chicken CR1 sequences evidenced by PRINS in Galliformes and some other birds. *Chromosome Res.* 13, 665–673. doi: 10.1007/s10577-005-1004-7

de la Herrán, R., Robles, F., Cuñado, N., Santos, J. L., Ruiz Rejón, M., Garrido-Ramos, M. A., et al. (2001). A heterochromatic satellite DNA is highly amplified in a single chromosome of *Muscari* (Hyacinthaceae). *Chromosoma* 110, 197–202. doi: 10.1007/s004120000115

D'Emerico, S., Grünanger, P., Scrugli, A., and Pignone, D. (1999). Karyomorphological parameters and C-band distribution suggest phyletic relationships within the subtribe Limodorinae (Orchidaceae). *Plant Syst. Evol.* 217, 147–161. doi: 10.1007/BF00984927

Demidov, D., Schubert, V., Kumke, K., Weiss, O., Karimi-Ashtiyani, R., Buttlar, J., et al. (2014). Anti-phosphorylated histone H2AThr120: a universal microscopic marker for centromeric chromatin of mono- and holocentric plant species. *Cytogenet. Genome Res.* 143, 150–156. doi: 10.1159/000360018

Doležel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2, 2233. doi: 10.1038/nprot.2007.310

Dong, Q., and Han, F. (2012). Phosphorylation of histone H2A is associated with centromere function and maintenance in meiosis. *Plant J.* 71, 800–809. doi: 10.1111/j.1365-313X.2012.05029.x

Dreissig, S., Fuchs, J., Cápal, P., Kettles, N., Byrne, E., and Houben, A. (2015). Measuring meiotic crossovers *via* multi-locus genotyping of single pollen grains in barley. *PLoS One* 10 (9), e0137677. doi: 10.1371/journal.pone.0137677

Feitoza, L., and Guerra, M. (2011). Different types of plant chromatin associated with modified histones H3 and H4 and methylated DNA. *Genetica* 139, 305–314. doi: 10.1007/s10709-011-9550-8

Feliner, G. N., and Rosselló, J. A. (2012). Concerted evolution of multigene families and homoeologous recombination, in *Plant genome diversity*, vol. 1. (Vienna: Springer), 171–193. doi: 10.1007/978-3-7091-1130-7_12

Fentaw, E., Dagne, K., Rønsted, N., Demissew, S., and Grace, O. M. (2013). Karyotypes in Ethiopian *Aloe* species (Xanthorrhoeaceae: Asphodeloideae). *Kew Bull.* 68, 599–607. doi: 10.1007/s12225-013-9475-8

Fiorin, F. G., Ruas, P. M., Ortiz, M. A., Urtubey, E., Matzenbacher, N. I., and Ruas, C. F. (2013). Karyotype studies on populations of two *Hypochaeris* species (*H. catharinensis* and *H. lutea*), Asteraceae, endemics to southern Brazil. *Genet. Mol. Res.* 12, 1849–1858. doi: 10.4238/2013.January.4.4

Gerlach, W. L., and Bedbrook, J. R. (1979). Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Res.* 7, 1869–1885. doi: 10.1093/nar/7.7.1869

Goldblatt, P., and Snow, N. (1991). Systematics and chromosome cytology of *Eleutherine* Herbert (Iridaceae). *Ann. Mo. Bot. Gard.* 78, 942–949. doi: 10.2307/2399735

Gole, J., Gore, A., Richards, A., Chiu, Y. J., Fung, H. L., Bushman, D., & Zhang, K. (2013). Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat. Biotechnol.* 31 (12), 1126. doi: 10.1038/nbt.2720

Griffin, D. K., Romanov, M. N., O'Connor, R., Fowler, K. E., and Larkin, D. M. (2015). Avian cytogenetics goes functional, in: Third Report on Chicken Genes and Chromosomes 2015. *Cytogenet. Genome Res.* 145 (2), 100–105. doi: 10.1159/000430927

Guerra, M. (1988). Mitotic and meiotic analysis of a pericentric inversion associated with a tandem duplication in *Eleutherine bulbosa. Chromosoma* 97, 80–87. doi: 10.1007/BF00331797

Guerra, M. (1991). *cis*-Acting regulation of NOR cistrons in *Eleutherine bulbosa* (Iridaceae). *Genetica* 83, 235–241. doi: 10.1007/BF00126229

Hemleben, V., Kovarik, A., Torres-Ruiz, R. A., Volkov, R. A., and Beridze, T. (2007). Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. *Syst. Biodivers.* 5, 277–289. doi: 10.1017/S147720000700240X

Heslop-Harrison, J. S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A. V., Alkhimova, E. G., et al. (1997). The chromosomal distributions of Ty1-*copia* group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* 100, 197–204. doi: 10.1023/A:1018337831039

Kato, A., Lamb, J. C., and Birchler, J. A. (2004). Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl. Acad. Sci. U. S. A.* 101 (37), 13554–13559. doi: 10.1073/pnas.0403659101

Kuipers, A. G., Heslop-Harrison, P. J., and Jacobsen, E. (1998). Characterisation and physical localisation of Ty1-copia-like retrotransposons in four *Alstroemeria* species. *Genome* 41, 357–367. doi: 10.1139/g98-048

Kumar, A., and Bennetzen, J. L. (1999). Plant retrotransposons. *Annu. Rev. Genet.* 33, 479–532. doi: 10.1146/annurev.genet.33.1.479

Lamb, J. C. (2006). Retroelement genome painting: cytological visualization of retroelement expansions in the genera *Zea* and *Tripsacum. Genetics* 173 (2), 1007–1021. doi: 10.1534/genetics.105.053165

Lamb, J. C., Meyer, J. M., Corcoran, B., Kato, A., Han, F., and Birchler, J. A. (2007). Distinct chromosomal distributions of highly repetitive sequences in maize. *Chromosome Res.* 15, 33–49. doi: 10.1007/s10577-006-1102-1

Masabanda, J. S., Burt, D. W., O'Brien, P. C., Vignal, A., Fillon, V., Walsh, P. S., et al. (2004). Molecular cytogenetic definition of the chicken genome: the first complete avian karyotype. *Genetics* 166, 1367–1373. doi: 10.1534/genetics.166.3.1367

McKain, M. R., Wickett, N., Zhang, Y., Ayyampalayam, S., McCombie, W. R., Chase, M. W., et al. (2012). Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). *Am. J. Bot.* 99, 397–406. doi: 10.3732/ajb.1100537

McQueen, H. A., Siriaco, G., and Bird, A. P. (1998). Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Res.* 8, 621–630. doi: 10.1101/gr.8.6.621

Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J. G., et al. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* 14, R10. doi: 10.1186/gb-2013-14-1-r10

Moraes, A. P., Souza-Chies, T. T., Stiehl-Alves, E. M., Burchardt, P., Eggers, L., Siljak-Yakovlev, S., et al. (2015). Evolutionary trends in Iridaceae: new cytogenetic findings from the New World. *Bot. J. Linn. Soc.* 177, 27–49. doi: 10.1111/boj.12232

Ni, Z., Schwartz, B. E., Werner, J., Suarez, J. R., and Lis, J. T. (2004). Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol. Cell* 13, 55–65. doi: 10.1016/S1097-2765(03)00526-4

Noronha, R. C. R., Barros, L. M. R., Araújo, R. E. F., Marques, D. F., Nagamachi, C. Y., Martins, C., et al. (2016). New insights of karyoevolution in

the Amazonian turtles *Podocnemis expansa* and *Podocnemis unifilis* (Testudines, Podocnemidae). *Mol. Cytogenet.* 9, 73. doi: 10.1186/s13039-016-0281-5

Novak, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793. doi: 10.1093/bioinformatics/btt054

Padeken, J., Zeller, P., and Gasser, S. M. (2015). Repeat DNA in genome organization and stability. *Curr. Opin. Genet. Dev.* 31, 12–19. doi: 10.1016/j.gde.2015.03.009

Palomino, G., Martínez, J., Cepeda-Cornejo, V., and Pimienta-Barrios, E. (2012). Nuclear genome size and cytotype analysis in *Agave cupreata* Trel. & Berger (*Agavaceae*). *Caryologia* 65 (4), 281–294. doi:10.1080/00087114.2012.752915

Pedrosa, A., Jantsch, M. F., Moscone, E. A., Ambros, P. F., and Schweizer, D. (2001). Characterisation of pericentromeric and sticky intercalary heterochromatin in *Ornithogalum longibracteatum* (Hyacinthaceae). *Chromosoma* 110, 203–213. doi: 10.1007/s004120000125

Pedrosa, A., Sandal, N., Stougaard, J., Schweizer, D., and Bachmair, A. (2002). Chromosomal map of the model legume *Lotus japonicus*. *Genetics* 161, 1661–1672.

Renny-Byfield, S., Chester, M., Kovařík, A., Le Comber, S. C., Grandbastien, M. A., Deloger, M., et al. (2011). Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. *Mol. Biol. Evol.* 28, 2843–2854. doi: 10.1093/molbev/msr112

Rodionov, A. V., Lukina, N. A., Galkina, S. A., Solovei, I., and Saccone, S. (2002). Crossing over in chicken oogenesis: cytological and chiasma-based genetic maps of the chicken lampbrush chromosome 1. *J. Hered.* 93, 125–129. doi: 10.1093/jhered/93.2.125

Ruas, C. F., Weiss-Schneeweiss, H., Stuessy, T. F., Samuel, M. R., Pedrosa-Harand, A., Tremetsberger, K., et al. (2008). Characterization, genomic organization and chromosomal distribution of Ty1-copia retrotransposons in species of *Hypochaeris* (Asteraceae). *Gene* 412, 39–49. doi: 10.1016/j.gene.2008.01.009

Santos, F. C., Guyot, R., do Valle, C. B., Chiari, L., Techio, V. H., Heslop-Harrison, P., et al. (2015). Chromosomal distribution and evolution of abundant retrotransposons in plants: gypsy elements in diploid and polyploid *Brachiaria* forage grasses. *Chromosome Res.* 23, 571–582. doi: 10.1007/s10577-015-9492-6

Schmid, M., and Steinlein, C. (2017). The hypermethylated regions in avian chromosomes. *Cytogenet. Genome Res.* 151, 216–227. doi: 10.1159/000464268

Schubert, I., and Lysak, M. A. (2011). Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* 27, 207–216. doi: 10.1016/j.tig.2011.03.004

Shirakawa, J., Nagano, K., and Hoshi, Y. (2012). Polyploid genome structure of *Drosera spatulata* complex (Droseraceae). *Cytologia* 77, 97–106. doi: 10.1508/cytologia.77.97

Smith, J., Bruley, C. K., Paton, I. R., Dunn, I., Jones, C. T., Windsor, D., et al. (2000). Differences in gene density on chicken macrochromosomes and microchromosomes. *Anim. Genet.* 31, 96–103. doi: 10.1046/j.1365-2052.2000.00565.x

Sonnhammer, E. L. L., and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167, GC1–GC10. doi: 10.1016/0378-1119(95)00714-8

Sousa, A., Bellot, S., Fuchs, J., Houben, A., and Renner, S. S. (2016). Analysis of transposable elements and organellar DNA in male and female genomes of a species with a huge Y chromosome reveals distinct Y centromeres. *Plant J.* 88, 387–396. doi: 10.1111/tpj.13254

Steflova, P., Tokan, V., Vogel, I., Lexa, M., Macas, J., Novak, P., et al. (2013). Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY1Y2) in the dioecious plant *Rumex acetosa*. *Genome Biol. Evol.* 5, 769–782. doi: 10.1093/gbe/evt049

Stock, A. D., and Mengden, G. A. (1975). Chromosome banding pattern conservatism in birds and nonhomology of chromosome banding patterns between birds, turtles, snakes and amphibians. *Chromosoma* 50, 69–77. doi: 10.1007/BF00284963

Tenaillon, M. I., 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

Vosa, C. G. (2005). On chromosome uniformity, bimodality and evolution in the tribe Aloineae (Asphodelaceae). *Caryologia* 58, 83–85. doi: 10.1080/00087114.2005.10589437

Watkins, G. M. (1936). Chromosome numbers and species characters in *Yucca*. *Am. J. Bot.* 23, 328–333. doi: 10.1002/j.1537-2197.1936.tb08990.x

Weiss-Schneeweiss, H., Leitch, A., McCann, J., Jang, T. S., and Macas, J., (2015). Employing next generation sequencing to explore the repeat landscape of the plant genome, in *Next-generation sequencing in plant systematics*. Eds. Hörandl, E., Appelhans, and M. S. [International Association for Plant Taxonomy (IAPT)].

Weisshart, K., Fuchs, J., and Schubert, V. (2016). Structured illumination microscopy (SIM) and photoactivated localization microscopy (PALM) to analyze the abundance and distribution of RNA polymerase II molecules in flow sorted *Arabidopsis* nuclei. *Bio. Protoc.* 6, e1725. doi: 10.21769/BioProtoc.1725

Yin, G. S., Yang, Z., Chiang, T., and Gong, X. (2014). Tracing the origin of the bimodal karyotype of the tribe Lilieae (Liliaceae) based on comparative karyotypes analyses. *Plant Div. Res.* 36, 737–746.

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.

The handling editor and reviewer HW-S declared their involvement as co-editors in the Research Topic, and confirm the absence of any other collaboration.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 20 November 2019
doi: 10.3389/fpls.2019.01503

Chromosome Painting Facilitates Anchoring Reference Genome Sequence to Chromosomes *In Situ* and Integrated Karyotyping in Banana (*Musa* Spp.)

*Denisa Šimoníková[1], Alžběta Němečková[1], Miroslava Karafiátová[1], Brigitte Uwimana[2], Rony Swennen[3,4,5], Jaroslav Doležel[1] and Eva Hřibová[1]**

[1] Institute of Experimental Botany, Czech Academy of Sciences, Centre of the Region Hana for Biotechnological and Agricultural Research, Olomouc, Czechia, [2] Banana Breeding, International Institute of Tropical Agriculture, Kampala, Uganda, [3] Bioversity International, Banana Genetic Resources, Heverlee, Belgium, [4] Division of Crop Biotechnics, Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Leuven, Belgium, [5] Banana Breeding, International Institute of Tropical Agriculture, Arusha, Tanzania

OPEN ACCESS

Edited by:
Martin A. Lysak, Masaryk University, Czechia

Reviewed by:
Alexander Betekhtin, University of Silesia of Katowice, Poland
Jiming Jiang, University of Wisconsin-Madison, United States

***Correspondence:**
Eva Hřibová
hribova@ueb.cas.cz

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

Received: *29 July 2019*
Accepted: *29 October 2019*
Published: *20 November 2019*

Citation:
Šimoníková D, Němečková A, Karafiátová M, Uwimana B, Swennen R, Doležel J and Hřibová E (2019) Chromosome Painting Facilitates Anchoring Reference Genome Sequence to Chromosomes In Situ and Integrated Karyotyping in Banana (Musa Spp.). Front. Plant Sci. 10:1503. doi: 10.3389/fpls.2019.01503

Oligo painting FISH was established to identify all chromosomes in banana (*Musa* spp.) and to anchor pseudomolecules of reference genome sequence of *Musa acuminata* spp. *malaccensis* "DH Pahang" to individual chromosomes *in situ*. A total of 19 chromosome/chromosome-arm specific oligo painting probes were developed and were shown to be suitable for molecular cytogenetic studies in genus *Musa*. For the first time, molecular karyotypes of diploid *M. acuminata* spp. *malaccensis* (A genome), *M. balbisiana* (B genome), and *M. schizocarpa* (S genome) from the Eumusa section of *Musa*, which contributed to the evolution of edible banana cultivars, were established. This was achieved after a combined use of oligo painting probes and a set of previously developed banana cytogenetic markers. The density of oligo painting probes was sufficient to study chromosomal rearrangements on mitotic as well as on meiotic pachytene chromosomes. This advance will enable comparative FISH mapping and identification of chromosomal translocations which accompanied genome evolution and speciation in the family *Musaceae*.

Keywords: banana, chromosome identification, fluorescence *in situ* hybridization, molecular karyotype, *Musa*, oligo painting FISH

INTRODUCTION

Bananas (*Musa* spp.) are grown in tropical and subtropical regions of South East Asia, Africa and South America (Häkkinen, 2013; Janssens et al., 2016). They are one of the world's major fruit crops, a staple and important export commodity for millions of people living mainly in developing countries. Despite the importance and breeding efforts (Ortiz and Swennen, 2014; Brown et al., 2017), little is known about banana genome structure, organization and evolution at chromosomal level across the whole *Musaceae* family.

The genus *Musa* comprises about 75 species and numerous cultivated edible clones. Based on a set of morphological descriptors (IPGRI-INIBAP/CIRAD, 1996) and basic chromosome number (x),

the genus *Musa* has traditionally been divided into four sections: Eumusa (x = 11), Rhodochlamys (x = 11), Australimusa (x = 10), and Callimusa (x = 9, 10) (Cheesman, 1947). Argent (1976) created a separate section Ingentimusa which contains a single species *Musa ingens* with the lowest basic chromosome number (x = 7). However, genotyping using molecular markers revealed close relationship of *M. ingens* with other species of sections Callimusa and Australimusa (Li et al., 2010). Most of the modern edible banana clones originated within section Eumusa after intra- and inter-specific crosses between two wild diploid species *M. acuminata* (donor of A genome) and *M. balbisiana* (donor of B genome). In some cases, diploid *M. schizocarpa* (S genome) contributed to the evolution of edible clones, mainly after cross-breeding with diploid *M. acuminata* (Carreel et al., 1994; Čížková et al., 2013; Němečková et al., 2018). The spontaneous intra- and inter-specific crosses gave arise to seed sterile diploid (AA, AB) and triploid (AAA, AAB, or ABB) edible banana cultivars. Although tetraploid clones (AAAB, AABB) that originated spontaneously are known (Simmonds and Shepherd, 1955; Simmonds, 1956), currently cultivated tetraploid bananas were obtained in the breeding programs.

Species of genus *Musa* have a relatively small genome, ranging from 550 to 750 Mbp/1C (Doležel et al., 1994; Lysák et al., 1999; Asif et al., 2001; Kamaté et al., 2001; Bartoš et al., 2005; Čížková et al., 2015) and until now it was possible to identify only a few chromosomes in their karyotypes. The attempts were hampered by the relatively high number of chromosomes, their small size at mitotic metaphase (1–2 μm) and morphological similarity (Doleželová et al., 1998; Osuji et al., 1998; D'Hont et al., 2000). Chromosome banding, which was found informative in plant species with large and repeat-rich genomes, including wheat and rye (Gill and Kimber, 1977; Gill et al., 1991), did not result in diagnostic chromosome banding patterns in *Musa*, similar to many other plant species (Greilhuber, 1977; Schubert et al., 2001).

The application of fluorescence *in situ* hybridization (FISH), usually done with probes for DNA repeats with chromosome-specific distribution, provided a powerful approach to identify chromosomes in a range of plant species and study chromosome structural changes (e.g., Liu et al., 2011; Danilova et al., 2014; Amosova et al., 2017; Hou et al., 2018). Unfortunately, its use in *Musa* was hampered by the lack of suitable probes (Doleželová et al., 1998; Osuji et al., 1998; Valárik et al., 2002; Hřibová et al., 2007; Čížková et al., 2013). Until now, only NOR-bearing satellite chromosome, two chromosomes with clusters of tandem repeats CL18 and CL33, and two chromosomes bearing 5S rDNA loci can be cytogenetically identified in *M. acuminata* and *M. balbisiana* (Čížková et al., 2013). In *M. schizocarpa*, one chromosome pair bearing NOR and two chromosome pairs bearing tandem repeats CL18 and CL33 and other four chromosome pairs with 5S rDNA loci can be identified cytogenetically (Čížková et al., 2013). Even the mining of the reference genome sequence of *M. acuminata* "DH Pahang" (D'Hont et al., 2012) did not result in identification of sequences suitable as FISH probes useful for unambiguous identification of all *Musa* chromosomes and their anchoring to the genome sequence.

A method for chromosome painting, which allows fluorescent labeling of whole chromosomes, was developed in the late 1980s. This advance revolutionized human cytogenetics and found numerous applications in animal cytogenetics (e.g., Speicher et al., 1996; Cremer and Cremer, 2001; Ferguson-Smith and Trifonov, 2007). The original method was based on FISH with whole chromosome probes obtained from chromosomes isolated by flow cytometric sorting or microdissection. This was the reason why the method failed in plants where a majority of DNA repeats is distributed across the whole genome and only a minority of sequences are unique and chromosome-specific (Schubert et al., 2001). A solution was to use pools of chromosome-specific BAC (Bacterial Artificial Chromosome) clones (Lysák et al., 2001). However, the development of chromosome BAC pools requires whole genome sequence obtained after clone by clone (BAC by BAC) sequencing to identify single or low copy BAC clones useful for painting. Thus, the method is suitable for species with small genomes and containing low amounts of DNA repeats. Till now, painting using chromosome-specific BAC pools was used in dicotyledonous species with small nuclear genomes— *Arabidopsis* and its closely related species (e.g., Lysák et al., 2001; Mandáková and Lysák, 2008; Mandáková et al., 2013) as well as in monocot *Brachypodium distachyon* (Idziak et al., 2014). The attempts to use BAC FISH in banana were not successful due to the lack of a larger number of BAC clones containing single or low copy sequences (Hřibová et al., 2008).

The recent progress in the production of reference genome sequences and in technologies for DNA synthesis provided an alternative opportunity for affordable preparation of whole chromosome probes (chromosome paints) for FISH. The method called oligo painting FISH (Han et al., 2015) is based on *in silico* identification of large numbers of short (usually 45–50 bp) and unique (single copy) sequences in pseudomolecules of individual chromosomes, or their parts, synthesis of oligonucleotides, and their fluorescent labeling. A pool of synthesized and fluorescently labeled oligonucleotides is then used as a probe for FISH. Thus, the oligo painting FISH provides an opportunity to identify individual chromosomes and chromosome regions in *Musa*, perform comparative chromosome analysis and characterize chromosomal rearrangements (Qu et al., 2017; Braz et al., 2018; Xin et al., 2018; Jiang, 2019).

The present study fills the important gap in molecular cytogenetics of *Musa*. The application of oligo painting FISH described here allows anchoring genome sequence to chromosomes *in situ* and unambiguous identification of all *Musa* chromosomes after development of molecular karyotypes by a combined use of oligo painting probes and existing cytogenetic landmarks. Molecular karyotypes are described and compared for the three main genomes of Eumusa section—*M. acuminata* ssp. *malaccensis*, *M. balbisiana*, and *M. schizocarpa*, which contributed to the evolution of many edible banana clones.

MATERIALS AND METHODS

Plant Material and Preparations of Chromosome Spreads

Representatives of three species from the section Eumusa were obtained as *in vitro* rooted plants from the International *Musa* Transit Centre (ITC, Bioversity International, Leuven, Belgium). *In vitro* plants were transferred to garden soil and maintained in a heated greenhouse. **Table 1** lists the accessions used in this study.

TABLE 1 | List of analyzed accessions, their genomic constitution, genome size, and the number of loci identified on mitotic metaphase chromosomes (data from Čížková et al., 2013).

Species	Accession name	ITC code[a]	Genomic constitution	Genome size (1C)	Chromosome number (2n)	The number of loci in diploid cells (2n = 22)				
						45S rDNA	5S rDNA	BAC 2G17	CL33	CL18
M. acuminata ssp. *malaccensis*	Pahang	0609	AA	594 Mbp[b]	22	2	6	2	4	2
M. balbisiana	Tani	1120	BB	551 Mbp[b]	22	2	6	2	0	4
M. schizocarpa	Schizocarpa	0560	SS	671 Mbp[b]	22	2	12	2	4	2

[a]*Code assigned by the International Transit Centre (ITC, Leuven, Belgium)*

[b]*DNA content was estimated by flow cytometry using Glycine max L. cv. Polanka (2C = 2.5pg DNA) which served as an internal reference standard (Čížková et al., 2013).*

Male buds of *M. acuminata* "Pahang" and *M. balbisiana* "Tani" were obtained from the research station of the International Institute of Tropical Agriculture in Sendusu, Uganda.

Actively growing root tips (~1 cm long) were collected into 50-mM phosphate buffer (pH 7.0) containing 0.2% (v/v) β-mercaptoethanol, pre-treated in 0.05% (w/v) 8-hydroxyquinoline for three hours at room temperature, fixed in 3:1 ethanol:acetic acid fixative overnight, and stored in 70% ethanol. Preparation of protoplast suspensions was performed according to Doležel et al. (1998). Briefly, after digesting root tip segments in a mixture of 2% (w/v) cellulase and 2% (w/v) pectinase in 75-mM KCl and 7.5-mM EDTA (pH 4) for 90 min at 30°C, the suspension of resulting protoplasts was filtered through a 150-µm nylon mesh, pelleted, and washed in 70% ethanol. For further use, the protoplast suspension was stored in 70% ethanol at −20°C. Mitotic metaphase chromosome spreads were prepared by dropping method according to Doležel et al. (1998), the slides were postfixed in 4% (v/v) formaldehyde solution in 2x SSC solution and used for FISH.

Preparation of pachytene chromosome spreads was performed according to Mandáková and Lysák (2008), with minor modifications. Male flowers were fixed in 3:1 ethanol:acetic acid fixative overnight and stored in 70% ethanol at −20°C. Anthers were incubated in 0.3% (w/v) mix of cellulase, cytohelicase, and pectolyase (Sigma Aldrich, Darmstadt, Germany) for 30 min at 37°C. After the incubation in the enzyme mixture, the anthers were dissected in a drop of 60% (v/v) acetic acid on a microscopic slide and spread on the slide placed on a metal hot plate (50°C) after adding 60% (v/v) acetic acid for 25 s. The preparations were fixed in 3:1 ethanol:acetic acid fixative, air-dried, and used for FISH.

Identification of Specific Oligomers and Labeling of the Oligo Probes

Oligomers specific for individual chromosome arms were identified in the reference genome sequence of *M. acuminata* "DH Pahang" v.2 (Martin et al., 2016) using Chorus pipeline (Han et al., 2015). Sets of 20,000 oligomers (45-mers) per one library were synthesized by Arbor Biosciences (Ann Arbor, Michigan, USA). Labeled oligomer probes were prepared according to Han et al. (2015). Briefly, the oligomer libraries were amplified using emulsion PCR (Murgha et al., 2014), where F primer contained T7 RNA polymerase promoter. The emulsified PCR product was washed with water-saturated diethyl ether and ethyl acetate and purified with QIAquick PCR purification kit (Qiagen, Hilden, Germany). The product (480 ng DNA) was used for T7 *in vitro* transcription with MEGAshortscript T7 Kit (ThermoFisher Scientific/Invitrogen, Waltham, Massachusetts, USA) at 37°C for 4 h. The RNA product was purified using RNeasy Mini Kit (Qiagen) and 42 µg of RNA was reverse-transcribed with either digoxigenin-, biotin-, or CY5-labeled R primer (Eurofins Genomics, Ebersberg, Germany) using Superscript II Reverse Transcriptase and SUPERase-In RNase inhibitor (ThermoFisher Scientific/Invitrogen). The RNA : DNA hybrids were cleaned with Quick-RNA MiniPrep Kit (Zymo Research, Freiburg im Breisgau, Germany) and hydrolyzed with RNase H (New England Biolabs, Ipswich, Massachusetts, USA) and finally with RNase A (ThermoFisher Scientific/Invitrogen). The products were purified with Quick-RNA MiniPrep Kit (Zymo Research) and eluted with nuclease-free water to obtain single-stranded labeled oligomers, which were used as FISH probes.

Preparation of Other Cytogenetic Markers for FISH

Probes specific for ribosomal DNA sequences were prepared by labeling *Radka1* (part of 26S rRNA gene) and *Radka2* (contains 5S rRNA gene and non-transcribed spacer) DNA clones (Valárik et al., 2002) with biotin-16-dUTP (Roche Applied Science, Penzberg, Germany) or aminoallyl-dUTP-CY5 (Jena Biosciences, Jena, Germany) by PCR using T3 (forward) and T7 (reverse) primers (Invitrogen). Probes for tandem repeats CL18 and CL33 (Hřibová et al., 2010) were amplified using specific primers and labeled with aminoallyl-dUTP-CY5 or fluorescein-12-dUTP (Jena Biosciences, Jena, Germany) by PCR according to Čížková et al. (2013). Single copy BAC clone 2G17 (Hřibová et al., 2008) was labeled by digoxigenin-11-dUTP nick translation following manufacturer's recommendation (Roche Applied Science, Penzberg, Germany).

Fluorescence *In Situ* Hybridization and Image Analysis

Hybridization mix (30 µl) containing 50% (v/v) formamide, 10% (w/v) dextran sulfate in 2x SSC and 10 ng/µl of labeled probe was added onto slide and denatured for 3 min at 80°C. Hybridization was carried out overnight at 37°C. The sites of

hybridization of digoxigenin- and biotin-labeled probes were detected using anti-digoxigenin-FITC (Roche Applied Science) and streptavidin-Cy3 (ThermoFisher Scientific/Invitrogen), respectively. Chromosomes were counterstained with DAPI and mounted in VECTASHIELD Antifade Mounting Medium (Vector Laboratories, Burlingame, CA, USA). The slides were examined with Axio Imager Z.2 Zeiss microscope (Zeiss, Oberkochen, Germany) equipped with Cool Cube 1 camera (Metasystems, Altlussheim, Germany) and appropriate optical filters. The capture of fluorescence signals, merging the layers, and measurement of chromosome length were performed with ISIS software 5.4.7 (Metasystems), the final image adjustment and creation of idiograms were done in Adobe Photoshop CS5.

RESULTS

Development of Chromosome Painting Probes and *In Situ* Hybridization

In order to produce chromosome arm-specific painting probes, unique *k*- mers were identified according to Han et al. (2015) in the reference genome sequence of the doubled haploid banana (*M. acuminata* "DH Pahang"; Martin et al., 2016) and analyzed with the Chorus program (https://github.com/forrestzhang/Chorus). While eight pseudomolecules corresponded to metacentric chromosomes, pseudomolecules 1, 2, and 10 appeared to be acrocentric with peri-centromeric region occupying an entire chromosome arm. The density of unique oligomers was lower in peri-centromeric regions in all pseudomolecules (**Supplementary Figure S1**) and these regions were excluded from the selection of oligomers for painting probes. The number of unique oligomers ranged from 79,896 to 127,835 for pseudomolecules 2 and 6, respectively. Sets of 20,000 45-mers specific to individual chromosome arms were then selected in Chorus, synthesized as so called immortal libraries and labeled directly by Cy5 or indirectly by biotin or digoxigenin as described in Materials and Methods. Oligomer libraries were designed to achieve a density of 0.9 to 2.1 oligomers per 1-kb chromosome sequence (**Supplementary Table S1**). To confirm that it is not possible to paint peri-centromeric regions with low oligomer densities and large gaps between low copy oligomers, a painting probe was prepared from peri-centromeric region of pseudomolecule 3. In total, 8,317 oligomers spanning this region (~10.5 Mb long) ensured an average density of ~0.8 oligomers/kb.

First, the painting probes were hybridized to mitotic metaphase chromosomes spreads of *M. acuminata* ssp. *malaccensis* (A genome)—the genotype from which the *Musa* reference genome sequence was developed. FISH with the painting probes resulted in visible signals covering chromosome arms along their lengths (**Figures 1A–F**). This observation confirmed that the probes had the expected parameters. Moreover, because the painting highlighted individual chromosome arms, it was possible to anchor pseudomolecules to individual chromosome arms. This work revealed that in the assembly, pseudomolecules 1, 6, 7 start with long arms and end with short arms, i.e., they are oriented inversely to the way karyotypes are presented, where the short arm of the chromosome is on top and the long arm on the bottom.

Following this, the painting probes were used for FISH in *M. balbisiana* (B genome) and *M. schizocarpa* (S genome) (**Figures 1B**, **G**, **H**). Comparison of chromosome and/or chromosome-arm painting in *M. acuminata* ssp. *malaccensis* and *M. schizocarpa* did not reveal any large chromosome translocations differentiating both species. On the other hand, a large translocation of the long arm of chromosome 3 to long (painted) arm of chromosome 1 was found in *M. balbisiana* (**Figure 1B**).

The small size of condensed mitotic metaphase chromosomes reduces the longitudinal resolution of chromosome painting and hence a chance to discover small structural rearrangements. An alternative is to perform chromosome painting with meiotic pachytene chromosomes (**Figure 2**) which are approximately fifty times longer. When hybridized to pachytene chromosome spreads of *M. acuminata* ssp. *malaccensis*, painting probes provided visible signals and the opportunity to analyze chromosome structure in more detail. This experiment showed that banana chromosomes do not contain large blocks of heterochromatin in distal and subtelomeric regions (**Figure 2**). Taking the advantage of higher spatial resolution, the set of oligo painting probes developed in this work will be suitable to visualize meiotic processes such as crossing over and synapsis. Following this, pachytene chromosome spreads of *M. acuminata* ssp. *malaccensis* were used to evaluate the signal of peri-centromeric painting probe designed for chromosome 3. FISH with the probe did not result in a continuous signal along the whole region. Instead, discontinuous signals, with signal-free gaps along most of the (peri-)centromeric region of chromosome 3 (**Figure 2B**), were observed. Based on this observation, painting probes were not designed for (peri-)centromeric regions of the remaining ten banana chromosomes.

Integration of Cytogenetic Landmarks and Oligopaints

In order to utilize the existing probes for FISH in *Musa* and develop a highly informative toolbox to characterize *Musa* chromosome structure, the existing cytogenetic landmarks were integrated with the painting probes.

45S rRNA genes mapped to secondary constriction located on non-painted arm of chromosome 10 in all three *Musa* species. The probe for 5S rRNA genes localized to different chromosome regions and on different chromosomes in the three *Musa* species studied. In *M. acuminata* ssp. *malaccensis*, six signals of 5S rDNA were observed on mitotic metaphase plates and were localized in sub-telomeric region of chromosome 1 and long arm of chromosome 8, and in peri-centromeric region on short arm of chromosome 3. Six hybridization signals with 5S rDNA probe were observed also in mitotic metaphase plate of *M. balbisiana*. Two pairs of strong signals were localized in sub-telomeric region of chromosome 2 (non-painted arm) and in peri-centromeric region of the long arm of chromosome 3. Additional weak signal was observed in peri-centromeric region of the long arm of chromosome 6 (**Figures 3** and **4**). In *M. schizocarpa*, three pairs of strong signals and three pairs of weaker signals were observed after FISH with 5S rDNA probe on mitotic metaphase plate. Sub-telomeric region of chromosome 1 (non-painted arm) and peri-centromeric region of short arm of chromosome 3 and long arm of chromosome 4

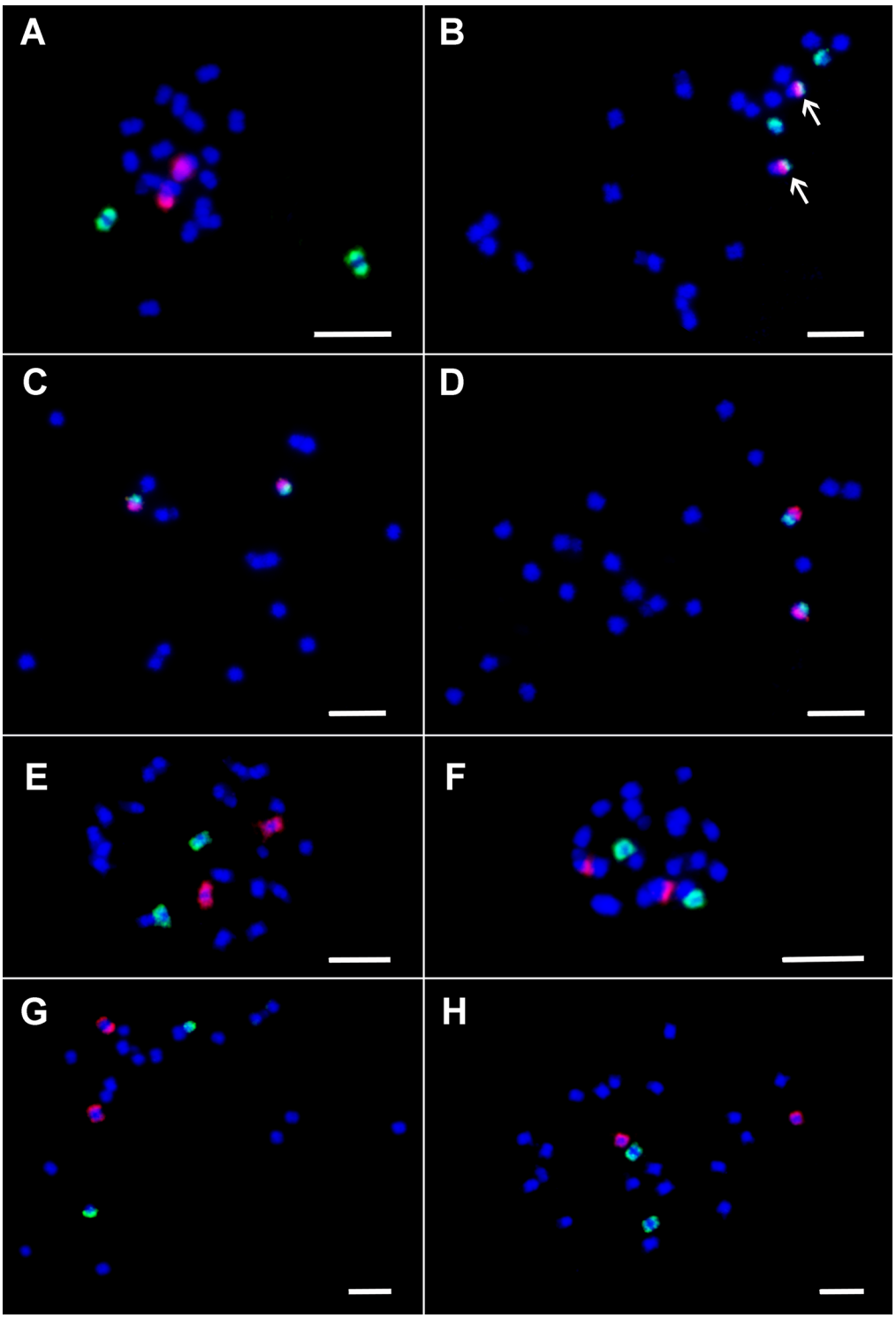

FIGURE 1 | Oligo painting FISH on mitotic metaphase plates of three species of *Musa*. **(A)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; chromosome 1 in red, chromosome 3 in green). **(B)** *M. balbisiana* "Tani" (2n = 22, BB; chromosome 1 in red, chromosome 3 in green). **(C)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; short arm of chromosome 4 in green, its long arm in red). **(D)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; short arm of chromosome 5 in red, its long arm in green). **(E)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; chromosome 6 in red, chromosome 7 in green. **(F)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; chromosome 10 in red, chromosome 11 in green. **(G)** *Musa schizocarpa* "Schizocarpa" (2n = 2x = 22, SS; chromosome 8 in red, chromosome 2 in green). **(H)** *Musa schizocarpa* "Schizocarpa" (2n = 2x = 22, SS; chromosome 11 in red, chromosome 9 in green). Chromosomes were counterstained with DAPI (blue). Arrows point to the region of chromosome 3 translocated to chromosome 1 in *M. balbisiana*. Bars = 5 μm.

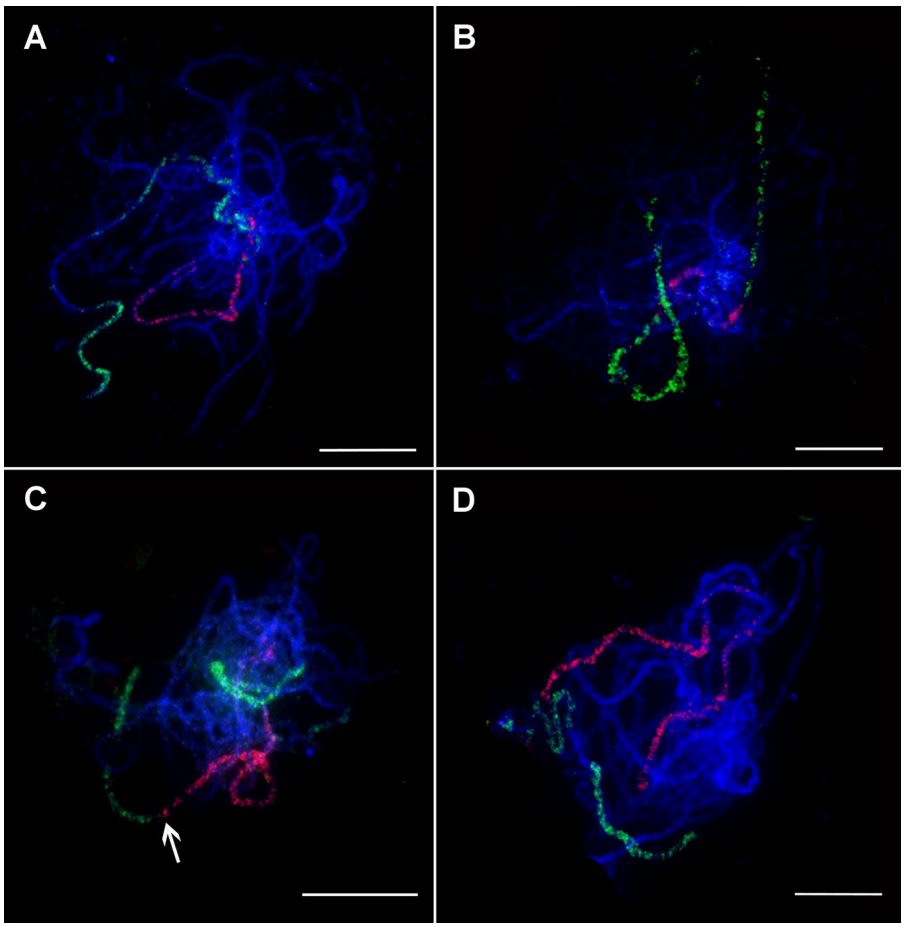

FIGURE 2 | Oligo painting FISH on meiotic pachytene chromosome spreads of *Musa*. **(A)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; chromosome 1 in red, chromosome 4 in green). **(B)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; (peri-)centromeric region in red, chromosome 3 in green). **(C)** *M. balbisiana* "Tani" (2n = 22, BB; chromosome 1 in red, chromosome 3 in green). **(D)** *M. balbisiana* "Tani" (2n = 22, BB; chromosome 5 in red, chromosome 11 in green). Chromosomes were counterstained with DAPI (blue). Arrows point to the region translocated from chromosome 3 to chromosome 1 in *M. balbisiana*. Bars = 10 µm.

contained strong signals of 5S rDNA. Additional weaker signals of 5S rDNA probe were observed in peri-centromeric regions of short arm of chromosome 8 and short arm of chromosome 11, as well as on the non-painted arm of chromosome 10 (**Figures 3** and **4**).

Tandem organized repeats CL18, CL33, and BAC clone 2G17 were localized on non-painted arm of chromosome 1 in all three *Musa* species, except satellite CL33, which was not detected on any chromosome in *M. balbisiana*. In contrast, additional signal of tandem repeat CL33 located on non-painted arm of chromosome 2 was observed in *M. acuminata* ssp. *malaccensis* and in *M. schizocarpa* (**Figures 3** and **4**). Finally, additional signal of tandem repeat CL18 was observed in *M. balbisiana* on the non-painted arm of chromosome 2.

DISCUSSION

Until recently, chromosome painting could be used only in plants whose genomes were sequenced clone by clone (The Arabidopsis Genome Initiative, 2000; The International Brachypodium Initiative, 2010) and chromosome painting was achieved by FISH with pools of single copy BAC clones that covered entire chromosomes. Importantly, chromosome paints developed in one species could be used in related species, providing a powerful approach for comparative karyotype analysis and for tracing karyotype changes during the evolution and speciation (e.g., Lysák et al., 2001; Idziak et al., 2014). Unfortunately, this painting method cannot be used in species with large genomes due to the prevalence of repetitive DNA and in species not closely related to those for which painting using BAC pools was developed.

The progress in DNA sequencing technology and assembly algorithms resulted in a shift from the clone by clone sequencing to shotgun sequencing and a majority of plant genomes has been sequenced in this way (Hamilton and Buell, 2012; Zimin et al., 2017; Belser et al., 2018). The availability of reference genome sequences and the affordable cost of synthesizing short oligonucleotides offered a direct way to develop chromosome paints (Han et al., 2015). Here, thousands of short single copy sequences are identified, bulk synthesized, fluorescently labeled and used as probes for FISH (Han et al., 2015). Pools of labeled

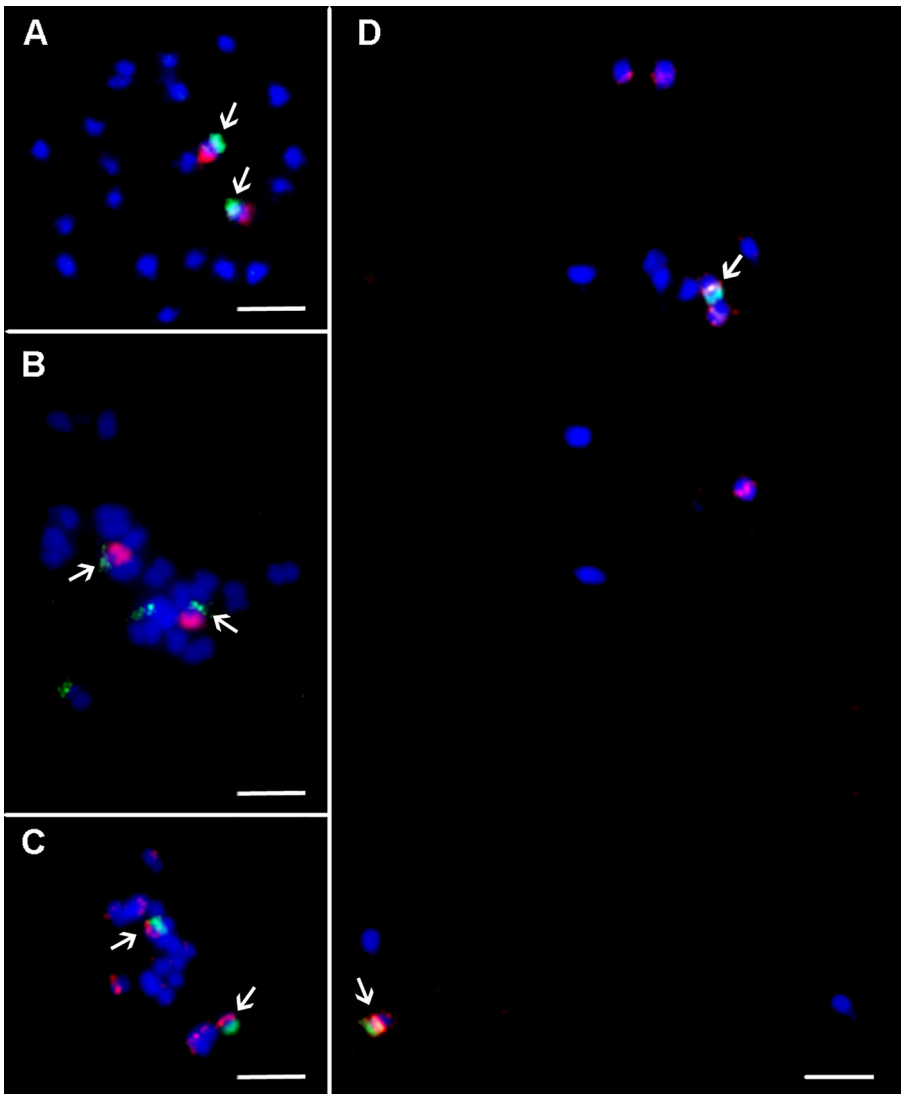

FIGURE 3 | Integration of oligo painting FISH and existing cytogenetic markers on mitotic metaphase plates of *Musa*. **(A)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; chromosome 1 in red, BAC clone 2G17 in green). **(B)** *M. schizocarpa* "Schizocarpa" (2n = 22, SS; chromosome 2 in red, tandem repeat CL33 in green). **(C)** *M. schizocarpa* "Schizocarpa" (2n = 22, SS; short arm of chromosome 4 in green, 5S rRNA in red—two loci are localized on long arm of chromosome 4). **(D)** *M. acuminata* ssp. *malaccensis* "Pahang" (2n = 22, AA; 5S rRNA in red, short arm of chromosome 3 in green bears 5S rRNA). Chromosomes were counterstained with DAPI (blue). Bars = 5 µm. Arrows indicate colocalization of oligo painting FISH probes with existing cytogenetic markers.

oligonucleotide probes were found suitable for FISH on somatic metaphase chromosomes, meiotic pachytene chromosomes and interphase nuclei (Han et al., 2015; Filiault et al., 2018; Xin et al., 2018; Albert et al., 2019; Jiang, 2019). Successful applications of oligo painting FISH include construction of molecular cytogenetic karyotypes (Braz et al., 2018; Qu et al., 2017; Meng et al., 2018), identification of large chromosomal rearrangements, analysis of chromosome pairing in meiosis (Han et al., 2015; He et al., 2018; Albert et al., 2019), as well as the visualization of the arrangement of chromosomes in 3D space of interphase nuclei (Albert et al., 2019).

Despite the availability of a reference genome sequence of *M. acuminata ssp. malaccensis* (D'Hont et al., 2012; Martin et al., 2016) and resequencing of more than 120 accessions of *Musa* (Dupouy et al., 2019), DNA pseudomolecules have not been anchored to individual chromosomes and molecular karyotype of *Musa* has not been developed to date. In many plant species, tandem organized repeats serve as useful probes for FISH to identify individual chromosomes and their regions (Hřibová et al., 2007; Badaeva et al., 2015; Koo et al., 2016; Křivánková et al., 2017; Said et al., 2018). The nuclear genome of *Musa* species is relatively small (1C ~ 500–750 Mb; Doležel et al., 1994; Lysák et al., 1999; Asif et al., 2001; Kamaté et al., 2001; Bartoš et al., 2005; Čížková et al., 2015) and until now, only a few tandem organized repeats and rDNA sequences were successfully used as cytogenetic landmarks (Balint-Kurti et al., 2000; Valárik et al., 2002; Hřibová et al., 2007; Hřibová et al., 2010; Čížková et al., 2013; Novák et al., 2014). Moreover, only

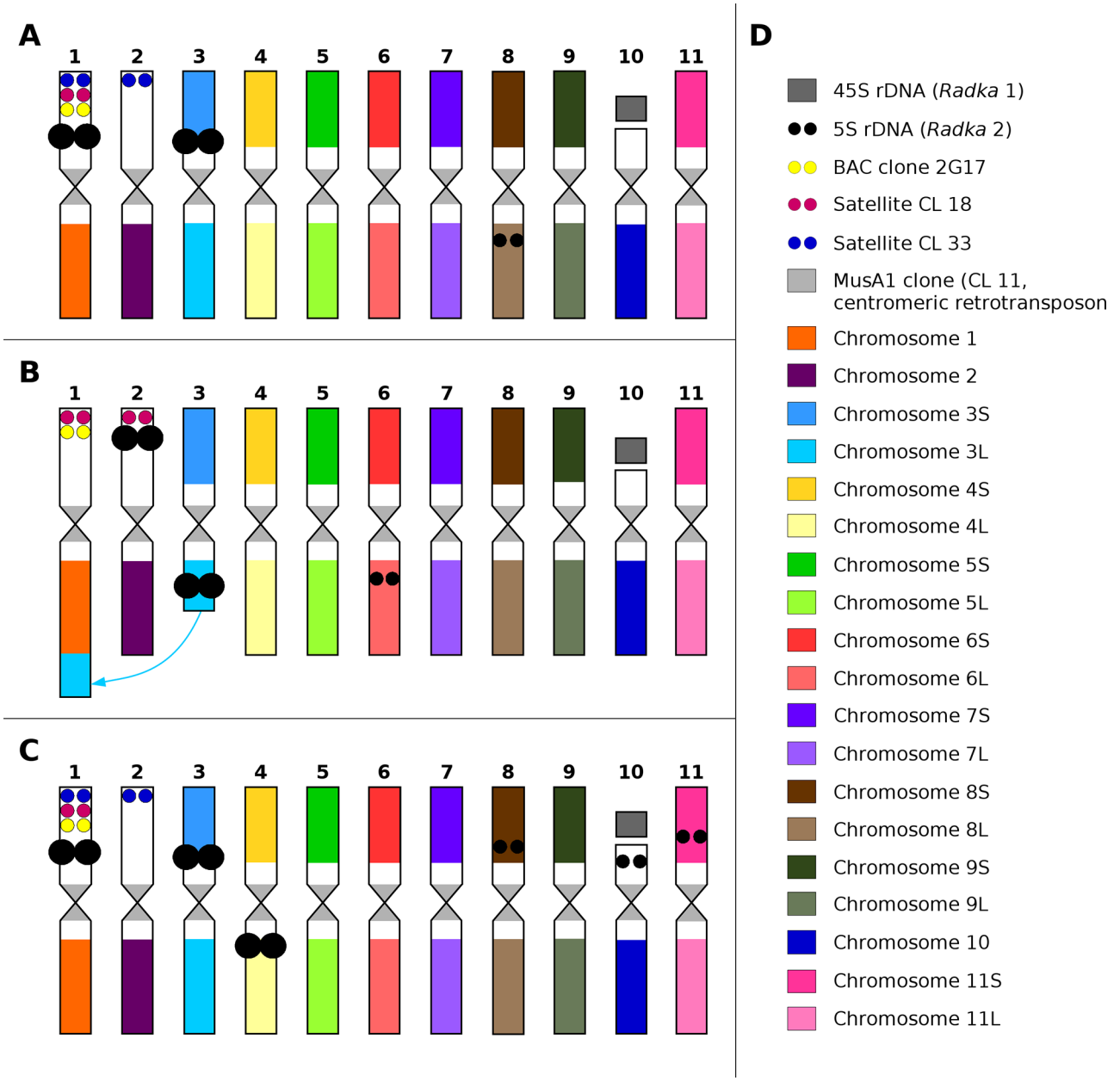

FIGURE 4 | Idiograms of three diploid species of Eumusa section of *Musa*. **(A)** *Musa acuminata* ssp. *malaccensis* "Pahang" (genome A) ITC 0609. **(B)** *Musa balbisiana* "Tani" (genome B) ITC 1120. **(C)** *Musa schizocarpa* "Schizocarpa" (genome S) ITC 0560. **(D)** multicolored scheme.

one BAC clone has been used as a cytogenetic marker in *Musa* (Hřibová et al., 2008) and only four BAC clones were localized on pachytene chromosomes (De Capdeville et al., 2009). Thus, the attempts to use of BAC clones for anchoring pseudomolecules to chromosomes in banana as has been done in other species (Jiang et al., 1995; Lapitan et al., 1997; Kim et al., 2002; Idziak et al., 2014), were not successful.

Unlike the previous approaches, chromosome painting using pools of single copy oligomers offers the opportunity to establish a molecular karyotype of *Musa*, making it possible to identify individual chromosomes, follow their behavior during somatic cell cycle and meiosis, perform comparative karyotype analysis, and identify structural chromosome changes. FISH with oligo painting probes developed in this work resulted in visible hybridization signals along chromosomal arms on condensed mitotic metaphase chromosomes (**Figure 1**) as well as on less condensed pachytene chromosomes (**Figure 2**) confirming their usefulness as painting probes in *Musa*.

Only small regions on pachytene chromosomes were free of painting signals. This could be either due to the presence of heterochromatin blocks, or due to gaps in the genome sequence (**Figure 2**). In contrast to chromosome arms, peri-centromeric regions were not labeled. These regions contain large gaps in the genome sequence and large proportion of repetitive DNA sequences in peri-centromeric regions (Hřibová et al., 2010; Neumann et al., 2011; D'Hont et al., 2012; Martin et al., 2016).

We demonstrate that chromosome/chromosome-arm specific oligo painting libraries designed for *M. acuminata* ssp. *malaccensis* can be used for cytogenetic analysis of related species *M. balbisiana* and *M. schizocarpa*, which played an important role in the evolution of many edible banana clones (Carreel et al., 1994; D'Hont et al., 2012; Davey et al., 2013; Čížková et al., 2013). This observation provided an opportunity for comparative karyotype analysis and identification of putative chromosome translocations. In our study, we observed translocation of long arm of chromosome 3 to long arm of chromosome 1 in *M. balbisiana* (B genome) (**Figures 1B** and **2C**). This observation confirms the result of Baurens et al. (2019), which were obtained after anchoring a dense genetic map of *M. balbisiana* "Pisang Klutuk Wulung" to *M. acuminata* ssp. *malaccensis* reference genome sequence (Martin et al., 2016). The authors estimated the size of the translocated region of long arm of chromosome 3 to be ~8 Mb, confirming the sensitivity of oligo chromosome painting.

Co-localization of chromosome painting probes with cytogenetic markers developed earlier for *Musa* (Valárik et al., 2002; Hřibová et al., 2010; Čížková et al., 2013) offered an opportunity to create molecular karyotypes suitable for comparative analysis. The presence of 5S rRNA genes on non-collinear chromosomes in the A, B, and S genomes of *Musa* as described here indicates small chromosomal rearrangements which occurred during *Musa* speciation. On the other hand, the location of tandem organized repeats CL18, CL33, and BAC clone 2G17 on collinear chromosome arms in all three species indicates their structural homology of the chromosome arms. These observations imply that chromosomes containing a particular DNA sequence, e.g., 5S rDNA, cannot be considered as collinear. This shows a potential weakness of comparative karyotype analysis of using only a few cytogenetic markers (Fukui et al., 1994; Murata et al., 1997).

Tandem organized repeats CL18 and CL33 (Hřibová et al., 2010) were located together with 5S rRNA genes on short arms of chromosomes 1 and 2, which lacked oligopainting signals. Genome sequence of *M. acuminata* ssp. *malaccensis* includes three pseudomolecules which are represented by two large regions differing in DNA repeat composition and in density of unique oligomers (**Supplementary Figure 1**, Martin et al., 2016). The constitution of banana pseudomolecules 1, 2, and 10 indicates that they cover only one chromosome arm and a peri-centromeric region. Painting probes created for the three pseudomolecules localized to only one chromosomal arm. One of the pseudomolecules is collinear with acrocentric chromosome 10 and bears 45S rRNA locus on its short arm. The two remaining pseudomolecules represent chromosomes 1 and 2, which seem to be meta or sub-metacentric thus could miss a large sequence region. These observations indicate that these genomic regions were not completely assembled and are missing due to the presence of a large number of various tandem organized sequences.

The improved version of *M. acuminata* "DH Pahang" reference genome sequence represents 450.7 Mbp which corresponds to ~81% of its nuclear genome size estimated by flow cytometry (Čížková et al., 2013). In addition, the reference genome sequence contains a total of 56.6-Mbp sequences, which were not anchored to the 11 pseudomolecules. The most plausible explanation why these sequences were not included in pseudomolecules is that they represent heterochromatin regions, which are difficult to sequence. However, relatively high number of unique oligomers in unanchored scaffolds as observed in this work (**Supplementary Figure 1**) indicates that the unanchored part of the reference genome sequence contains low copy sequences from euchromatic regions. Thus, these regions were probably not anchored due to the absence of DNA markers, or they were too short to be anchored using Bionano optical mapping. The use of long-read sequencing technologies such as Oxford Nanopore in combination with optical mapping (Belser et al., 2018) should further improve the current assembly and shed light on the difficult parts of *M. acuminata* ssp. *malaccensis* genome.

CONCLUSIONS

In this work, chromosome painting probes were developed for banana (*Musa* spp.) and used to establish molecular karyotypes for three species of *Musa* that were the parents of a majority of cultivated edible banana clones. This advance made it possible to anchor reference genome sequence of banana, *Musa acuminata* spp. *malaccensis* to individual chromosomes. The study also demonstrates the potential of oligo painting FISH for comparative karyotype analysis and identification of structural chromosome changes that accompanied the evolution and speciation in the genus *Musa*.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

EH and JD conceived the experiments. DŠ, AN, and MK conducted the study and processed the data. BU and RS provided the banana materials. DŠ and EH wrote the manuscript. EH, JD, and DŠ discussed the results and contributed to

manuscript writing. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Dr. Ines van den Houwe for providing the plant material and Ms. Radka Tušková for excellent technical assistance. This work was supported by the Czech Science Foundation (award No. 19-20303S). The computing was supported by the National Grid Infrastructure MetaCentrum (grant No. LM2010005 under the program Projects of Large Infrastructure for Research, Development, and Innovations). The authors thank all donors who supported this work through their contributions to the CGIAR Fund (http://www.cgiar.org/who-we-are/cgiar-fund/fund-donors-2/), and in particular to the CGIAR Research Program Roots, Tubers and Bananas (RTB-CRP).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01503/full#supplementary-material

SUPPLEMENTARY FIGURE S1 | Oligomer coverage of 11 pseudomolecules (labeled **A** – **K**) and concatenated unanchored scaffolds **(L)** in the reference genome of *M. acuminata* 'DH Pahang' (Martin et al., 2016). The oligomers (45 bp) were designed using the Chorus program (Han et al., 2015) and are depicted in black. Position and coverage of tandem repeats CL18 (pink) and CL33 (green) are also shown.

REFERENCES

Albert, P. S., Zhang, T., Semrau, K., Rouillard, J. M., Kao, Y. H., and Wang, C. R. (2019). Whole-chromosome paints in maize reveal rearrangements, nuclear domains, and chromosomal relationships. *Proc. Natl. Acad. Sci. U. S. A.* 116 (5), 1679–1685. doi: 10.1073/pnas.1813957116

Amosova, A. V., Bolsheva, N. L., Zoshchuk, S. A., Twardovska, M. O., Yurkevich, O. Y., Andreev, I. O., et al. (2017). (A) Comparative molecular cytogenetic characterization of seven *Deschampsia (Poaceae)* species. *PloS One* 12 (4), e0175760. doi: 10.1371/journal.pone.0175760

Argent, G. C. G. (1976). The wild bananas of Papua New Guinea. *Notes R. Bot. Gard. Edinburgh.* 35 (1), 77–114.

Asif, M. J., Mak, C., and Othman, R. Y. (2001). Characterization of indigenous *Musa* species based on flow cytometric analysis of ploidy and nuclear DNA content. *Caryologia.* 54 (2), 161–168. doi: 10.1080/00087114.2001.10589223

Badaeva, E. D., Amosova, A. V., Goncharov, N. P., Macas, J., Ruban, A. S., Grechishnikova, I. V., et al. (2015). A set of cytogenetic markers allows the precise identification of all A-genome chromosomes in diploid and polyploid wheat. *Cytogenet. Genome Res.* 146 (1), 71–79. doi: 10.1159/000433458

Balint-Kurti, P., Clendennen, S., Doleželová, M., Valárik, M., Doležel, J., Beetham, P. R., et al. (2000). Identification and chromosomal localization of the *monkey* retrotransposon in *Musa* sp. *Mol. Gen. Genet.* 263 (6), 908–915. doi: 10.1007/s004380000265

Bartoš, J., Alkhimova, O., Doleželová, M., De Langhe, E., and Doležel, J. (2005). Nuclear genome size and genomic distribution of ribosomal DNA in *Musa* and *Ensete* (*Musaceae*): taxonomic implications. *Cytogenet. Genome Res.* 109, 50–57. doi: 10.1159/000082381

Baurens, F. C., Martin, G., Hervouet, C., Salmon, F., Yohomé, D., and Ricci, S. (2019). Recombination and large structural variations shape interspecific edible bananas genomes. *Mol. Biol. Evol.* 36 (1), 97–111. doi: 10.1093/molbev/msy199

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., and Falentin, C. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants.* 4 (11), 879–887. doi: 10.1038/s41477-018-0289-4

Braz, G. T., He, L., Zhao, H., Zhang, T., Semrau, K., and Rouillard, J. M. (2018). Comparative oligo-FISH mapping: An efficient and powerful methodology to reveal karyotypic and chromosomal evolution. *Genetics.* 208 (2), 513–523. doi: 10.1534/genetics.117.300344

Brown, A., Tumuhimbise, R., Amah, D., Uwimana, B., Nyine, M., and Mduma, H. (2017). *The genetic improvement of bananas and plantains (Musa spp.) In Genetic Improvement of Tropical Crops* Vol. pp. Campos, HCaligari, PDS. Cham: Springer, 219–240.

Carreel, F., Fauré, S., González de León, D., Lagoda, P. J. L., Perrier, X., and Bakry, F. (1994). Evaluation of the genetic diversity in diploid bananas (*Musa* sp.). *Genet. Sel. Evol.* 26 (Suppl 1), 125–136. doi: 10.1051/gse:19940709

Cheesman, E. E. (1947). Classification of the bananas. The genus *Ensete* Horan and the genus *Musa* L. Kew. *Bull.* 2 (2), 97–117. doi: 10.2307/4109206

and the genus Musa L. Kew Bull. 2(2):97-117

Čížková, J., Hřibová, E., Humplíková, L., Christelová, P., Suchánková, P., Doležel J. (2013). Molecular analysis and genomic organization of major DNA satellites in banana (Musa spp.). PLoS ONE. 8:e54808. doi: 10.1371/journal.pone.0054808

Čížková, J., Hřibová, E., Christelová, P., Van den Houwe, I., Häkkinen, M., Roux, N., et al. (2015). Molecular and cytogenetic characterization of wild *Musa* species. *PLoS One* 10:e0134096. doi: 10.1371/journal.pone.0134096

Cremer, T., and Cremer, S. (2001). (B) Chromosome territories, nuclear architecture and gene regulation in mammalian cells.. *Nat. Rev. Genet.* 2 (4), 292–301. doi: 10.1038/35066075

D'Hont, A., Paget-Goy, A., Escoute, J., and Carreel, F. (2000). The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genomic DNA *in situ* hybridization. *Theor. Appl. Genet.* 100, 177–183. doi: 10.1007/s001220050024

D'Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217. doi: 10.1038/nature11241

Danilova, T. V., Friebe, B., and Gill, B. S. (2014). Development of wheat single gene FISH map for analyzing homoeologous relationship and chromosomal rearrangements within the triticeae. *Theor. Appl. Genet.* 127 (3), 715–730. doi: 10.1007/s00122-013-2253-z

Davey, M. W., Gudimella, R., Harikrishna, J. A., Sin, L. W., Khalid, N., and Keulemans, J. (2013). A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids. *BMC Genomics* 14, 683. doi: 10.1186/1471-2164-14-683

De Capdeville, G., Souza Junior, M. T., Szinay, D., Diniz, L. E. C., Wijnker, E., Swennen, R., et al. (2009). The potential of high-resolution BAC-FISH in banana breeding. *Euphytica* 166, 431–443. doi: 10.1007/s10681-008-9830-2

Doležel, J., Doleželová, M., and Novák, F. J. (1994). Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *Musa balbisiana*). *Biol. Plant* 36, 351–357. doi: 10.1007/BF02920930

Doležel, J., Doleželová, M., Roux, N., and Van den houwe, I. (1998). A novel method to prepare slides for high resolution chromosome studies in *Musa* spp. *Infomusa* 7, 3–4.

Doleželová, M., Valárik, M., Swennen, R., Horry, J. P., and Doležel, J. (1998). Physical mapping of the 18S-25S and 5S ribosomal RNA genes in diploid bananas. *Biol. Plant* 41, 497–505. doi: 10.1023/A:1001880030275

Dupouy, M., Baurens, F. C., Derouault, P., Hervouet, C., Cardi, C., Cruaud, C., et al. (2019). Two large reciprocal translocations characterized in the disease resistance-rich *burmannica* genetic group of *Musa acuminata*. *Ann. Bot.* XX, 1–11. doi: 10.1093/aob/mcz078

Ferguson-Smith, M. A., and Trifonov, V. (2007). Mammalian karyotype evolution. *Nat. Rev. Genet.* 8 (12), 950–962. doi: 10.1038/nrg2199

Filiault, D. L., Ballerini, E. S., Mandáková, T., Aköz, G., Derieg, N. J., Schmutz, J., et al. (2018). The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *eLife* 7, e36426. doi: 10.7554/eLife.36426

Fukui, K., Kamisugi, Y., and Sakai, F. (1994). Physical mapping of 5S rDNA loci by direct-cloned biotinylated probes in barley chromosomes. *Genome.* 37 (1), 105–111. doi: 10.1139/g94-013

Gill, B. S., and Kimber, G. (1977). Recognition of translocations and alien chromosome transfers in wheat by the Giemsa C-banding technique. *Crop Sci.* 17, 264–266. doi: 10.2135/cropsci1977.0011183X001700020008x

Gill, B. S., Friebe, B., and Endo, T. R. (1991). Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). *Genome* 34, 830–839. doi: 10.1139/g91-128

Greilhuber, J. (1977). Why plant chromosomes do not show G-bands. *Theor. Appl. Genet.* 50 (3), 121–124. doi: 10.1007/BF00276805

Häkkinen, M. (2013). Reappraisal of sectional taxonomy in *Musa* (*Musaceae*). *Taxon.* 62 (4), 809–813. doi: 10.12705/624.3

Hamilton, J. P., and Buell, C. R. (2012). Advances in plant genome sequencing. *Plant J.* 70 (1), 177–190. doi: 10.1111/j.1365-313X.2012.04894.x

Han, Y., Zhang, T., Thammapichai, P., Weng, Y., and Jiang, J. (2015). Chromosome-specific painting in *Cucumis* species using bulked oligonucleotides. *Genetics.* 200, 771–779. doi: 10.1534/genetics.115.177642

He, L., Braz, G. T., Torres, G. A., and Jiang, J. M. (2018). Chromosome painting in meiosis reveals pairing of specific chromosomes in polyploid *Solanum* species. *Chromosoma* 127, 505–513. doi: 10.1007/s00412-018-0682-9

Hou, L., Xu, M., Zhang, T., Xu, Z., Wang, W., Zhang, J., et al. (2018). *BMC Plant Biol.* 18 (1), 110. doi: 10.1186/s12870-018-1325-2

Hřibová, E., Doleželová, M., Town, C. D., Macas, J., and Doležel, J. (2007). Isolation and characterization of the highly repeated fraction of the banana genome. *Cytogenet. Genome Res.* 119 (3-4), 268–274. doi: 10.1159/000112073

Hřibová, E., Doleželová, M., and Doležel, J. (2008). Localization of BAC clones on mitotic chromosomes of *Musa acuminata* using fluorescence *in situ* hybridization. *Biol. Plant* 52, 445–452. doi: 10.1007/s10535-008-0089-1

Hřibová, E., Neumann, P., Matsumoto, T., Roux, N., Macas, J., and Doležel, J. (2010). Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biol.* 10, 204. doi: 10.1186/1471-2229-10-204

Idziak, D., Hazuka, I., Poliwczak, B., Wiszynska, A., Wolny, E., and Hasterok, R. (2014). Insight into the karyotype evolution of *Brachypodium* species using comparative chromosome barcoding. *PloS One* 9 (3), e93503. doi: 10.1371/journal.pone.0093503

International Plant Genetic Resources Institute-International Network for the Improvement of Banana and Plantain/Centre de Coopération internationale en recherche agronomique pour le développement [IPGRI-INIBAP/CIRAD] International Plant Genetic Resources Institute-International Network for the Improvement of Banana and Plantain/Centre de Coopération internationale en recherche agronomique pour le développement [IPGRI-INIBAP/CIRAD] (1996). Description for Banana (*Musa* spp.). Int. Network for the Improvement of Banana and Plantain, Montpellier, France; Centre de coopération int. en recherche agronomique pour le développement, Montpellier, France; International Plant Genetic Resources Institute Press, Rome.

Janssens, S. B., Vandelook, F., De Langhe, E., Verstraete, B., Smets, E., Van den Houwe, I., et al. (2016). Evolutionary dynamics and biogeography of *Musaceae* reveal a correlation between the diversification of the banana family and the geological and climatic history of Southeast Asia. *New Phytol.* 210 (4), 1453–1465. doi: 10.1111/nph.13856

Jiang, J., Gill, B. S., Wang, G. L., Ronald, P. C., and Ward, D. C. (1995). Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 92 (10), 4487–4491. doi: 10.1073/pnas.92.10.4487

Jiang, J. (2019). Fluorescence *in situ* hybridization in plants: recent developments and future applications. *Chromosom. Res.* 27 (3), 153–165. doi: 10.1007/s10577-019-09607-z

Křivánková, A., Kopecký, D., Stočes, Š., Doležel, J., and Hřibová, E. (2017). Repetitive DNA: a versatile tool for karyotyping in *Festuca pratensis* Huds. *Cytogenet. Genome Res.* 151 (2), 96–105. doi: 10.1159/000462915

Kamaté, K., Brown, S., Durand, P., Bureau, J. M., De Nay, D., and Trinh, T. H. (2001). Nuclear DNA content and base composition in 28 taxa of *Musa*. *Genome.* 44, 622–627. doi: 10.1139/g01-058

Kim, J. S., Childs, K. L., Islam-Faridi, M. N., Menz, M. A., Klein, R. R., and Klein, P. E. (2002). Integrated karyotyping of sorghum by *in situ* hybridization of landed BACs. *Genome.* 45, 402–412. doi: 10.1139/g01-141

Koo, D. H., Zhao, H., and Jiang, J. (2016). Chromatin-associated transcripts of tandemly repetitive DNA sequences revealed by RNA-FISH. *Chromosome Res.* 24 (4), 467–480. doi: 10.1007/s10577-016-9537-5

Lapitan, N. L. V., Brown, S. E., Kennard, W., Stephens, J. L., and Knudson, D. L. (1997). FISH physical mapping with barley BAC clones. *Plant J.* 11, 149–156. doi: 10.1046/j.1365313X.1997.11010149.x

Li, L. F., Häkkinen, M., Yuan, Y. M., Hao, G., and Ge, X. J. (2010). Molecular phylogeny and systematics of the banana family (*Musaceae*) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. *Mol. Phylogenet. Evol.* 57 (1), 1–10. doi: 10.1016/j.ympev.2010.06.021

Liu, W., Rouse, M., Friebe, B., Jin, Y., Gill, B., and Pumphrey, M. O. (2011). Discovery and molecular mapping of a new gene conferring resistance to stem rust, Sr53, derived from *Aegilops geniculata* and characterization of spontaneous translocation stocks with reduced alien chromatin. *Chromosome Res.* 19 (5), 669–682. doi: 10.1007/s10577-011-9226-3

Lysák, M. A., Doleželová, M., Horry, J. P., Swennen, R., and Doležel, J. (1999). Flow cytometric analysis of nuclear DNA content in *Musa*. *Theor. Appl. Genet.* 98 (8), 1344–1350. doi: 10.1007/s001220051201

Lysák, M. A., Fransz, P. F., Ali, H. B. M., and Schubert, I. (2001). Chromosome painting in *A. thaliana*. *Plant J.* 28, 689–697. doi: 10.1046/j.1365-313x.2001.01194.x

Mandáková, T., and Lysák, M. A. (2008). Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (*Brassicaceae*). *Plant Cell* 20, 2559–2570. doi: 10.1105/tpc.108.062166

Mandáková, T., Marhold, K., and Lysák, M. A. (2013). The widespread crucifer species *Cardamine flexulosa* is an allotetraploid with a conserved subgenomic structure. *New Phytol.* 201, 982–992. doi: 10.1111/nph.12567

Martin, G., Baurens, F. C., Droc, G., Rouard, M., Cenci, A., Kilian, A., et al. (2016). Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics* 17, 1–12. doi: 10.1186/s12864-016-2579-4

Meng, Z., Zhang, Z. L., Yan, T. Y., Lin, Q. F., Wang, Y., Huang, W. Y., et al. (2018). Comprehensively characterizing the cytological features of *Saccharum spontaneum* by the development of a complete set of chromosome-specific oligo probes. *Front. Plant Sci.* 9, 1624. doi: 10.3389/fpls.2018.01624

Murata, M., Heslop-Harrison, J. S., and Motoyoshi, F. (1997). Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multi-color fluorescence *in situ* hybridization with cosmid clones. *Plant J.* 12 (1), 31–37. doi: 10.1046/j.1365-313X.1997.12010031.x

Murgha, Y. E., Rouillard, J. M., and Gulari, E. (2014). Methods for the preparation of large quantities of complex single-stranded oligonucleotide libraries. *PloS One* 9, e94752. doi: 10.1371/journal.pone.0094752

Němečková, A., Christelová, P., Čížková, J., Nyine, M., Van den houwe, I., Svačina, R., et al. (2018). Molecular and cytogenetic study of East African Highland Banana. *Front. Plant Sci.* 9, 1371. doi: 10.3389/fpls.2018.01371

Neumann, P., Navrátilová, A., Koblížková, A., Kejnovský, E., Hřibová, E., Hobza, R., et al. (2011). Plant cytogenetic perspective. *Mob. DNA* 2 (1), 4. doi: 10.1186/1759-8753-2-4

Novák, P., Hřibová, E., Neumann, P., Koblížková, A., Doležel, J., and Macas, J. (2014). Genome-wide analysis of repeat diversity across the family *Musaceae*. *PloS One* 9 (6), e98918. doi: 10.1371/journal.pone.0098918

Ortiz, R., and Swennen, R. (2014). From crossbreeding to biotechnology-facilitated improvement of banana and plantain. *Biotechnol. Adv.* 32, 158–169. doi: 10.1016/j.biotechadv.2013.09.010

Osuji, J. O., Crouch, J., Harrison, G., and Heslop-Harrison, J. S. (1998). Molecular cytogenetics of *Musa* species, cultivars and hybrids: location of 18S-5.8S-25S and 5S rDNA and telomere-like sequences. *Ann. Bot.* 82, 243–248. doi: 10.1006/anbo.1998.0674

Qu, M., Li, K., Han, Y., Chen, L., Li, Z., and Han, Y. (2017). Integrated karyotyping of woodland strawberry (*Fragaria vesca*) with oligopaint FISH probes. *Cytogenet. Genome Res.* 153, 158–164. doi: 10.1159/000485283

Said, M., Hřibová, E., Danilova, T. V., Karafiátová, M., Čížková, J., Friebe, B., et al. (2018). The *Agropyron cristatum* karyotype, chromosome structure and cross-genome homoeology as revealed by fluorescence *in situ* hybridization with tandem repeats and wheat single-gene probes. *Theor. Appl. Genet.* 131 (10), 2213–2227. doi: 10.1007/s00122-018-3148-9

Schubert, I., Fransz, P. F., Fuchs, J., and De Jong, J. H. (2001). Chromosome painting in plants. *Methods Cell Sci.* 23, 57–69. doi: 10.1023/A:1013137415093

Simmonds, N. W., and Shepherd, K. (1955). The taxonomy and origins of the cultivated bananas. *J. Linn. Soc. Bot.* 55, 302–312. doi: 10.1111/j.1095-8339.1955.tb00015.x

Simmonds, N. W. (1956). Botanical results of the banana collecting expeditions, 1954-5. *Kew Bull.* 11, 463–489. doi: 10.2307/4109131

Speicher, M. R., Ballard, S. G., and Ward, D. C. (1996). Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat. Genet.* 12 (4), 368–375. doi: 10.1038/ng0496-368

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 6814, 796–815. doi: 10.1038/35048692

The International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463 (7282), 763–768. doi: 10.1038/nature08747

Valárik, M., Šimková, H., Hřibová, E., Šafář, J., Doleželová, M., and Doležel, J. (2002). Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (*Musa* spp.). *Chromosome Res.* 10 (2), 89–100. doi: 10.1023/A:1014945730035

Xin, H., Zhang, T., Han, Y., Wu, Y., Shi, J., Xi, M., et al. (2018). Chromosome painting and comparative physical mapping of the sex chromosomes in *Populus tomentosa* and *Populus deltoides*. *Chromosoma.* 127, 313–321. doi: 10.1007/s00412-018-0664-y

Zimin, A. V., Puiu, D., Hall, R., Kingan, S., Clavijo, B. J., and Salzberg, S. L. (2017). The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience.* 6 (11), 1–7. doi: 10.1093/gigascience/gix097

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: 27 November 2019
doi: 10.3389/fpls.2019.01541

The Genome Sequence of *Gossypioides kirkii* Illustrates a Descending Dysploidy in Plants

Joshua A. Udall[1], Evan Long[2], Thiruvarangan Ramaraj[3,4], Justin L. Conover[5], Daojun Yuan[5,6], Corrinne E. Grover[5], Lei Gong[7], Mark A. Arick II[8], Rick E. Masonbrink[9], Daniel G. Peterson[8] and Jonathan F. Wendel[5]**

[1] Crop Germplasm Research, USDA, College Station, TX, United States, [2] Plant Breeding and Genetics, Cornell University, Ithaca, NY, United States, [3] National Center of Genome Resources, Santa Fe, NM, United States, [4] School of Computing, DePaul University, Chicago, IL, United States, [5] EEOB Department, Iowa State University, Ames, IA, United States, [6] College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [7] Key Laboratory of Molecular Epigenetics of the Ministry of Education, Northeast Normal University, Changchun, China, [8] Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Mississippi State, MS, United States, [9] Genome Informatics Facility, Iowa State University, Ames, IA, United States

OPEN ACCESS

Edited by:
Martin A. Lysak, Masaryk University, Czechia

Reviewed by:
Robin Van Velzen, Wageningen University & Research, Netherlands
Tae-Soo Jang, Chungnam National University, South Korea
Adam Lukaszewski, University of California, Riverside, United States

***Correspondence:**
Joshua A. Udall
Joshua.udall@usda.gov
Jonathan F. Wendel
jfw@iastate.edu

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

Received: *16 May 2019*
Accepted: *05 November 2019*
Published: *27 November 2019*

Citation:
Udall JA, Long E, Ramaraj T, Conover JL, Yuan D, Grover CE, Gong L, Arick MA II, Masonbrink RE, Peterson DG and Wendel JF (2019) The Genome Sequence of Gossypioides kirkii Illustrates a Descending Dysploidy in Plants. Front. Plant Sci. 10:1541. doi: 10.3389/fpls.2019.01541

One of the extraordinary aspects of plant genome evolution is variation in chromosome number, particularly that among closely related species. This is exemplified by the cotton genus (*Gossypium*) and its relatives, where most species and genera have a base chromosome number of 13. The two exceptions are sister genera that have n = 12 (the Hawaiian *Kokia* and the East African and Madagascan *Gossypioides*). We generated a high-quality genome sequence of *Gossypioides kirkii* (n = 12) using PacBio, Bionano, and Hi-C technologies, and compared this assembly to genome sequences of *Kokia* (n = 12) and *Gossypium* diploids (n = 13). Previous analysis demonstrated that the directionality of their reduced chromosome number was through large structural rearrangements. A series of structural rearrangements were identified comparing the *de novo G. kirkii* genome sequence to genome sequences of *Gossypium,* including chromosome fusions and inversions. Genome comparison between *G. kirkii* and *Gossypium* suggests that multiple steps are required to generate the extant structural differences.

Keywords: speciation, chromosome evolution, cotton, structural rearrangements, *Gossypieae*

INTRODUCTION

One of the extraordinary aspects of plant genomes is how variable they are in terms of chromosome number. Haploid chromosome counts among angiosperms span more than two orders of magnitude, from a low of n = 2 in six different species spread among four angiosperm families (Vanzela et al., 1996; Roberto, 2005), to 320 in the genus *Sedum* (Crassulaceae) (Uhl, 1978). Driving this diversity are mechanisms that both expand and shrink chromosome numbers, either saltationally *via* polyploidy, or in a more stepwise fashion *via* ascending or descending dysploidy. These processes have long been recognized as important in speciation (Stebbins, 1971; Grant, 1981) because of the impact of chromosome number divergence on reproductive isolation. Reflective of this, it is not uncommon for congeneric species to display either ascending or descending chromosome counts. From a mechanistic perspective, ascending or descending dysploidy can arise from several chromosome rearrangement processes (Jones, 1998; Guerra, 2008; Heslop-Harrison and Schwarzacher, 2011;

Lysák and Schubert, 2013; Weiss-Schneeweiss and Schneeweiss, 2013; Hoang and Schubert, 2017), including ascending dysploidy *via* chromosome fission along with the evolution of neocentromeres (Giannuzzi et al., 2013; Lysák and Schubert, 2013), and descending dysploidy through various chromosome fusion processes, including the difficult to distinguish telomere-to-telomere fusions and Robertsonian translocations (Schubert, 1992; Lysák and Schubert, 2013; Chiatante et al., 2017; Jarvis et al., 2017), and the acquisition of chromosome segments into other chromosomes (Luo et al., 2009; Murat et al., 2010; Vogel et al., 2010; Wang and Bennetzen, 2012; Fonsêca et al., 2016).

A prerequisite for understanding the directionality of chromosome number change in any taxonomic group is the availability of a well-established phylogenetic framework, so that hypotheses regarding ancestral and derived conditions are phylogenetically justified. Illustrative of this is the small monophyletic tribe *Gossypieae*, which contains the economically important cotton genus (*Gossypium*) as well as eight other lesser known genera (including *Thepparatia*) (Fryxell, 1979; Seelanan et al., 1997; Phuphathanaphong, 2006). More than 20 years ago, the Hawaiian *Kokia* and the East African/Madagascan *Gossypioides* were shown to belong to a single clade (**Figure 1**). Because these two genera have one fewer chromosomes (n = 12) than their sister genus *Gossypium* (n = 13), and because this assemblage is nested within other genera (e.g., *Hampea*, *Thespesia*) with a chromosome number of 13, they proposed an explanation involving aneuploid reduction in the lineage leading to *Kokia* and *Gossypioides* after divergence of this branch from *Gossypium*. Temporal perspectives to this reduction are the recent divergence time estimates of 5 million years (MY) for *Kokia* and *Gossypioides* and about 10 MY for the divergence of this clade from *Gossypium* (Wendel and Cronn, 2003).

Here we describe the genomic consequences of descending dysploidy in the *Kokia/Gossypioides* clade. We present a high quality *de novo* genome assembly for *Gossypioides kirkii* and compare this assembly to *Gossypium*, for which multiple assemblies have been generated. Comparison of our high quality genome assembly to other *Gossypium* genomes suggests that aneuploid reduction was accompanied by chromosome fusion and other structural rearrangements. Assuming the *Gossypium* genome was representative of the ancestral genome, we developed a model of aneuploid reduction that included several structural rearrangements reducing three chromosomes to two chromosomes during the evolution of the ancestor to the *Kokia* and *Gossypioides* genera.

MATERIALS AND METHODS

Plant Material, Sequencing, and Assembly

G. kirkii leaves were collected from the Pohl Conservatory at Iowa State University and shipped to Brigham Young University for DNA extraction. Seven PacBio cells were sequenced at BYU from two libraries created from the same DNA source. Sequence reads were assembled (**Table S1**) using Canu V1.6 (Koren et al., 2017).

Leaf tissue of *G. kirkii* was also shipped to Phase Genomics (Seattle, WA) for DNA extraction and construction of HiC sequencing libraries. The sequenced HiC libraries generated 47× coverage of 125 bp paired-end Illumina reads; these were used to establish connections between contigs (**Table S2**). Illumina reads were mapped to the reference genome and a proximity guided assembly (PGA) was performed by Phase Genomics. High-molecular weight (HMW) DNA was extracted and labeled following the Bionano Plant protocol for the Irys system.

The optical map was aligned to the PGA assembly using an *in silico* labeled reference sequence. Conflicts between the Bionano map and the PGA assembly were manually identified in the Bionano Access software by comparing the mapped Bionano contigs and reference sequence to a bed file containing sequence contigs. These inter-species alignments, along with the Bionano alignments, guided manual rearrangements of scaffolds. Corrections to the PGA assembly removed conflicts between datasets by repositioning and reorienting sequence contigs in PGA ordering files. Corrections to the HiC scaffolding were made if more than one other genome agreed with the rearrangement and if the rearrangements coincided with contig breakpoints (i.e. scaffolding rearrangements). The contig order was arranged

FIGURE 1 | The clade containing species (*Gossypioides kirkii* and *Kokia drynarioides*) with n = 12 is nested among genera that have n = 13, suggesting that these two species have one fewer chromosome compared to their close relatives. Six different species are used as examples. The haploid chromosome number for each species and group is indicated in the yellow box. Aggregate geographic distribution (Mad. refers to Madagascar) and species richness (number of species in parentheses) are shown next to each genus. Phylogenetic tree is based off Wendel et al. (2002).

to maximize the frequency of close linkages throughout the genome. The resulting fasta file of the scaffolded assembly was produced by concatenating PacBio contigs with 100 N bases between them. Several iterations of correction and realignment resolved nearly all of the conflicts between the sequence and Bionano assemblies. Similar iterations of HiC interaction maps were created using Juicer v1.5 and Juicebox v1.8.8, respectively, for the final manual adjustments to the genome sequence. Specifically, HiC reads were re-mapped to the modified sequence and the association frequency between each paired-end was used to adjust the genome sequence using JuiceBox (Durand et al., 2016). A custom python script from Phase Genomics was used to adjust the initially assembled pseudomolecules with the changes made to the genome *via* JuiceBox. Based on the HiC data, the *G. kirkii* pseudomolecule corrections consisted of two inversions and three translocations (involving seven of 12 chromosomes). These corrections established near-complete congruence between mapped paired-ends along the entire genome. The *G. kirkii* genome sequence is available from GenBank (Accession numbers: CP032244–CP032255).

Genome Alignments

The *G. kirkii* assembly was separately aligned by Minimap2 (Li, 2018) to other genomes in *Gossypium*, including *G. arboreum* (Du et al., 2018), *G. raimondii* (Paterson et al., 2012), and *G. hirsutum* (Zhang et al., 2015), and visualized using dotPlotly (Poorten, 2018). The alignments identified assembly errors in chromosomes Chr09 and Chr12 of *G. raimondii*. Telomere sequences were also used to confirm assembly completeness and structure by searching for the canonical telomere repeat (McKnight et al., 1997; Fajkus et al., 2005; Watson and Riha, 2010) in the *G. kirkii* genome using Geneious (Biomatters, New Zealand). The telomere repeat were also visualized and manually annotated in Geneious to verify telomere location on each chromosome.

Phase Genomics also constructed and sequenced a Hi-C library made from leaf tissue of *Kokia drynarioides*, a member of the genus sister to *Gossypioides* that also shares n = 12, to further verify the structure. The Hi-C reads were mapped to the final, corrected version of the *G. kirkii* genome assembly using BWA. Approximately 10.7 M contacts (11% of the total paired reads) passed mapping filters and were used as Hi-C interaction evidence. Contact maps were visualized using Juicer v1.5 and Juicebox v1.8.8.

Transcriptomic Sequencing and Gene Annotation

Total RNA was extracted from 3-cm seedling leaves. Illumina TruSeq RNA-sequencing libraries were prepared for each replicate and were sequenced (Paired-end 150 bp) Berry Genomics Co. Ltd. (Beijing, China). Gene annotations were created using GenSAS 5.0 (Lee et al., 2011), an online integrated genome sequence annotation pipeline. BUSCO analysis was conducted to test for annotation completeness. Repetitive elements were detected by RepeatModeler (Smit and Hubley, 2013) and RepeatMasker (Smit et al., 2019). AgriGO tested for enrichment of Gene Ontology categories of gene functions in the rearranged segments (Tian et al., 2017). RNA-seq of *G. kirkii* is available from GenBank under SRX5894875. Gene and repetitive annotations are available from CottonGen under https://www.cottongen.org/analysis/213.

Analysis of Paleo-Genome Duplications

Protein sequences of *G. kirkii* and the D_T genome of *G. hirsutum* were clustered using OrthoFinder v.2.1 (Emms and Kelly, 2015) with the Diamond alignment tool. Single copy orthologs from OrthoFinder were used as input to MCScanX_h (.homology file), with default settings (Wang et al., 2012). The collinearity plots between chromosomes Chr02 (Chr15, if the tetraploid chromosomes were numbered sequentially), Chr04 (Chr17) and Chr06 (Chr19) in *G. hirsutum* and chromosomes KI_2_4 and KI_06 in *G. kirkii* were created using the circle_plotter downstream tool of the MCScanX package. From the OrthoFinder output, all of the intraspecific paralogs were extracted for *G. kirkii*. Within each group of putatively orthologous genes, Ks values for every possible pairwise combination of paralogs were calculated using the codeml package of PAML, using custom python scripts.

ChIP-seq

Leaves and leaf buds were also collected from *G. kirkii* (specimen voucher ISC 418555, Ada Hayden Herbarium, Iowa State University). Rabbit polyclonal CenH3 antibody was made to the CenH3 amino acids 9–20 and conjugated to KLH (Covance, Inc.), a conserved peptide in *Gossypium* species of CenH3. Immunostain on *G. raimondii* root tips ensured centromere specificity of the CenH3 antibody. Chromatin immunoprecipitation was performed using the Epigentek EpiQuik Plant ChIP Kit (P-2014) with modifications. DIECA (2%) and PVP-40 (4%) were added to the fixative and to final solutions of CP3C, CP3D, and CP3E. DNA samples were sonicated at 60% amplitude for three total minutes of sonication/rest (15 s/15 s). Divided samples were incubated with either rabbit pre-immune sera, anti-CenH3, or polyclonal H3K9ac (ABCam, ab10812, LOT GR171780). Four replications of each reaction was pooled for whole genome amplification using the SeqPlex Enhanced DNA Amplification Kit (SeqXE, Sigma) then sequenced (Illumina PE150 bp) at the Beijing Genomics Institute (BGI). ChIP-seq reads were mapped to the genome using BWA (Li and Durbin, 2009). ChIP-seq data are available from NCBI under SRX5894872–SRX5894874.

FISH

Preparation of chromosomes and staining were performed as previously described for maize (Masonbrink and Birchler, 2010). FISH was performed as specifically described for cotton (Wang et al., 2006).

RESULTS

Sequencing and *De Novo* Assembly of the *G. kirkii* Genome

Two different genome technologies were used to assemble the *G. kirkii* genome sequence (**Figure 1**). First, approximately

68× coverage of raw SMRT data (40 Gb) was generated using the PacBio Sequel System (**Table S1**). The contig-level assembly was 544 Mb composed of 389 contigs with a contig N50 of 9.92 Mb and a maximum contig size of 31.1 Mb (**Table 1**). After scaffolding with HiC (Burton et al., 2013), the 12 pseudomolecules assembly was 92.5% of the expected genome size of 588 Mb (Wendel et al., 2002) with only 277 gaps (**Table 1**, **Table S2** and **Figure S1**). Chromosomes were manually adjusted (**Figures S2** and **S3**) and named according to the convention used in *Gossypium hirsutum* (Zhang et al., 2015). These pseudomolecules represented the 12 chromosomes of the *G. kirkii* genome (**Figure 2**, **Table S3**). Chromosome KI_2_4 contained the largest number of sequence contigs (65 contigs, 41.5 Mb) and Chromosome KI_08 contained the fewest (seven contigs, 39.6 Mb), even though these two chromosomes contained approximately the same total sequence length. Chromosome KI_06 was the largest chromosome (see below).

TABLE 1 | Assembly metrics of the *Gossypioides kirkii* genome.

Genome Statistics	Stats for Contigs	Stats for Pseudomolecules (Hi-C)	Stats for Chromosome Assembly (BioNano)
Assembly size (in Mb)	544	538	354
Number of sequences (number of gaps)	389 (0)	12 (277)	1,079 (3,721)
Longest Scaffold Length (Mb)	31.1	60.3	16.7
N50 in Mb (mean number of contigs per chromosome)	9.92 (17)	42.97 (17)	0.296

An optical map (Bionano Genomics, Inc.) was used to validate the assembly of individual contigs and the HiC connections between contigs (**Table S4**). Optical map data typically serves as an independent validation of the assembled sequence because the image data of Bionano labeled DNA molecules is assembled independently and aligned to DNA sequences using restriction patterns matching the labels in the Bionano contigs (Udall and Dawe, 2017). While the percentage of alignments between optical maps and contigs was relatively low, we note that over half of the genome sequence was validated by optical map alignment. The Bionano alignments also spanned 62% of the 71 eligible sequence gaps (i.e. gaps flanked by contigs >100 kb on each side, since

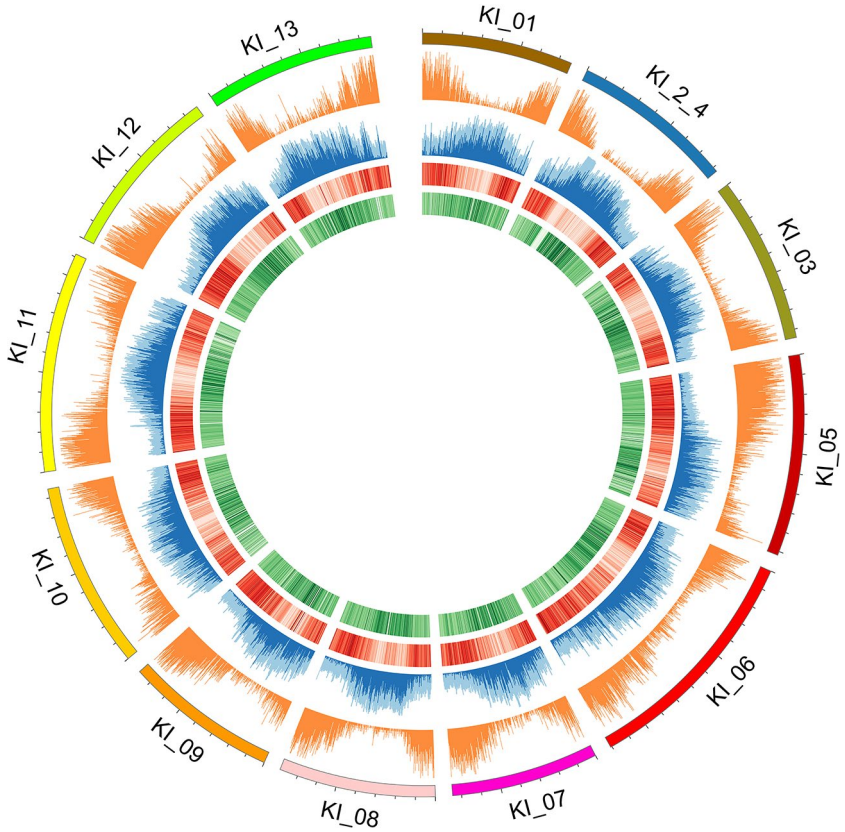

FIGURE 2 | Individual chromosomes (KI_ labels) of the *G. kirkii* genome are illustrated by 5 tracks in a Circos plot. Darker shades of colors represent a higher value or frequency of genomic features with the 100 kb window. From outside to inside: chromosome graph and scale; plot of gene density; TE content (light blue is total TE content, dark blue is Gypsy content), ChIP-seq of H3K9ac (Darker red lines of H3K9ac indicate a higher frequency of H3K9 acetylation); and ChIP-seq of CENH3 (Darker green lines of the CENH3 track indicate a higher frequency of CENH3 binding). Centromeres are inferred in the regions where H3K9ac is low (light red) and CENH3 is high (dark green).

Bionano contigs do not generally match smaller contigs due to limitations in nick-pattern matching) (**Dataset S1**).

A common measure of genome quality is the percentage of expected genes recovered in an annotated assembly. Here, the percentage of genes identified in the *G. kirkii* genome sequence provided confidence that nearly the entire genome was represented, with 95% (1,364/1,440) of conserved genes from BUSCO identified (Simão et al., 2015). The remaining genes were either fragmented (n = 18, 1%) or missing (n = 58, 4%). That a few of the BUSCO genes were missing is not surprising due to the previously reported genome downsizing and gene loss in this species (Grover et al., 2017), and therefore may reflect a combination of genome completeness as well as historical evolution. Of the 36,669 gene annotations, 64% had RNA-seq reads (> 20 reads) mapping to them, suggesting that we assembled much of the leaf transcriptome.

Collinearity With Other *Gossypieae* Genomes

The integrity of the *G. kirkii* genome assembly was also assessed by comparing it to genome sequences recently published for *Gossypium* (Paterson et al., 2012; Du et al., 2018; Wang et al., 2019) (**Figures S4**–**S7**). Occasionally, we used these comparisons to correct scaffolding errors in the *G. kirkii* genome if the *G. kirkii* contigs and optical map contigs supported such corrections. These manual rearrangements (**Dataset S2**) utilized evidence from both contig ends of each initial non-colinear placement of *G. kirkii* sequence.

We further assessed the completeness of the assembly and the orientation of terminal scaffolds by searching for telomere sequences in the *G. kirkii* pseudomolecules. We identified 20 loci with characteristic sequence of telomere repeats at the ends of our pseudomolecules (**Table S5**); eight pseudomolecules had telomere repeats on both chromosome arms, four had telomere repeats on a single arm, and two pseudomolecules had telomere repeats that were confidently embedded within a single scaffold. The longest telomeric repeat (> 24 kb) was identified on KI_04. Since this length was longer than most of the length of our trimmed reads used for assembly (N50 = 16,192), it is likely that many reads containing a majority of telomeric sequence collapsed during sequence assembly. Indeed, these regions had a higher read coverage compared to the adjacent chromosome sequence (data not shown). Different telomere sequences were identified in different combinations on each of the chromosome ends, suggesting the existence of multiple telomerases or at a minimum multiple guide RNAs in *Gossypioides*.

Because typical centromeres do not have conserved sequences (Ma et al., 2007; Lysak, 2014; Birchler and Han, 2018), we leveraged additional data to identify centromeric regions. That is, we evaluated the density of both ChIP-seq reads and gene density to infer putative centromeric regions. Euchromatic and histone modifications of H3K9ac and CENH3, respectively were used to estimate centromeric regions (Masonbrink et al., 2014). Typical distributions of epigenetic marks were identified (e.g. increasing frequency of CENH3 marks near the centromeric regions, **Figure 2**). In some cases, chromosomes had a single contig assembled across the centromeric region (e.g. chromosomes KI_06, KI_10, KI_11) suggesting proper assembly and density of CENH3 marks in centromeric regions. The centromeric regions of other chromosomes contained multiple contigs. While their assembly depended on both correct sequence assembly and correct scaffolding, their density of CENH3 marks was similar to those regions composed of a single contig.

Annotation of Genes and Repetitive Elements

Gene annotation recognized 36,669 genes, somewhat higher than previously reported (Grover et al., 2017); these differences are likely due to both genome quality and annotation method. All *G. kirkii* genes were aligned to their closest intragenomic paralog to calculate synonymous substitutions (Ks); the plot of these pairwise Ks values exhibits a peak congruent with previous findings (Conover et al., 2019) of an ancient polyploidization event shared with *K. drynarioides* and all members of *Gossypium* (**Figure S8**). Because genes comprise useful genomic anchors, gene annotations were used to inform analyses of the chromosome rearrangements in *G. kirkii* (below).

Repetitive elements were detected by RepeatModeler (Smit and Hubley, 2018) and RepeatMasker (Smit et al., 2019). As a whole, the genome contained ~30% interspersed repeats and 1.7% simple repeats. The interspersed repetitive elements corresponded to transposable elements, namely *Gypsy* and *Copia* retrotransposons (**Table S3**). We detected the TEs on each chromosome to assess the class distribution of TE elements throughout the genome (Bailly-Bechet et al., 2014). While TEs can be associated with chromosome rearrangements, we found no bias in terms of TE number, total length, or class between chromosomes. In general, the number and total length of *Gypsy* elements greatly outweighs *Copia* elements, as is common for many plant genomes (**Table S3**).

Comparative Genomics Between *G. Kirkii* and Related Genomes

The base chromosome number (x) of *G. kirkii* and *K. drynarioides* is x = 12, but the remainder of the cotton tribe (*Gossypieae*) in which this lineage is nested has a base chromosome number of x = 13 (**Figure 1**). To explore which chromosomes may be involved in this derived state, we identified chromosome rearrangements that occurred after divergence between *G. kirkii* and *Gossypium* (represented by the ancestral "G" chromosomes). In this analysis, the genome of *Gossypium* was assumed to represent the ancestral genome to the *Gossypium*–*Gossypioides*–*Kokia clade*, and the *G. kirkii* genome was considered derived due to the presence of necessary changes during chromosome reduction, although we cannot discount the possibility of some structural changes in *Gossypium*. Whole genome comparisons suggested that an entire arm of chromosome G2 and an entire arm of chromosome G4 (intact within modern-day *G. raimondii* and *G. arboreum*) fused to form a single chromosome, while the other chromosome arms were fragmented and inserted into KI_06 (**Figure 3A**). A comparison of annotated genes to *G. hirsutum* in these regions

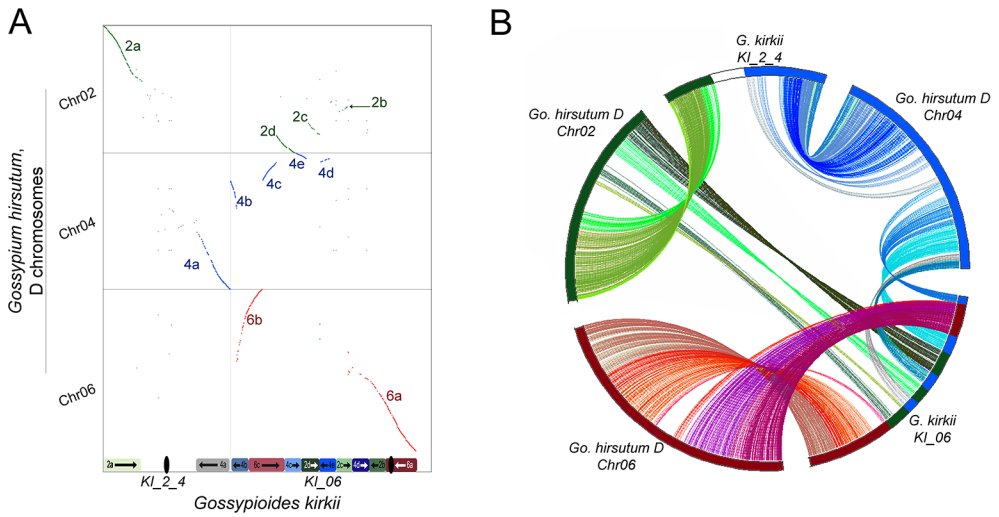

FIGURE 3 | (A) Whole-genome dot-plot alignment between the aneuploid-reduced chromosomes of *Gossypioides kirkii* (Chr2_4 and Chr06; x-axis) and the ancestral state, represented here by the D_T-genome of *Gossypium hirsutum* (Chr02, Chr04, Chr06; y-axis) because of an assembly error in *Gossypium raimondii* affecting a key chromosome. Diagrams of the current chromosome configurations are represented next to each of their respective axis **(B)** Circos plot illustrating the rearrangement based on conserved single copy orthologs. Colors in each plot were produced to illustrate matching segments between the whole-genome and gene-based illustrations.

also supports our inferred genome alignments (**Figure 3B**). The insertion of these G2 and G4 fragments into KI_06 explain the absence of a single chromosome that is twice the size of other metacentric chromosomes, as might be expected if a simple chromosome fusion had occurred (Hutchinson, 1943). We confirmed the absence of an unusually long chromosome in *G. kirkii* by chromosome staining (**Figure S9**). The inserted portion on KI_06 consists of alternating segments of ancient chromosomes G2 and G4 with six segments accounting for approximately 30 MB of the chromosome. Details of the rearranged segments are found in **Table S6**. These segments each contained between 117 and 458 genes. A GO enrichment test of each segment found no enriched GO categories.

These findings may be summarized as three salient facts regarding the genomic history of *G. kirkii*. First, one chromosome arm each of G2 and G4 were inserted into what became part of *Gossypioides kirkii* chromosome KI_06. Second, before or after the insertion, segments of these two chromosome arms were interleaved through unknown evolutionary processes. It is worth noting that all of the G2/G4 'junctions' in KI_06 have strong support of PacBio and Bionano coverage. Third, the remaining, entire chromosome arms of G2 and G4 fused to create KI_2_4. We further support these inferences by mapping a *K. drynarioides* HiC library to the *G. kirkii* assembly. The resulting HiC contact heatmap (**Figure S10**) also shows a linear contact pattern along KI_06 and KI_2_4 suggesting that the chromosome rearrangements we describe are shared between these sister genera *Kokia* and *Gossypioides*.

DISCUSSION

Among the many opportunities afforded by genome sequencing is the possibility of gaining insight into long-standing cytogenetic phenomena that remain unexplained at the sequence level. A promising example is dysploid evolution, which is a well-known and common pattern of cytogenetic variation in both plants and animals (Grant, 1971; Stebbins, 1971; White, 1973). From a mechanistic standpoint, it has long been thought that dysploidy arises primarily from chromosome translocations. This view was promulgated in George Ledyard Stebbins' 1971 classic *Chromosomal Evolution in Higher Plants*, in which he stated that "*aneuploid alterations of the basic chromosome number are usually the outcome of successive translocations*" (5, pg. 86). Similarly, in Verne Grant's widely used 1971 textbook "*Plant Speciation*", he stated "*the mechanism of aneuploid reduction at the diploid level involves unequal reciprocal translocations*" (4, pg. 359). More recently, telomeric (end-to-end) fusion and Robertsonian translocation have been recognized as processes leading to aneuploid reduction (Schubert, 1992; Lysák and Schubert, 2013; Chiatante et al., 2017; Jarvis et al., 2017), as has the insertion of one chromosome into another (Luo et al., 2009; Murat et al., 2010; Vogel et al., 2010; Wang and Bennetzen, 2012; Fonsêca et al., 2016). Remarkably, while we were completing the present work, Birchler and Han (Birchler and Han, 2018) published a thought-provoking explication of how the Breakage-Fusion-Bridge cycle, as illuminated by McClintock 80 years ago for understanding various chromosome anomalies in maize (McClintock, 1939; McClintock, 1941), likely has causal connections to common mechanisms of karyotypic evolution in plants, and by extension possibly all eukaryotes.

Here we provide sequence-based evidence for chromosome number reduction where related members of the cotton tribe establish the polarity of the descending dysploidy (from x = 13 to x = 12). The foundation for our conclusions is the high-quality assembly of the *G. kirkii* genome sequence presented here. The accuracy of this assembly was determined by multiple congruent datasets (PacBio, HiC, and Bionano) and

by comparative analyses that demonstrate consistency with previously published cotton genomes (Paterson et al., 2012; Zhang et al., 2015; Du et al., 2018; Wang et al., 2019). Analyses of colinearity revealed a complex pattern of inter-digitating chromosome segments. The identified rearrangements also are congruent with previous cytogenetic observations of *Gossypioides brevilanatum*, in the fact that *G. kirkii* does not display an 'extra-large' chromosome (Hutchinson, 1943) as might be expected from a simpler scenario of a 2 to 1 chromosome fusion event.

Explanations for the inferences depicted in **Figure 3** for the derivation of the reduced chromosome number in *G. kirkii* ("KI" chromosomes) relative to the G chromosomes of ancestral *Gossypieae* need to account for the following observations: (1) identification of end-to-end G2 and G4 (2d and 4e) segments (including internal telomeric sequences) in KI_06 that implicate a historical end-to-end fusion of ancestral chromosome arms G2 and G4 (2) chromosome KI_2_4 contains entire chromosome arms of G2 and G4, suggestive of chromosome fusion at, or close to, the centromeres for the intact G2 and G4 arms; and (3) because the terminal inversion on KI_06 included 2.0Mb of the G4 chromosome (in addition to 8.4 Mb of the original G6 chromosome arm), it must have occurred after the insertion event above.

We recognize that by assuming the *Gossypium* genome represented the ancestral *Gossypieae* genome, unique *Gossypium* changes were confounded with the differences between *G. kirkii* and *Gossypieae*. We are comfortable with this assumption based on previous cytological work that prompted the previous generation of botanists to coin the phrase 'cryptic structural differentiation' when working with the cotton tribe (Fryxell, 1979). They understood the chromosomes were different based on pairing data, but observable structural differentiation was not sufficient to differentiate members of the cotton tribe, other than the descending dysploidy of *Gossypioides* and *Kokia*.

While other interpretations could be made when additional genomes from tribe *Gossypieae* (e.g., *Thespesia*, *Hampea*, or *Lebronnecia*) are sequenced, we use the above three key

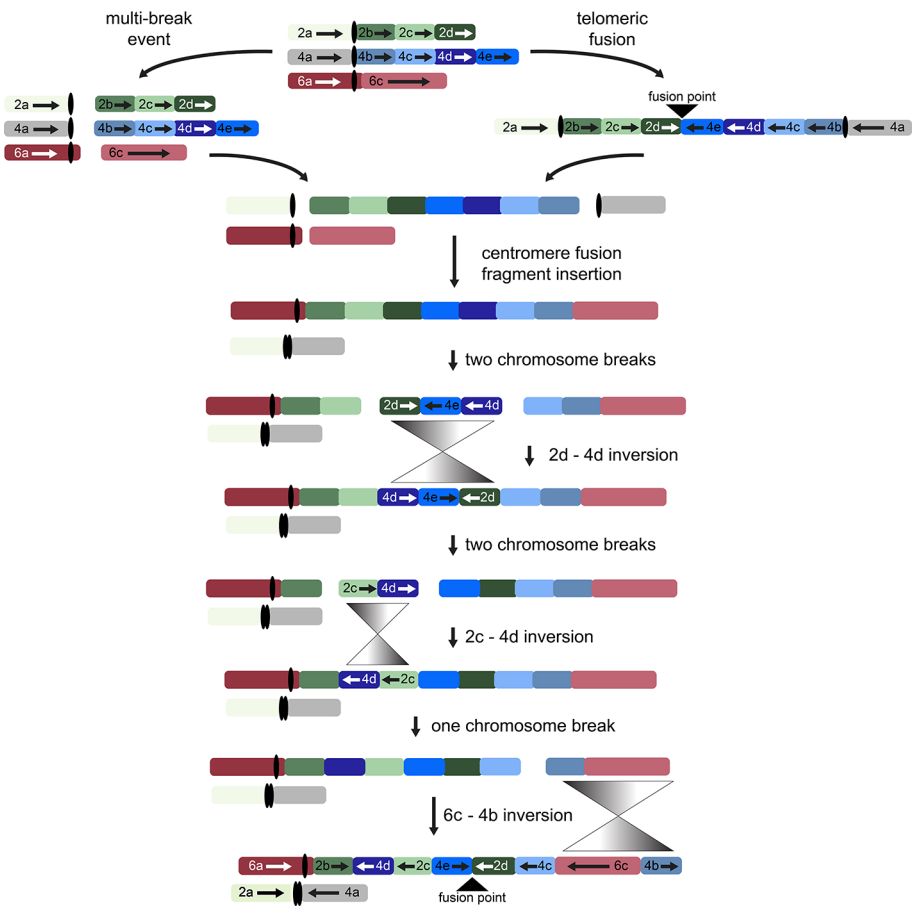

FIGURE 4 | Possible evolutionary model for the origin of descending dysploidy in the ancestor of *Gossypioides* and *Kokia* (x = 12) from a progenitor with n = 13. Ancestral chromosomes involved in the aneuploid reduction are pictured at the top. Two possible paths are shown, which include a multi-break event near the centromeres of each ancestral chromosome (left) or an end-to-end fusion of ancestral chromosomes G2 and G4 (right). After either a multi-break event (left) or the generation and subsequent breakage of a dicentric chromosome, chromosome segments 2a and 4a fused to generate one chromosome, while the remaining fragments of G2 and G4 were inserted between segments of G6 (here, 6a and 6c). Three inversions (grey triangles) are required to rearrange the order of the original chromosome blocks into the pattern seen in the extant *G. kirkii* genome.

observations (and one modest assumption) to create a hypothesis for the order of events following initial dicentric chromosome formation (**Figure 4**). The end-to-end fusion of G2 and G4 strongly supports evolutionary models that begin with a dicentric chromosome, although we note that a multi-break-fusion event could bypass the need for a dicentric chromosome (as noted in **Figure 4**). Myriad alternative models are also possible where end-to-end fusion are coincidental instead of contributional; however, they are not considered further because of the key evidence of the telomeric sequence and directionality of the end-to-end fusion fragments.

Chromosome comparisons between *Gossypioides* and *Gossypium* suggest that the origin of *G. kirkii* KI_06 involved both fusion and a series of inversions to generate the observed interleaved pattern (**Figure 4**). While this fusion could have been the result of breaks occurring on each of the involved chromosomes, the presence of internal telomeres supports an end-to-end fusion, generating a dicentric chromosome. Perhaps, the nascent dicentric chromosome somehow contributed to Subsequent breaks near each centromere (G2- and G4-derived). If an additional break was concurrent in G6, then translocations followed by subsequent paracentric inversions could create the extant chromosomes of *Gossypioides*. As depicted in **Figure 4**, two-breaks of a dicentric chromosome created an acentric fragment containing most of the arms of G2 and G4, which inserted into G6, and centromeric fusion between the G2- and G4-chromosome arms containing centromeric sequence. Three inversions are then required to transform the initial fusion of *G. kirkii* Chr06 into the extant chromosome morphology. The two unshared inversions would only involve portions of the inserted segment. Notably, the two chromosome inversions of *G. kirkii* were approximately 9.6 Mb and 12.7 Mb, respectively (**Table S6**), which is similar to the average inversion size for plants and animals (i.e., 8.4 Mb, (Wellenreuther and Bernatchez, 2018)). While inversions often are associated with TEs (Kidwell and Lisch, 2000), we do not find an increased density of TEs in KI_06 (**Figure 2**) to support this for *G. kirkii*. Although the responsible inversion mechanism is not known, it is possible that recombination between a hemizygous insertion KI_06 and a normal KI_06 could have played a role.

The foregoing hypothesis explains a novel "3" to "2" route for chromosome number reduction, as opposed to the more conventional "2" to "1". It certainly invokes a series of seemingly unlikely events, including formation of multiple inversions requiring two simultaneous double-strand breaks and repair (Kirkpatrick, 2010), either through a known mechanism such as breakage-fusion-bridge or unknown accidents of aberrant recombination (as described here). Because the likelihood of each rare event multiplies when each is considered as independent of the others, perhaps it is more parsimonious to postulate that chromosome number reduction occurred within a single generation, in a series of germ-line cell divisions with subsequent 'healing' in the sporophyte. In contrast, it remains possible that this entire process unfolded in a stepwise fashion during long evolutionary timescales. Unfortunately, we lack surviving intermediates that might testify to this temporal possibility, and we are unaware of other methods that might be used to distinguish between the "fast" and "slow" scenarios. Several other studies have detected aneuploid reduction between related species in the Brassicaceae (Lysak et al., 2006), or in the genomes of grasses (Zhang et al., 2008; Murat et al., 2010; Wang et al., 2015; Luo et al., 2017), based on patterns of FISH or using sequence comparisons, and others have noted instances of "chromosome shattering" with possible mechanisms (Zhang et al., 2008; Tan et al., 2015; Mandáková et al., 2019). As more plant and animal genomes are sequenced and assembled by robust methods, the spectrum of causative mechanisms and their frequency in explaining patterns of karyotypic evolution are likely to become much clearer.

DATA AVAILABILITY STATEMENT

The *G. kirkii* genome sequence is available on GenBank (Accession numbers: CP032244-CP032255).

AUTHOR CONTRIBUTIONS

JU and JW developed the idea. JU, JW, JC, and CG designed the project. EL, DY, LG, MA, RM, and DP generated the data. JU, TR, JC, DY, CG, and MA analysed the data. JU, JC, CG, and JW wrote the manuscript. All authors read, edited, and approved the final manuscript.

FUNDING

Primary funding was provided by the National Science Foundation Plant Genome Research Program (#1339412). Additional support was provided from Cotton Incorporated and USDA-ARS (58-6402-1-644 and 58-6066-6-59).

ACKNOWLEDGMENTS

We thank the BYU Fulton SuperComputer lab for use of their computation resources. We thank ResearchIT (https://researchit.las.iastate.edu/) for computational support at Iowa State University. We thank Rise Services for office accommodations in Orem, UT.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01541/full#supplementary-material

Supplementary material for this article can be found in (Udall_Supplemental.docx).

REFERENCES

Bailly-Bechet, M., Haudry, A., and Lerat, E. (2014). "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. *Mob. DNA* 5, 13. doi: 10.1186/1759-8753-5-13

Birchler, J. A., and Han, F. (2018). Barbara McClintock's unsolved chromosomal mysteries: parallels to common rearrangements and karyotype evolution. *Plant Cell* 30, 771–779. doi: 10.1105/tpc.17.00989

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt.2727

Chiatante, G., Giannuzzi, G., Calabrese, F. M., Eichler, E. E., and Ventura, M. (2017). Centromere destiny in dicentric chromosomes: new insights from the evolution of human chromosome 2 ancestral centromeric region. *Mol. Biol. Evol.* 34, 1669–1681. doi: 10.1093/molbev/msx108

Conover, J. L., Grover, C. E., Wendel, J. F., Karimi, N., Stenz, N., Ané, C., et al. (2019). A Malvaceae mystery: a mallow maelstrom of genome multiplications and maybe misleading methods? *J. Integr. Plant Biol.* 61 (1), 12–31. doi: 10.1111/jipb.12746

Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* 50 (6), 796–802. doi: 10.1038/s41588-018-0116-x

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3 (1), 99–101. doi: 10.1016/j.cels.2015.07.012

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. doi: 10.1186/s13059-015-0721-2

Fajkus, J., Sýkorová, E., and Leitch, A. R. (2005). Telomeres in evolution and evolution of telomeres. *Chromosom. Res.* 13, 469–479. doi: 10.1007/s10577-005-0997-2

Fonsêca, A., Ferraz, M. E., and Pedrosa-Harand, A. (2016). Speeding up chromosome evolution in Phaseolus: multiple rearrangements associated with a one-step descending dysploidy. *Chromosoma* 125, 413–421. doi: 10.1007/s00412-015-0548-3

Fryxell, P. A. (1979). *The natural history of the cotton tribe (Malvaceae, tribe Gossypieae). 1st ed*. College Station: Texas A&M University Press.

Giannuzzi, G., Pazienza, M., Huddleston, J., Antonacci, F., Malig, M., Vives, L., et al. (2013). Hominoid fission of chromosome 14/15 and the role of segmental duplications. *Genome Res.* 23, 1763–1773. doi: 10.1101/gr.156240.113

Grant, V. (1971). *Plant speciation*. Columbia University Press.

Grant, V. (1981). *Plant speciation*. New York: Columbia University Press.

Grover, C. E., Arick, M. A., Conover, J. L., Thrash, A., Hu, G., Sanders, W. S., et al. (2017). Comparative genomics of an unusual biogeographic disjunction in the cotton tribe (Gossypieae) yields insights into genome downsizing. *Genome Biol. Evol.* 9, 3328–3344. doi: 10.1093/gbe/evx248

Guerra, M. (2008). Chromosome numbers in plant cytotaxonomy: concepts and implications. *Cytogenet. Genome Res.* 120, 339–350. doi: 10.1159/000121083

Heslop-Harrison, J. S. P. P., and Schwarzacher, T. (2011). Organisation of the plant genome in chromosomes. *Plant J.* 66, 18–33. doi: 10.1111/j.1365-313X.2011.04544.x

Hoang, P. T. N., and Schubert, I. (2017). Reconstruction of chromosome rearrangements between the two most ancestral duckweed species Spirodela polyrhiza and S. intermedia. *Chromosoma* 126, 729–739. doi: 10.1007/s00412-017-0636-7

Hutchinson, J. B. (1943). A note on Gossypium brevilanatum Hochr. *Trop. Agric. Agric.* 20, 4.

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J. A., et al. (2017). The genome of Chenopodium quinoa. *Nature* 542, 307–312. doi: 10.1038/nature21370

Jones, K. (1998). Robertsonian fusion and centric fission in karyotype evolution of higher plants. *Bot. Rev.* 64, 273–289. doi: 10.1007/BF02856567

Kidwell, M. G., and Lisch, D. R. (2000). Transposable elements and host genome evolution. *Trends Ecol. Evol.* 15 (3), 95–99. doi: 10.1016/S0169-5347(99)01817-0

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PloS Biol.* 8 (9), e1000501. doi: 10.1371/journal.pbio.1000501 pii.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Lee, T., Peace, C., Jung, S., Zheng, P., Main, D., and Cho, I. (2011). "GenSAS - An online integrated genome sequence annotation pipeline," in *Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011*. doi: 10.1109/BMEI.2011.6098712

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. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100. doi: 10.1093/bioinformatics/bty191

Luo, M. C., Deal, K. R., Akhunov, E. D., Akhunova, a R., Anderson, O. D., Anderson, J. A., et al. (2009). Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proc. Natl. Acad. Sci. U. S. A.* 106, 15780–15785. doi: 10.1073/pnas.0908195106

Luo, M.-C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. *Nature* 551, 498. doi: 10.1038/nature24486

Lysák, M. A., and Schubert, I. (2013). "Mechanisms of Chromosome Rearrangements," in *Plant Genome Diversity Volume 2* (Vienna: Springer Vienna), 137–147. doi: 10.1007/978-3-7091-1160-4_9

Lysak, M., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., and Schubert, I. (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5224–5229. doi: 10.1073/pnas.0510791103

Lysak, M. A. (2014). Live and let die: centromere loss during evolution of plant chromosomes. *New Phytol.* 203, 1082–1089. doi: 10.1111/nph.12885

Ma, J., Wing, R., Bennetzen, J. L., and Jackson, S. A. (2007). Plant centromere organization: a dynamic structure with conserved functions. *Trends Genet.* 23, 134–139. doi: 10.1016/j.tig.2007.01.004

Mandáková, T., Pouch, M., Brock, J. R., Al-Shehbaz, I. A., and Lysak, M. A. (2019). Origin and evolution of Diploid and Allopolyploid Camelina genomes was accompanied by chromosome shattering. *Plant Cell* 31 (11), 2596–2612. doi: 10.1105/tpc.19.00366

Masonbrink, R. E., and Birchler, J. A. (2010). Sporophytic nondisjunction of the maize B chromosome at high copy numbers. *J. Genet. Genomics* 37, 79–84. doi: 10.1016/S1673-8527(09)60027-8

Masonbrink, R. E., Gallagher, J. P., Jareczek, J. J., Renny-Byfield, S., Grover, C. E., Gong, L., et al. (2014). CenH3 evolution in diploids and polyploids of three angiosperm genera. *BMC Plant Biol.* 14, 383. doi: 10.1186/s12870-014-0383-3

McClintock, B. (1939). The behavior in successive nuclear divisions of a chromosome broken at Meiosis. *Proc. Natl. Acad. Sci. U. S. A.* 25 (8), 405–416. doi: 10.1073/pnas.25.8.405

McClintock, B. (1941). The stability of broken ends of chromosomes in Zea mays. *Genetics* 26 (2), 234–282.

McKnight, T. D., Fitzgerald, M. S., and Shippen, D. E. (1997). Plant telomeres and telomerases. A review. *Biochem. (Mosc).* 62, 1224–1231. http://www.ncbi.nlm.nih.gov/pubmed/9467846.

Murat, F., Xu, J. H., Tannier, E., Abrouk, M., Guilhot, N., Pont, C., et al. (2010). Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Res.* 20, 1545–1557. doi: 10.1101/gr.109744.110

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427. doi: 10.1038/nature11798

Phuphathanaphong, L. (2006). Thepparatia (Malvaceae), a new genus from Thailand. *Thai. For. Bull.* 34, 195–200.

Poorten, T. (2018). https://github.com/tpoorten/dotPlotly.dotPlotly. Available at: .

Roberto, C. (2005). Low chromosome number angiosperms. *Caryologia* 58 (4), 403–409. doi: 10.1080/00087114.2005.10589480

Schubert, I. (1992). Telomeric polymorphism in Vicia faba. *Biol. Zentralblatt* 111, 164–168.

Seelanan, T., Schnabel, A., and Wendel, J. F. (1997). Congruence and consensus in the cotton tribe (Malvaceae). *Syst. Bot.* 22, 259. doi: 10.2307/2419457

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., and Hubley, R. (2013). doi: 10.1007/s00572-016-0720-5 RepeatModeler. RepeatModeler.

Smit, A., and Hubley, R. (2018). RepeatModeler Open-1.0. GitHub.

Smit, A. F. A., Hubley, R., and Green, P. (2019). http://www.repeatmasker.org. RepeatMasker Open-4.0. Available at: .

Stebbins, G. L. (1971). *Chromosomal evolution in higher plants*. London: Edward Arnold LTD. 87–89.

Tan, E. H., Henry, I. M., Ravi, M., Bradnam, K. R., Mandakova, T., Marimuthu, M. P., et al. (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *Elife* 4, e06516. doi: 10.7554/eLife.06516

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45 (W1), W122–W129. doi: 10.1093/nar/gkx382

Udall, J. A., and Dawe, R. K. (2017). Is it ordered correctly? Validating genome assemblies by optical mapping. *Plant Cell* 30, 7–14. doi: 10.1105/tpc.17.00514

Uhl, C. H. (1978). Chromosomes of Mexican Sedum II. Section Pachysedum. *Rhodora* 80, 491–512. http://www.jstor.org/stable/2331126. Available at: .

Vanzela, A. L. L., Guerra, M., and Luceno, M. (1996). Rhynchospora Tenuis Link (Cyperaceae), a species with the lowest number of holocentric chromosomes. *Cytobios* 88, 219–228. Available at: isi:A1996XT51300004.

Vogel, J. P., Garvin, D. F., Mockler, T. C., Schmutz, J., Rokhsar, D., Bevan, M. W., et al. (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* 463 (7282), 763–768. doi: 10.1038/nature08747

Wang, H., and Bennetzen, J. L. (2012). Centromere retention and loss during the descent of maize from a tetraploid ancestor. *Proc. Natl. Acad. Sci. U. S. A.* 109 (51), 21004–21009. doi: 10.1073/pnas.1218668109

Wang, K., Song, X., Han, Z., Guo, W., Yu, J. Z., Sun, J., et al. (2006). Complete assignment of the chromosomes of Gossypium hirsutum L. by translocation and fluorescence in situ hybridization mapping. *Theor. Appl. Genet.* 113, 73–80. doi: 10.1007/s00122-006-0273-7

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 (7), e49. doi: 10.1093/nar/gkr1293

Wang, X., Jin, D., Wang, Z., Guo, H., Zhang, L., Wang, L., et al. (2015). Telomere-centric genome repatterning determines recurring chromosome number reductions during the evolution of eukaryotes. *New Phytol.* 205 (1), 378–389. doi: 10.1111/nph.12985

Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., et al. (2019). Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. *Nat. Genet.* 51, 224–229. doi: 10.1038/s41588-018-0282-x

Watson, J. M., and Riha, K. (2010). Comparative biology of telomeres: where plants stand. *FEBS Lett.* 584, 3752–3759. doi: 10.1016/J.FEBSLET.2010.06.017

Weiss-Schneeweiss, H., and Schneeweiss, G. M. (2013). "Karyotype diversity and evolutionary trends in angiosperms," in *Plant Genome Diversity*, Springer-Verlag Wien. 209–230. doi: 10.1007/978-3-7091-1160-4_13

Wellenreuther, M., and Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* 33 (6), 427–440. doi: 10.1016/j.tree.2018.04.002

Wendel, J. F., and Cronn, R. C. (2003). Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78, 139–186. doi: 10.1016/S0065-2113(02)78004-8

Wendel, J. F., Cronn, R. C., Spencer Johnston, J., and James Price, H. (2002). Feast and famine in plant genomes. *Genetica* 115, 37–47. doi: 10.1023/a:1016020030189

White, M. (1973). *Animal cytology and evolution*. London: Cambridge University Press.

Zhang, P., Li, W., Friebe, B., and Gill, B. S. (2008). The origin of a "Zebra" chromosome in wheat suggests nonhomologous recombination as a novel mechanism for new chromosome evolution and step changes in chromosome number. *Genetics* 179 (3), 1169–1177. doi: 10.1534/genetics.108.089599

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33, 531–537. doi: 10.1038/nbt.3207

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 January 2020
doi: 10.3389/fpls.2019.01799

Mitotic Spindle Attachment to the Holocentric Chromosomes of *Cuscuta europaea* Does Not Correlate With the Distribution of CENH3 Chromatin

Ludmila Oliveira[1†], *Pavel Neumann*[1†], *Tae-Soo Jang*[1‡], *Sonja Klemme*[1], *Veit Schubert*[2], *Andrea Koblížková*[1], *Andreas Houben*[2] *and Jiří Macas*[1*]

[1] Biology Centre, Czech Academy of Sciences, Institute of Plant Molecular Biology, České Budějovice, Czechia, [2] Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

The centromere is the region on a chromosome where the kinetochore assembles and spindle microtubules attach during mitosis and meiosis. In the vast majority of eukaryotes, the centromere position is determined epigenetically by the presence of the centromere-specific histone H3 variant CENH3. In species with monocentric chromosomes, CENH3 is confined to a single chromosomal region corresponding to the primary constriction on metaphase chromosomes. By contrast, in holocentrics, CENH3 (and thus centromere activity) is distributed along the entire chromosome length. Here, we report a unique pattern of CENH3 distribution in the holocentric plant *Cuscuta europaea*. This species expressed two major variants of CENH3, both of which were deposited into one to three discrete regions per chromosome, whereas the rest of the chromatin appeared to be devoid of CENH3. The two CENH3 variants fully co-localized, and their immunodetection signals overlapped with the positions of DAPI-positive heterochromatic bands containing the highly amplified satellite repeat CUS-TR24. This CENH3 distribution pattern contrasted with the distribution of the mitotic spindle microtubules, which attached at uniform density along the entire chromosome length. This distribution of spindle attachment sites proves the holocentric nature of *C. europaea* chromosomes and also suggests that, in this species, CENH3 either lost its function or acts in parallel to an additional CENH3-free mechanism of kinetochore positioning.

Keywords: centromere, kinetochore, CENH3, holocentric chromosomes, repetitive DNA analysis, satellite DNA, *Cuscuta*

Abbreviations: FISH, fluorescence *in situ* hybridization; RT-PCR, reverse transcription - polymerase chain reaction; 3′RACE, rapid amplification of 3′ cDNA ends; SIM, structured illumination microscopy.

OPEN ACCESS

Edited by:
Hanna Weiss-Schneeweiss, University of Vienna, Austria

Reviewed by:
James A. Birchler, University of Missouri, United States
Steven Henikoff, Fred Hutchinson Cancer Research Center, United States

***Correspondence:**
Jiří Macas
macas@umbr.cas.cz

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

[‡]***Present address:***
Tae-Soo Jang, Department of Biological Sciences, Chungnam National University, Daejeon, South Korea

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

Received: *13 November 2019*
Accepted: *23 December 2019*
Published: *24 January 2020*

Citation:
Oliveira L, Neumann P, Jang T-S, Klemme S, Schubert V, Koblížková A, Houben A and Macas J (2020) Mitotic Spindle Attachment to the Holocentric Chromosomes of Cuscuta europaea Does Not Correlate With the Distribution of CENH3 Chromatin. Front. Plant Sci. 10:1799. doi: 10.3389/fpls.2019.01799

INTRODUCTION

Centromeres are chromosomal regions that facilitate faithful chromosome segregation during cell division. This is achieved by providing an anchor point for assembly of a kinetochore, a protein complex that connects centromeric chromatin to the spindle microtubules (Cheeseman, 2014). Most higher plant species possess monocentric chromosomes, in which the mitotic spindle binds to a single region on each chromosome that is discernible as the primary constriction on condensed metaphase chromosomes. On the other hand, holocentric chromosomes lack primary constrictions and have spindle binding sites distributed along almost the entire chromosome length. Holocentric taxa have a broad phylogenetic distribution, including various groups of nematodes, arthropods, and plants (Melters et al., 2012). In flowering plants, they represent a minor fraction, including, for example, families Juncaceae (Bozek et al., 2012), Cyperaceae (Luceño et al., 1998; Roalson et al., 2007; Håkansson, 2010), Droseraceae (Kolodin et al., 2018); genus *Chionographis* [Liliaceae; (Tanaka and Tanaka, 1979)]; and some species from the genus *Cuscuta* (Pazy and Plitmann, 1994; Pazy and Plitmann, 1995; Pazy and Plitmann, 2002). Because holocentric taxa are often embedded within broader phylogenetic lineages possessing monocentric chromosomes, it is thought that holocentric chromosome organization originated from the monocentric format and that this transition occurred independently in multiple phylogenetic lineages (Melters et al., 2012). However, the factors that induced this transition and its mechanisms are currently unknown.

In most organisms, centromeres are determined epigenetically by the presence of the centromere-specific histone variant CENH3 (Allshire and Karpen, 2008). In monocentric chromosomes, CENH3 is confined to the primary constrictions, whereas in holocentrics it is distributed along the chromosomes concurrently with spindle attachment sites. In exceptional cases, CENH3 genes have been lost altogether in some holocentric organisms (Drinnenberg and Akiyoshi, 2017). These include four lineages of insects in which the loss of CENH3 correlated with the transition to holocentricity, suggesting a causal relationship between the two events (Drinnenberg et al., 2014). In plants, only a few holocentric species have been studied in detail, including representatives of families Juncaceae (*Luzula*) and Cyperaceae (*Rhynchospora*). Chromosomes in these species are characterized by a longitudinal centromere groove that correlates with the presence of CENH3 and attachment of spindle microtubules (Nagaki et al., 2005; Heckmann et al., 2011; Marques et al., 2015; Wanner et al., 2015). In *Rhynchospora pubera*, specific repeats are associated with CENH3-containing chromatin, containing the satellite DNA family TYBA; two different TYBA-containing repeats, TCR1 and TCR2; and the LTR-retrotransposon CRRh (Marques et al., 2015; Ribeiro et al., 2017). By contrast, no centromere-specific repeats have been identified in *Luzula elegans* (Heckmann et al., 2013).

Investigation of the changes associated with the transition from monocentric to holocentric chromosome organization is, in theory, most informative when comparing phylogenetically closely related species that differ in centromere type. The best-documented such case is genus *Cuscuta*, in which holocentric chromosomes have been identified in particular species using classical cytogenetics techniques (Pazy and Plitmann, 1991; Pazy and Plitmann, 1994; Pazy and Plitmann, 1995; Pazy and Plitmann, 2002). Evidence for the existence of holocentric chromosomes includes 1) absence of primary constrictions, 2) orientation of chromosomes parallel to the equatorial plane during mitotic metaphase and anaphase, 3) tolerance of chromosomes to fragmentation, and 4) inverted meiosis. The genus includes about 200 species of obligatory parasitic plants that depend on their host plants for water and nutrients (García et al., 2014). Three distinct phylogenetic lineages exist within the genus, corresponding to subgenera *Monogynella*, *Grammica*, and *Cuscuta*. The first two subgenera include monocentric species, whereas subgenus *Cuscuta* presumably consists entirely of holocentrics (Pazy and Plitmann, 1994; Pazy and Plitmann, 1995; Pazy and Plitmann, 2002; McNeal et al., 2007; García et al., 2014). However, these species have never been studied in detail using molecular techniques.

In this work, we performed a detailed molecular-cytogenetic characterization of *Cuscuta europaea* as a representative holocentric *Cuscuta* species. To investigate centromeric chromatin, we identified *CENH3* genes in *C. europaea*, as well as in two representatives of monocentric species, and investigated the distribution of CENH3 chromatin and microtubule attachment sites along mitotic chromosomes. In addition, we characterized repetitive DNA sequences and mapped their distribution along chromosomes with regard to the patterns of CENH3 distribution.

MATERIALS AND METHODS

Plant Materials

Seeds of *C. europaea* (serial number: 0101147) were obtained from the Royal Botanic Garden (Ardingly, UK). Seeds of *C. campestris* and *C. japonica* were provided by Dr. Chnar Fathoulla (University of Salahaddin, Kurdistan Region, Iraq) and Dr. Takeshi Furuhashi (RIKEN Center for Sustainable Resource Science, Yokohama, Japan), respectively. To ensure rapid germination, seeds of *C. europaea* were treated in sulfuric acid for 1 h at room temperature, washed three times in distilled water, sterilized in 0.1× SAVO (UNILEVER, Prague, Czech Republic; 1× SAVO contains 4.7% NaOCl) for 20 min, and washed three times in sterile distilled water. Finally, the sterile seeds were germinated on a solid 0.5× Murashige and Skoog medium (Duchefa, Haarlem, the Netherlands) containing 0.8% agar supplemented with 3% sucrose. Seeds of the other two species were abraded with a sandpaper or scalpel and germinated on a damp paper towel in a Petri dish. *Cuscuta* seedlings that were 1.5–2 cm long were transferred onto their host plants: *Urtica dioica* (*C. europaea*), *Ocimum basilicum* (*C. campestris*), and *Pelargonium zonale* (*C. japonica*). All *Cuscuta* species were grown in isolation to prevent their accidental spread.

RNA Isolation, Sequencing, and Identification of CENH3-Coding Sequences

Total RNA was isolated from shoots and inflorescences of *C. europaea* and *C. campestris* using Trizol reagent (Invitrogen, Carlsbad, CA, USA) and treated with DNase I (Ambion, Austin, TX, USA). Before sequencing, RNA from the two tissues was mixed in a 1:1 ratio and subjected to two rounds of poly-A mRNA subtraction using the Dynabeads mRNA purification kit (Thermo Fisher Scientific, Waltham, MA, USA). RNA sequencing was performed at GATC Biotech AG (Konstanz, Germany) using Illumina technology to produce 50 nt paired-end reads. The data were deposited in the Short Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) under accessions ERR3651372 and ERR3651373. Illumina RNA-seq data from *C. japonica* were downloaded from SRA (run accessions DRR021689 and DRR021687). *De novo* transcriptome assemblies for all three *Cuscuta* species were built using Trinity (Grabherr et al., 2011). Contig sequences with similarity to CENH3 were identified using tBLASTn (Altschul, 1997), with query containing a set of CENH3 sequences published previously or downloaded from GenBank. Because the histone fold domain (HFD) at the C-terminus of CENH3 shares relatively high similarity with canonical H3 histones, many of the tBLASTn hits belonged to the latter histone variant. However, CENH3 and H3 could be clearly distinguished because the entire H3 histone sequence is nearly invariant across all eukaryotes, whereas CENH3 histones vary among species and differ from classical H3 histones at the N-terminus. The existence of multiple CENH3 variants in *C. europaea* ($CENH3^{CEURO\text{-}1a}$, $CENH3^{CEURO\text{-}1b}$, and $CENH3^{CEURO\text{-}2}$) and *C. campestris* ($CENH3^{CCAMP\text{-}a}$ and $CENH3^{CCAMP\text{-}b}$) was confirmed experimentally by RT-PCR and 3'RACE amplification, followed by sequencing of the cloned products. RT-PCR and 3'RACE were performed as described previously (Neumann et al., 2015). First-strand synthesis was performed using the SuperScript III First-Strand Synthesis System for the RT-PCR kit (Thermo Fisher Scientific). Sequences of primers used for reverse transcription and PCR, along with details of the amplification conditions, are provided in **Supplementary Table 1**. Sequences of selected clones representing all CENH3 variants in the two *Cuscuta* species were deposited into GenBank under accession numbers MN625517-MN625524. The CENH3 sequence from *C. japonica* was not verified experimentally because the same sequence was assembled using two different RNA-seq datasets, and the RNA-seq reads did not indicate the presence of multiple variants.

Analysis of Repetitive DNA Sequences

Genomic DNA used for sequencing was extracted from young shoots of *C. europaea* as described previously (Dellaporta et al., 1983). Shotgun sequencing of the DNA was performed by University of Rochester Genomics Research Center (New York, NY, USA), employing an Illumina platform to generate 100 nt paired-end reads from ~300–500 bp fragment libraries. The sequence data was deposited in SRA under run accession ERR3528104. Repetitive sequences were identified by similarity-based clustering of Illumina paired-end reads using the RepeatExplorer (Novák et al., 2013) and TAREAN (Novák et al., 2017) pipelines. The numbers of analyzed reads were 2,565,388 and 254,708, corresponding to 0.22× and 0.02× genome coverage, respectively. All clusters representing at least 0.01% of the genome were manually checked, and their automated annotation was corrected if needed; finally, the clusters were used to characterize and quantify the most abundant repeats. Genome proportions of the major repeat types (**Supplementary Table 2**) were calculated based on the proportion of reads in individual annotated clusters. Abundances of simple sequence repeats, such as (TAA)n, were calculated using Tandem Repeats Finder (TRF) (Benson, 1999) and TRAP (Sobreira et al., 2006). The input for TRF was prepared by concatenating one million randomly selected reads, each of which was separated by a stretch of 50 Ns.

In Situ Immunodetection and FISH

Affinity-purified rabbit polyclonal antibodies against peptides corresponding to N-terminal sequences of individual CENH3 variants were custom-produced by GenScript (Piscataway, NJ, USA; antibodies against $CENH3^{CEURO\text{-}1a}$, $CENH3^{CEURO\text{-}2}$, and $CENH3^{CCAMP\text{-}a/b}$) or Biomatik (Cambridge, Canada; antibody against $CENH3^{CJAPO}$). Peptide sequences are highlighted in **Figure 1**. Because the $CENH3^{CEURO\text{-}1a}$ and $CENH3^{CEURO\text{-}1b}$ variants differ at only two of 28 amino acid residues of the peptide sequence, the antibody raised primarily to $CENH3^{CEURO\text{-}1a}$ is likely to have recognized both $CENH3^{CEURO\text{-}1}$ variants. Mouse monoclonal antibody to α-tubulin was purchased from Sigma Aldrich (St. Louis, MO, USA; catalog number: T6199).

In situ immunodetection of CENH3 and α-tubulin was done as described by (Neumann et al., 2012; Neumann et al., 2015) with the following modifications. For detection of CENH3, chromosome preparations were made in LB01 lysis buffer (15 mM Tris, 2 mM Na_2EDTA, 80 mM KCl, 20 mM NaCl, 0.5 mM spermine, 15 mM mercaptoethanol, 0.1% Triton X-100, pH 7.5) by squashing shoot apical meristems fixed in TRIS-fix buffer (4% formaldehyde, 10 mM Tris, 10 mM Na_2EDTA, 100 mM NaCl, pH 7.5) for 30 min at 10°C (infiltration of the fixative was improved by application of vacuum for the first 5 min), followed by digestion with 2% cellulase ONOZUKA R10 (SERVA Electrophoresis, Heidelberg, Germany) and 2% pectinase (MP Biomedicals, Santa Ana, CA, USA) in PBS buffer for 110 min at 27.4°C. For simultaneous detection of CENH3 and α-tubulin, suspension of chromosomes and nuclei isolated from fixed shoot apical meristems were prepared as described (Neumann et al., 2002), and then were spun on slides using a Hettich centrifuge equipped with cytospin chambers. Rabbit and mouse primary antibodies were detected using goat anti-rabbit Rhodamine Red-X (1:500 dilution; Jackson ImmunoResearch, Suffolk, UK; catalog number: 111-295-144) and goat anti-mouse Alexa Fluor 488 (1:500 dilution; Jackson ImmunoResearch; catalog number: 115-545-166), respectively. For combined detection of CENH3 and CUS-TR24, shoot apical meristems were fixed in 3:1 ethanol: glacial acetic acid for 30 min at 10°C, digested with 2% cellulase and 2% pectinase in PBS for 110 min at 27.4°C, and finally

squashed in 45% acetic acid. Immunodetection and FISH were performed in consecutive steps as described (Neumann et al., 2012).

Mitotic chromosomes for FISH experiments were prepared from shoot apical meristems synchronized using ice-cold water for 17 h and fixed in a 3:1 solution of methanol:glacial acetic acid for at least 1 day. The fixed meristems were washed three times for 5 min in distilled water (5 min each), and then incubated in a solution of 2% cellulase and 2% pectinase in PBS for 70 min at 37°C. The samples were then washed carefully with cold distilled water, transferred to a glass slide, and macerated in a drop of cold 3:1 fixative solution (ethanol:glacial acetic acid) using fine-pointed forceps. Finally, the slides were warmed over an alcohol flame, air-dried, and stored at 4°C for up to 3 months. The oligonucleotide probes labeled at the 5' end with fluorescein isothiocyanate (FITC; CUS-TR24) or biotin [CUS-TR2, CUS-TR25, and $(TAA)_n$] were purchased from Integrated DNA Technologies (Leuven, Belgium). Fragments of the other repetitive sequences were amplified from genomic DNA of *C. europaea* (CUS-TR65, CUS-TR66, and CUS-TR67) or *Pisum sativum* (5S and 45S rDNA), and were cloned into pCR4-TOPO vector (Thermo Fisher Scientific). Sequences of the clones are provided in **Supplementary Data S1**. These probes were labeled with biotin-16-dUTP (Roche, Mannheim, Germany) using nick translation as described (Karafiátová et al., 2016). FISH was performed according to Macas et al. (2007), with hybridization and washing temperatures adjusted to account for AT/GC content and hybridization stringency allowing for 10%–20% mismatches. Biotin-labeled probes were detected using Streptavidin-Alexa Fluor 488 (Jackson ImmunoResearch) or Streptavidin-Alexa Fluor 568 (Thermo Fisher Scientific). The slides were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) and mounted in Vectashield mounting medium (Vector Laboratories, Burlingame, CA, USA).

Microscopy

Conventional wide-field fluorescence microscopy was performed using a Zeiss AxioImager.Z2 microscope equipped with an AxioCam 506 mono-color camera. The microscope was also equipped with an Apotome2.0 device for better resolution in the z-axis, which was needed when the images were composed of multiple optical sections. Images were generated using the ZEN 2 blue software (Carl Zeiss GmbH). To analyze the distribution of microtubules along the chromosomes at the super-resolution level (~120 nm using a 488-nm laser), spatial structured illumination microscopy (3D-SIM) was performed using a 63×/1.4 Oil Plan-Apochromat objective on an Elyra PS.1 microscope system, controlled by the ZENblack software (Carl Zeiss GmbH). Images were captured using the 405-, 488-, and 561-nm laser lines for excitation and the appropriate emission filters. Three-dimensional movies were produced from SIM image stacks using the Imaris 8.0 (Bitplane) and ZENblack software.

RESULTS

Because our strategy for investigating holocentric chromosomes of *C. europaea* required the visualization of CENH3 protein, which is a universal marker of the centromeric chromatin in plants, we first identified CENH3-coding genes in this species. In parallel, we also investigated representatives of two phylogenetic lineages of monocentric *Cuscuta* species, *C. japonica* (subgenus *Monogynella*) and *C. campestris* (subgenus *Grammica*). Putative CENH3 sequences were identified in short-read transcriptomic data and then verified by cloning complete transcripts obtained by RT-PCR and 3'RACE. In monocentric species, we found a single CENH3 sequence in *C. japonica* ($CENH3^{CJAPO}$) and two sequence variants in *C. campestris* ($CENH3^{CCAMP\text{-}a}$ and $CENH3^{CCAMP\text{-}b}$), which corresponded to their diploid and tetraploid chromosome numbers, respectively. By contrast, the diploid holocentric species *C. europaea* yielded three variants of CENH3 ($CENH3^{CEURO\text{-}1a}$, $CENH3^{CEURO\text{-}1b}$, and $CENH3^{CEURO\text{-}2}$). The CENH3 sequences (**Figure 1**) differed considerably between species, as they did between the $CENH3^{CEURO\text{-}1a/b}$ and $CENH3^{CEURO\text{-}2}$ variants in *C. europaea* (**Table 1**). We then designed polyclonal antibodies for immunodetection of CENH3 on chromosomes, targeting the N-terminal regions, which were the most variable parts of these proteins (**Figure 1**).

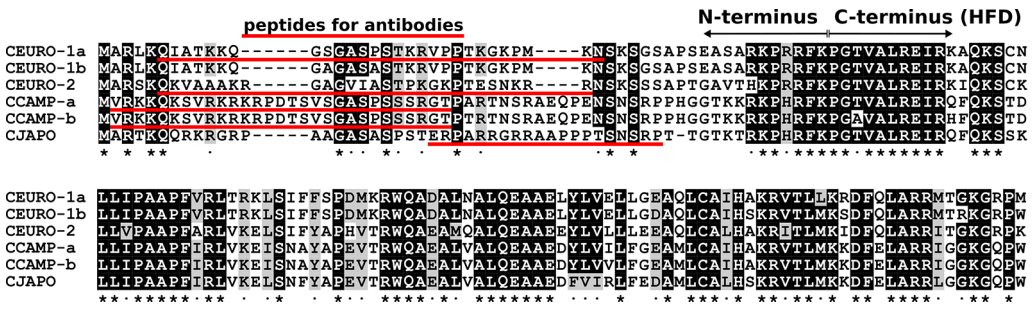

FIGURE 1 | Comparison of CENH3 protein sequences identified in *C. europaea* (CEURO), *C. campestris* (CCAMP), and *C. japonica* (CJAPO). The alignment of CENH3 protein sequences is shaded to highlight invariable sites (black) and sites containing similar amino acid residues (gray). These sites are also marked below the alignment by asterisks and dots, respectively. Note that the sequences are relatively conserved at their C-termini, which contain the histone fold domain (HFD), but are highly divergent at the N-terminus. Sequences of peptides used to produce CENH3-specific antibodies are underlined.

TABLE 1 | Pairwise similarities between CENH3 sequences.

	CEURO-1a	CEURO-1b	CEURO-2	CCAMP-a	CCAMP-b	CJAPO
CEURO-1a	–	94.6	65.5	60.1	60.1	54.4
CEURO-1b	96.2/93.8	–	65.5	58.1	58.1	57.8
CEURO-2	48.1/75.0	51.9/72.9	–	61.7	61.7	62.2
CCAMP-a	42.3/69.8	38.5/68.8	37.7/75.0	–	97.5	75.7
CCAMP-b	44.2/68.8	40.4/67.7	39.6/74.0	98.4/96.9	–	73.7
CJAPO	35.3/64.6	39.2/67.7	40.4/74.0	57.1/86.5	55.4/84.4	–

All values are percentages. The percent similarities between entire CENH3 protein sequences are shown above the diagonal, whereas percent similarities between the N- and C-terminal parts are shown below the diagonal (N/C).

In *in situ* immunodetection, these antibodies revealed unexpected patterns of CENH3 distribution on mitotic metaphase chromosomes of holocentric *C. europaea* (**Figure 2**). Contrary to previously characterized holocentric plant species, in which CENH3 is always distributed in narrow stripes positioned on the lateral sides of the sister chromatids and extending along almost the entire chromosome length, *C. europaea* chromosomes exhibited an uneven labeling, with CENH3 signals concentrated in one to three bands arranged across the chromosome width (**Figures 2A, C**). The number of CENH3 bands did not correlate with chromosome size: the largest chromosome (chromosome 1) of the *C. europaea* karyotype displayed a single subterminal CENH3 band, whereas the smaller chromosomes each had one to three CENH3 bands. The remaining parts of the chromosomes were devoid of detectable CENH3 signal. The immunodetection signals overlapped with DAPI-positive heterochromatic bands, with the exception of one of the two DAPI bands on chromosome 1, which was free of CENH3 signal (**Figures 2A, C, E**). A clear association between CENH3 proteins and heterochromatin was also observed in interphase nuclei (**Figures 2B, D**). The same labeling patterns were obtained using antibodies for both major variants, $CENH3^{CEURO\text{-}1a/b}$ and $CENH3^{CEURO\text{-}2}$. This atypical distribution of CENH3 raised doubts about the protein's role in kinetochore establishment in holocentric *Cuscuta* species. On the other hand, CENH3 function was maintained in the two monocentric *Cuscuta* species, which both displayed the expected patterns of CENH3 localization in the primary constrictions of mitotic chromosomes (**Figures 2F–G**).

Prompted by the atypical distribution of CENH3 on *C. europaea* chromosomes, we next investigated the distribution of mitotic spindle attachment sites along the chromosomes. Should CENH3 act as an epigenetic mark of centromeric chromatin, as it does in other plant species, kinetochore formation (and thus the spindle attachment sites) would be expected to co-localize with the CENH3 bands. However, spatial structured illumination microscopy (3D-SIM, super-resolution) revealed that the spindle microtubules visualized by α-tubulin antibody were evenly attached along the entire chromosome length, regardless of the distribution of the CENH3 signals (**Figure 3** and **Supplementary Movies 1** and **2**). This was clearest for chromosome 1, where CENH3 was concentrated into a single band at one chromosome end, but the spindle microtubules attached along the entire chromosome (**Figures 3B, C**). On the other hand, the observed spindle attachment patterns confirmed the holocentric nature of the *C. europaea* chromosomes, which was also supported by the anaphase arrangement of chromosomes parallel to the equatorial plane of the cell division, which is typical of holocentromeres (**Supplementary Movie 3**). In *C. japonica*, the microtubules attached to chromosomes exclusively at CENH3 chromatin domains (**Figure 3D**).

In light of the association of CENH3 with heterochromatin, we next sought to determine whether the CENH3 distribution correlated with the presence of specific families of repetitive sequences. To identify such sequences in the *C. europaea* genome, we performed low-pass Illumina sequencing of genomic DNA followed by repeat characterization from the Illumina reads using the RepeatExplorer (Novák et al., 2013) and TAREAN (Novák et al., 2017) pipelines. A particular abundance of satellite DNA (satDNA), constituting 18% of the genome, was found beside all other major types of repetitive sequences (**Supplementary Table 2**). Although six satDNA families and one abundant microsatellite were identified (**Table 2**), this high proportion of satellite DNA was mainly due to the amplification of a single family, CUS-TR24, which alone made up 15.5% of the genome. Because satDNA is a typical constituent of heterochromatic bands on plant chromosomes, we localized all identified tandem repeats, along with rRNA genes, to test for their presence in the DAPI-positive bands (**Figure 4**). Fluorescence *in situ* hybridization (FISH) experiments revealed that all but one of these heterochromatic bands contained the CUS-TR24 satellite (**Figure 4A**), along with the (TAA)n repeat (**Figure 4B**). The same hybridization sites were also associated with the CENH3 chromatin (**Figure 5**). An additional minor satellite repeat, CUS-TR25, was located within the CUS-TR24/CENH3 band on chromosome 7 (**Figure 4C**). The remaining prominent DAPI band on chromosome 1 that was free of CENH3 consisted of satellite CUS-TR2 (**Figure 4D**). The other three satellites were detected as minor loci, apart from the heterochromatic bands (**Figures 4E, F** and data not shown), as schematically depicted in **Figure 4H**.

DISCUSSION

In this work, we demonstrated that *C. europaea* chromosomes can be classified as holocentric, based on the attachment of the mitotic spindle microtubules along the entire chromosome

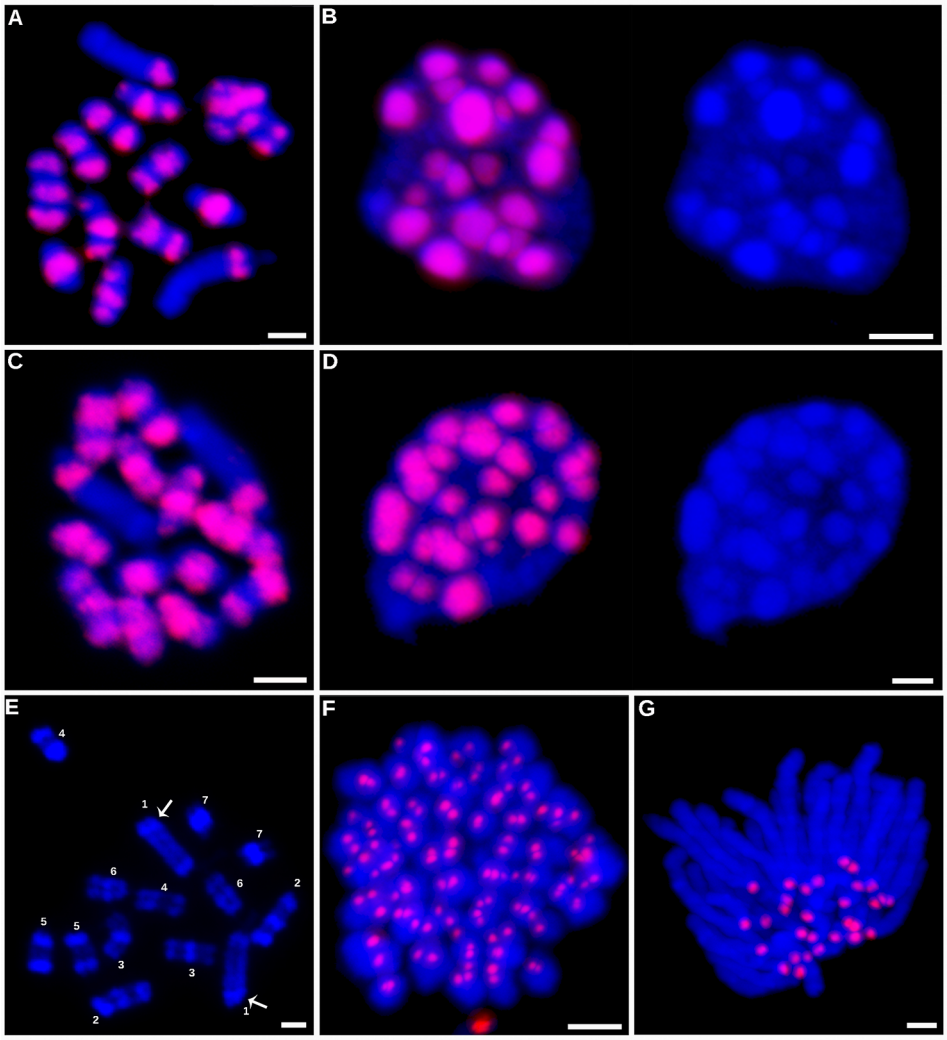

FIGURE 2 | Immunodetection of CENH3 proteins. Immunodetection of CENH3 proteins in *C. europaea* **(A–D).** Detection of the CENH3$^{CEURO-1a/b}$ variant on metaphase chromosomes **(A)** and interphase nucleus **(B)**. Detection of the CENH3$^{CEURO-2}$ variant on metaphase chromosomes **(C)** and interphase nucleus **(D)**. Nuclei in the panels **(B** and **D)** are shown with and without the CENH3 signals to demonstrate correlation between the distribution of CENH3 and DAPI-positive heterochromatin domains. **(E)** Distribution of DAPI-positive heterochromatin domains on *C. europaea* metaphase chromosomes, prepared from 3:1 fixed meristems to achieve better contrast between heterochromatic and euchromatic regions. Distribution of the heterochromatin domains allowed to distinguish all seven chromosome pairs in this species. Arrows mark the heterochromatin band on chromosome 1 that lacked CENH3 signal. **(F)** Detection of CENH3$^{CCAMP-a/b}$ on metaphase chromosomes of *C. campestris*. **(G)** Detection of CENH3^{CJAPO} on anaphase chromosomes of *C. japonica.* CENH3 signals and DAPI-stained chromosomes are shown in red and blue, respectively. Scale bars = 2 µm.

length, the absence of primary constrictions, and the orientation of sister chromatids during cell division. This is in agreement with previous reports on species from the same subgenus *Cuscuta*, which were deemed holocentric based on their chromosome morphology and behavior during mitosis and meiosis (Pazy and Plitmann, 1991; Pazy and Plitmann, 1994; Pazy and Plitmann, 1995). However, the distribution of CENH3 proteins on *C. europaea* chromosomes did not correspond to the patterns found in other holocentrics in either the plant or animal kingdoms. Contrary to the known role of CENH3 as a foundational kinetochore protein in most eukaryotes (Westermann and Schleiffer, 2013; Cheeseman, 2014; Hara and Fukagawa, 2017), its atypical confinement into one to three loci per chromosome did not impact the holocentromere-typical arrangement of the mitotic spindle microtubules. Instead, the microtubules also attached to chromosomes at sites where CENH3 was not detected, and their density was not higher at sites of CENH3 accumulation. In contrast to other holocentric plants of genera *Luzula* and *Rhynchospora*, where the holocentromeres form a longitudinal groove extending over almost the entire sister chromatids (Heckmann et al., 2011; Marques et al., 2015; Wanner et al., 2015), the chromosome-spindle interface in *C. europaea* mainly had a smooth surface (**Supplementary Movie 4**).

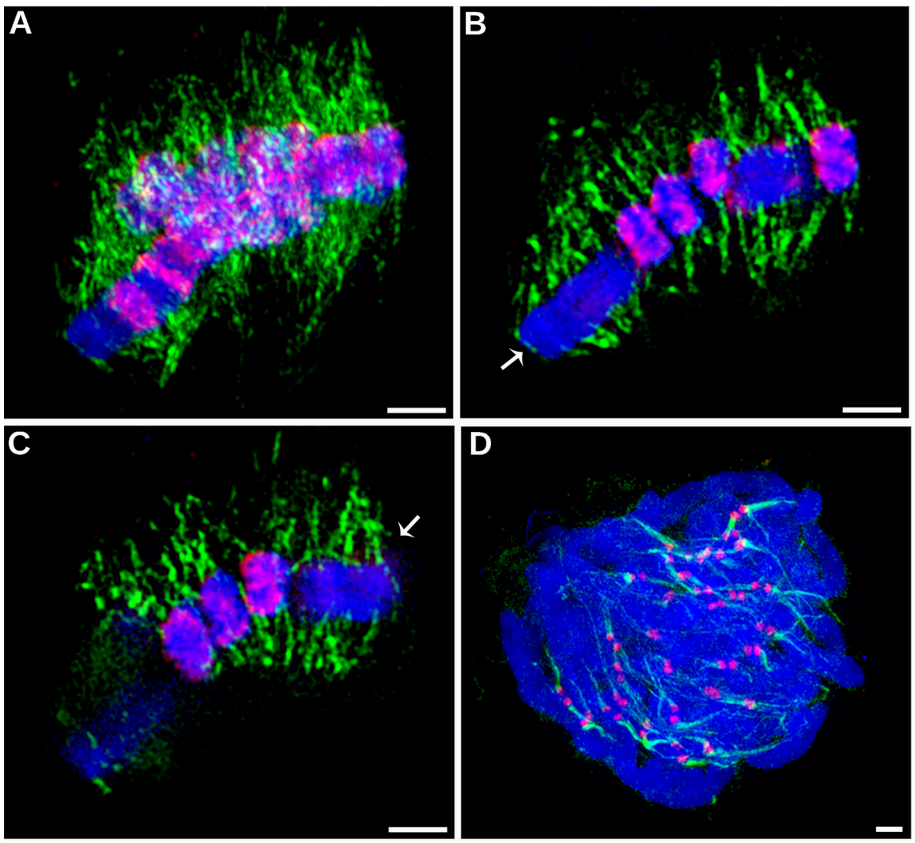

FIGURE 3 | Distribution of CENH3 and spindle attachment sites on chromosomes, investigated by super-resolution structured illumination microscopy (3D-SIM). **(A–C)** Detection of CENH3 (red) and tubulin (green) on metaphase chromosomes of *C. europaea*. The CENH3 was detected using mixed antibodies to $CENH3^{CEURO\text{-}1a/b}$ and $CENH3^{CEURO\text{-}2}$ (red). Microtubules were detected using α-tubulin antibody (green). **(A)** Maximum-intensity projection image reconstructed from 3D SIM image stack. **(B–C)** Two optical sections selected from the same 3D SIM image stack. They show that microtubules of the mitotic spindle are evenly attached to chromosomes at their poleward sides and along their entire lengths, regardless of the occurrence of CENH3 signals. Arrows mark chromosome 1 which exhibits large CENH3-free region. The full set of optical sections and corresponding projections are also available as **Supplementary Movies 1** and **2**. **(D)** Detection of CENH3 (red) and α-tubulin (green) on metaphase chromosomes of *C. japonica*. Note that microtubules attach to the chromosomes exclusively at CENH3-containing domains. Chromosomes were stained with DAPI (blue). Scale bars = 2 μm.

TABLE 2 | Characteristics of satellite DNA families identified in *C. europaea*.

	Genome proportion [%][a]	TAREAN confidence	monomer length [bp]	proportion of AT [%]
CUS-TR24	15.51	low	389	66.2
CUS-TR2	1.66	high	170	64.8
(TAA)n	0.68[a]	–	3	100.0
CUS-TR65	0.04	high	1714	69.1
CUS-TR25	0.02	high	173	74.5
CUS-TR66	0.02	high	1047	70.6
CUS-TR67	0.01	low	322	71.8

[a]*Genome proportions were estimated from the proportion of reads in the respective RepeatExplorer cluster relative to the total number of reads. The exception was the $(TAA)_n$ microsatellite, which was calculated using TRF and TRAP.*

Assuming that the mitotic spindle binds to chromosomes exclusively at sites where kinetochores are formed (Cheeseman, 2014), our observations suggest that none of the CENH3 variants is an integral component of mitotic kinetochores in *C. europaea*. This contrasts with the notion that CENH3 is essential for kinetochore formation and function (Hara and Fukagawa, 2017). Rare exceptions to this notion include holocentric insect species that lack the *CENH3* genes, implying that they evolved a CENH3-independent mechanism of kinetochore assembly (Drinnenberg et al., 2014; Mon et al., 2017). We speculate that this might also have taken place in *C. europaea*, except that, in this case, the CENH3 genes were preserved and continue to be expressed. An alternative explanation for the discrepancy between the arrangement of microtubules and the distribution of CENH3 is that, in *C. europaea*, these proteins are actually not restricted to sites where they were detected, but also occur in small domains below the limits of microscopy, which are scattered along the entire chromosome length. Theoretically,

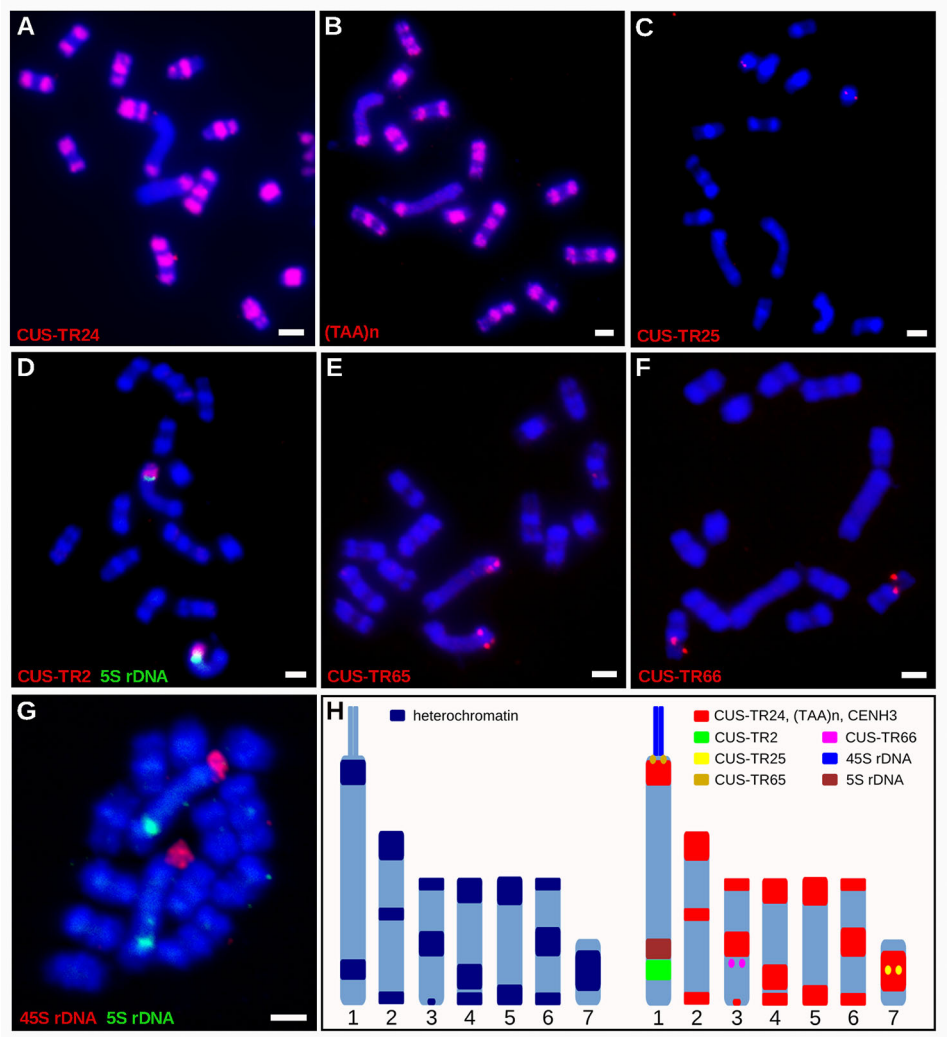

FIGURE 4 | Distribution of satDNA families on chromosomes of *C. europaea*, investigated by FISH. **(A)** CUS-TR24, **(B)** (TAA)n, **(C)** CUS-TR25, **(D)** CUS-TR2 and 5S rDNA, **(E)** CUS-TR65, **(F)** CUS-TR66, **(G)** 45S and 5S rDNAs. Note that 45S rDNA localized to a chromatin tail protruding from one terminus of the largest chromosome 1. The presence of the tail helped to distinguish the CENH3-containing (proximal to the tail) and CENH3-lacking (distal from the tail) heterochromatin domains on this chromosome. Chromosome morphology, size, and repeat distribution patterns allowed us to distinguish all seven chromosome types in *C. europaea*. **(H)** Ideograms summarizing distribution of heterochromatin domains (left), and satDNA families and rDNA repeats (right). CUS-TR67 is not included because it was difficult to detect and its distribution could not be precisely determined. Scale bars = 2 μm.

the kinetochore can assemble on a centromere unit as small as a single CENH3-containing nucleosome, as is the case for point centromeres in *Saccharomyces cerevisiae* (Westermann et al., 2007); therefore, the presence of small CENH3-dependent kinetochores cannot be excluded. However, on the assumption that the stoichiometry between the CENH3, other kinetochore proteins, and attached microtubules is preserved along the chromosome, this hypothesis would predict the higher density of microtubules in CENH3-enriched than in CENH3-depleted regions, which is not in accordance with our observations. Yet another explanation of our results is that CENH3-dependent and CENH3-independent pathways of mitotic kinetochore assembly co-exist in *C. europaea*, the former acting in most heterochromatin domains, and the latter in other parts of the chromosomes. Some of these explanations may be tested by simultaneous localization of other constitutive centromere proteins like CENP-X (Cheeseman, 2014; Hara and Fukagawa, 2017) that may be sufficient to direct kinetochore assembly. Another important question to be addressed in the future experiments is CENH3 distribution and organization of *C. europaea* chromosomes in meiosis, which was not addressed in the present study.

The occurrence of CENH3 exclusively in heterochromatin domains that contain satDNA families (CUS-TR24, CUS-TR25, and $(TAA)_n$) resembles the arrangement in most monocentric species, as well as in the holocentric *Rhynchospora pubera*, in which CENH3 is associated mainly with satDNA (Ugarković, 2009; Marques et al., 2015). Although the causes of this phenomenon

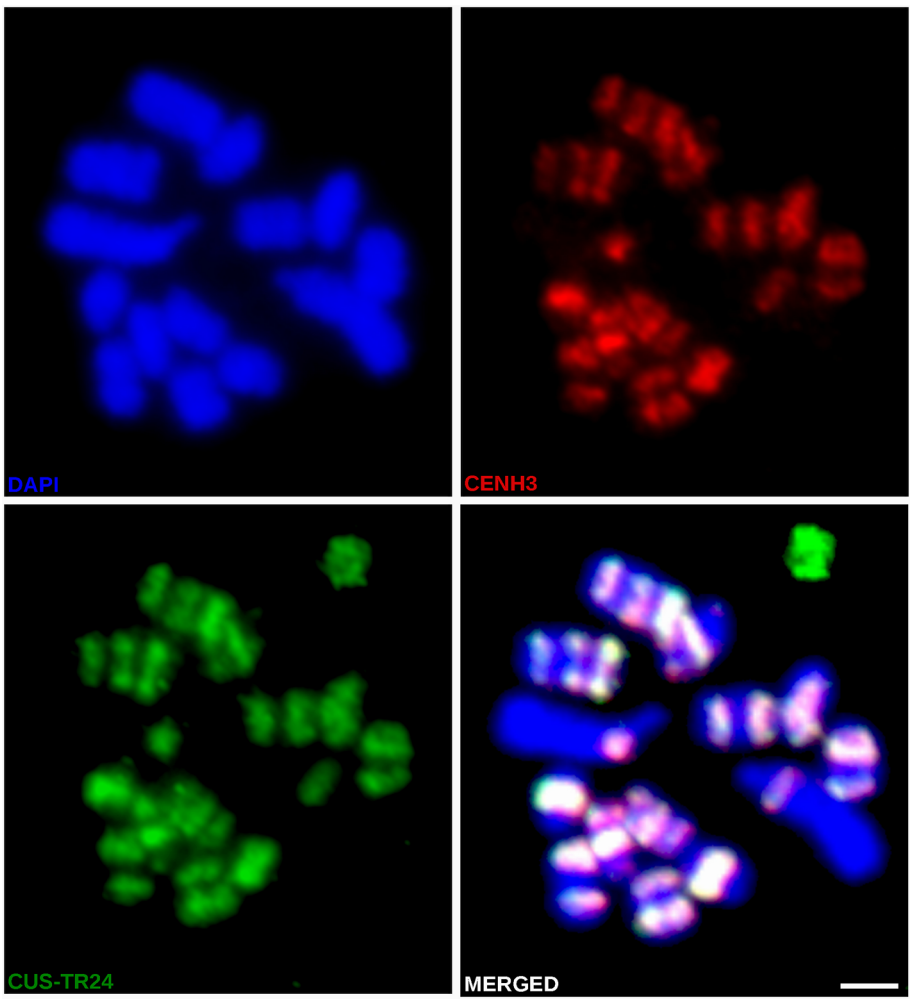

FIGURE 5 | Simultaneous immuno-FISH detection of CENH3 and CUS-TR24 on metaphase chromosomes of *C. europaea*. CENH3 (red) was detected using mixed antibodies against CENH3$^{CEURO-1a/b}$ and CENH3$^{CEURO-2}$. CUS-TR24 and DAPI-stained chromosomes are shown in green and blue, respectively. Scale bar = 2 µm.

are not yet well understood, the colocalization of CENH3 with the satDNA families in *C. europaea* suggests that the factors behind the satDNA amplification in CENH3-containing chromatin still are, or were until recently, also active in this species. It is not yet known, however, which DNA sequences constitute the CENH3-lacking holocentromeres in *C. europaea*, but it is clear that arrays of satDNA are not involved, as none of the satDNA families present in this species exhibited a chromosome-wide distribution reflecting the distribution of spindle attachment sites. Thus, the only holocentric organisms in which the centromere domains form on specific repetitive DNA sequences are the *Rhynchospora* species (Marques et al., 2015; Ribeiro et al., 2017). On the other hand, this is not surprising because centromeres are supposed to be determined epigenetically, and although they are usually found in highly repetitive regions, centromeric repeats are neither necessary nor sufficient for centromere specification (McKinley and Cheeseman, 2016).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Sequence Read Archive: ERR3651372 and ERR3651373, GenBank: MN625517-MN625524.

AUTHOR CONTRIBUTIONS

PN, AH, and JM conceived the study and designed the experiments. LO, T-SJ, and SK performed the cytogenetics experiments and conventional fluorescence microscopy. AK performed RT-PCR and 3'RACE experiments. PN analyzed the sequenced data. VS carried out the super-resolution microscopy. PN and JM wrote the manuscript with input from LO, T-SJ, SK, AK, VS, and AH. All authors read and approved the final manuscript.

FUNDING

This research was financially supported by grants from the Czech Science Foundation (17-09750S) and the Czech Academy of Sciences (RVO:60077344).

ACKNOWLEDGMENTS

Computational resources were provided by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01799/full#supplementary-material

SUPPLEMENTARY TABLE 1 | Primers used for RT-PCR and RACE reactions.

SUPPLEMENTARY TABLE 2 | Proportions of major types of repetitive DNA in the genome of *C. europaea*.

SUPPLEMENTARY DATA S1 | Sequences of the FISH probes in FASTA format.

SUPPLEMENTARY MOVIE 1 | Three-dimensional organization of CENH3 and tubulin on metaphase chromosomes of *C. europaea*. A two-dimensional image of these chromosomes is shown in **Figure 3**. CENH3 and tubulin are shown in red and green, respectively.

SUPPLEMENTARY MOVIE 2 | Organization of CENH3 and tubulin on metaphase chromosomes of *C. europaea*. The movie plays through optical sections that were acquired for the metaphase shown in **Figure 3** and **Supplementary Movie 1**. CENH3 and tubulin are shown in red and green, respectively.

SUPPLEMENTARY MOVIE 3 | Orientation of *C. europaea* chromosomes during anaphase. The chromosomes are oriented in parallel to the plane of mitotic division, indicating that the pulling force of the mitotic spindle is evenly distributed along the chromosomes.

SUPPLEMENTARY MOVIE 4 | Surface rendering of the DAPI-stained *C. europaea* metaphase chromosomes shown in **Supplementary Movie 1** indicates the absence of distinct structures associated with CENH3-positive chromatin and the microtubule attachment sites. The chromosomes surfaces were produced *via* the surface rendering tool of the Imaris 8.0 software.

REFERENCES

Allshire, R. C., and Karpen, G. H. (2008). Epigenetic regulation of centromeric chromatin: Old dogs, new tricks? *Nat. Rev. Genet.* 9, 923–937. doi: 10.1038/nrg2466

Altschul, S. (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

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Bozek, M., Leitch, A. R., Leitch, I. J., Záveská Drábková, L., and Kuta, E. (2012). Chromosome and genome size variation in *Luzula* (Juncaceae), a genus with holocentric chromosomes. *Bot. J. Linn. Soc* 170, 529–541. doi: 10.1111/j.1095-8339.2012.01314.x

Cheeseman, I. (2014). The Kinetochore. *Cold Spring Harb. Perspect. Biol.* 6, 1–18. doi: 10.1101/cshperspect.a015826

Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983). A plant DNA minipreparation: Version II. *Plant Mol. Biol. Rep.* 1, 19–21. doi: 10.1007/BF02712670

Drinnenberg, I. A., and Akiyoshi, B. (2017). "Evolutionary Lessons from Species with Unique Kinetochores," in *Centromeres and Kinetochores*, ed. B. Black (Cham: Springer), 111–138. doi: 10.1007/978-3-319-58592-5_5

Drinnenberg, I. A., DeYoung, D., Henikoff, S., and Malik, H. S. (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife* 3, e03676. doi: 10.7554/eLife.03676

García, M. A., Costea, M., Kuzmina, M., and Stefanović, S. (2014). Phylogeny, character evolution, and biogeography of *Cuscuta* (Dodders; Convolvulaceae) inferred from coding plastid and nuclear sequences. *Am. J. Bot.* 101, 670–690. doi: 10.3732/ajb.1300449

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

Håkansson, A. (2010). Holocentric chromosomes in *Eleocharis*. *Hereditas* 44, 531–540. doi: 10.1111/j.1601-5223.1958.tb03498.x

Hara, M., and Fukagawa, T. (2017). "Critical foundation of the Kinetochore: the constitutive centromere-associated network (CCAN)," in *Centromeres and Kinetochores*, ed. B. Black (Cham: Springer), 29–57. doi: 10.1007/978-3-319-58592-5_2

Heckmann, S., Schroeder-Reiter, E., Kumke, K., Ma, L., Nagaki, K., Murata, M., et al. (2011). Holocentric chromosomes of *Luzula elegans* are characterized by a longitudinal centromere groove, chromosome bending, and a terminal nucleolus organizer region. *Cytogenet. Genome Res.* 134, 220–228. doi: 10.1159/000327713

Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., et al. (2013). The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. *Plant J.* 73, 555–565. doi: 10.1111/tpj.12054

Karafiátová, M., Bartoš, J., and Doležel, J. (2016). "Localization of Low-Copy DNA Sequences on Mitotic Chromosomes by FISH," in Plant Cytogenetics eds. S. Kianian and P. Kianian (New York, NY: Humana Press), 49–64. doi: 10.1007/978-1-4939-3622-9_5

Kolodin, P., Cempírková, H., Bureš, P., Horová, L., Veleba, A., Francová, J., et al. (2018). Holocentric chromosomes may be an apomorphy of Droseraceae. *Plant Syst. Evol.* 304, 1289–1296. doi: 10.1007/s00606-018-1546-8

Luceño, M., Vanzela, A. L. L., and Guerra, M. (1998). Cytotaxonomic studies in Brazilian *Rhynchospora* (Cyperaceae), a genus exhibiting holocentric chromosomes. *Can. J. Bot.* 76, 440–449. doi: 10.1139/b98-013

Macas, J., Neumann, P., and Navrátilová, A. (2007). Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 8, 427. doi: 10.1186/1471-2164-8-427

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., et al. (2015). Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. *Proc. Natl. Acad. Sci. U. S. A.* 112, 13633–13638. doi: 10.1073/pnas.1512255112

McKinley, K. L., and Cheeseman, I. M. (2016). The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* 17, 16–29. doi: 10.1038/nrm.2015.5

McNeal, J. R., Arumugunathan, K., Kuehl, J. V., Boore, J. L., and Depamphilis, C. W. (2007). Systematics and plastid genome evolution of the cryptically photosynthetic parasitic plant genus *Cuscuta* (Convolvulaceae). *BMC Biol.* 5, 55. doi: 10.1186/1741-7007-5-55

Melters, D. P., Paliulis, L. V., Korf, I. F., and Chan, S. W. L. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosom. Res.* 20, 579–593. doi: 10.1007/s10577-012-9292-1

Mon, H., Lee, J. M., Sato, M., and Kusakabe, T. (2017). Identification and functional analysis of outer kinetochore genes in the holocentric insect *Bombyx mori*. *Insect Biochem. Mol. Biol.* 86, 1–8. doi: 10.1016/j.ibmb.2017.04.005

Nagaki, K., Kashihara, K., and Murata, M. (2005). Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant *Luzula nivea*. *Plant Cell* 17, 1886–1893. doi: 10.1105/tpc.105.032961

Neumann, P., Požárková, D., Vrána, J., Doležel, J., and Macas, J. (2002). Chromosome sorting and PCR-based physical mapping in pea (*Pisum sativum* L.). *Chromosom. Res.* 10, 63–71. doi: 10.1023/A:1014274328269

Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., et al. (2012). Stretching the rules: Monocentric chromosomes with multiple centromere domains. *PloS Genet.* 8, e1002777. doi: 10.1371/journal.pgen.1002777

Neumann, P., Pavlíková, Z., Koblížková, A., Fuková, I., Jedličková, V., Novák, P., et al. (2015). Centromeres off the hook: massive changes in centromere size and structure following duplication of CenH3 gene in *Fabeae* species. *Mol. Biol. Evol.* 32, 1862–1879. doi: 10.1093/molbev/msv070

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793. doi: 10.1093/bioinformatics/btt054

Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., and Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* 45, e111. doi: 10.1093/nar/gkx257

Pazy, B., and Plitmann, U. (1991). Unusual chromosome separation in meiosis of *Cuscuta* L. *Genome* 34, 533–536. doi: 10.1139/g91-082

Pazy, B., and Plitmann, U. (1994). Holocentric chromosome behaviour in *Cuscuta* (Cuscutaceae). *Plant Syst. Evol.* 191, 105–109. doi: 10.1007/BF00985345

Pazy, B., and Plitmann, U. (1995). Chromosome divergence in the genus *Cuscuta* and its systematic implications. *Caryologia* 48, 173–180. doi: 10.1080/00087114.1995.10797327

Pazy, B., and Plitmann, U. (2002). New perspectives on the mechanisms of chromosome evolution in parasitic flowering plants. *Bot. J. Linn. Soc* 138, 117–122. doi: 10.1046/j.1095-8339.2002.00006.x

Ribeiro, T., Marques, A., Novák, P., Schubert, V., Vanzela, A. L. L., Macas, J., et al. (2017). Centromeric and non-centromeric satellite DNA organisation differs in holocentric *Rhynchospora* species. *Chromosoma* 126, 325–335. doi: 10.1007/s00412-016-0616-3

Roalson, E., McCubbin, A., and Whitkus, R. (2007). Chromosome Evolution in Cyperales. *Aliso* 23, 62–71. doi: 10.5642/aliso.20072301.08

Sobreira, T. J. P., Durham, A. M., and Gruber, A. (2006). TRAP: automated classification, quantification and annotation of tandemly repeated sequences. *Bioinformatics* 22, 361–362. doi: 10.1093/bioinformatics/bti809

Tanaka, N., and Tanaka, N. (1979). Chromosome studies in *Chionographis* (Liliaceae). *Cytologia (Tokyo).* 44, 935–949. doi: 10.1508/cytologia.44.935

Ugarković, Đ. (2009). "Centromere-competent DNA: structure and evolution," in *Centromere*, ed. Ugarković, Đ. (Berlin, Heidelberg: Springer), 53–76. doi: 10.1007/978-3-642-00182-6_3

Wanner, G., Schroeder-Reiter, E., Ma, W., Houben, A., and Schubert, V. (2015). The ultrastructure of mono- and holocentric plant centromeres: an immunological investigation by structured illumination microscopy and scanning electron microscopy. *Chromosoma* 124, 503–517. doi: 10.1007/s00412-015-0521-1

Westermann, S., and Schleiffer, A. (2013). Family matters: Structural and functional conservation of centromere-associated proteins from yeast to humans. *Trends Cell Biol.* 23, 260–269. doi: 10.1016/j.tcb.2013.01.010

Westermann, S., Drubin, D. G., and Barnes, G. (2007). Structures and functions of yeast kinetochore complexes. *Annu. Rev. Biochem.* 76, 563–591. doi: 10.1146/annurev.biochem.76.052705.160607

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 February 2020
doi: 10.3389/fpls.2019.01789

Genome Synteny Has Been Conserved Among the Octoploid Progenitors of Cultivated Strawberry Over Millions of Years of Evolution

Michael A. Hardigan[1], *Mitchell J. Feldmann*[1], *Anne Lorant*[1], *Kevin A. Bird*[2], *Randi Famula*[1], *Charlotte Acharya*[1], *Glenn Cole*[1], *Patrick P. Edger*[2] *and Steven J. Knapp*[1]*

[1] Department of Plant Sciences, University of California, Davis, Davis, CA, United States, [2] Department of Horticulture, Michigan State University, East Lansing, MI, United States

OPEN ACCESS

Edited by:
Martin A. Lysak,
Masaryk University,
Czechia

Reviewed by:
Martin Mascher,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK),
Germany
Yongpeng Ma,
Chinese Academy of Sciences,
China

***Correspondence:**
Steven J. Knapp
sjknapp@ucdavis.edu

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

Received: 17 October 2019
Accepted: 20 December 2019
Published: 07 February 2020

Citation:
Hardigan MA, Feldmann MJ, Lorant A, Bird KA, Famula R, Acharya C, Cole G, Edger PP and Knapp SJ (2020) Genome Synteny Has Been Conserved Among the Octoploid Progenitors of Cultivated Strawberry Over Millions of Years of Evolution. Front. Plant Sci. 10:1789. doi: 10.3389/fpls.2019.01789

Allo-octoploid cultivated strawberry (*Fragaria* × *ananassa*) originated through a combination of polyploid and homoploid hybridization, domestication of an interspecific hybrid lineage, and continued admixture of wild species over the last 300 years. While genes appear to flow freely between the octoploid progenitors, the genome structures and diversity of the octoploid species remain poorly understood. The complexity and absence of an octoploid genome frustrated early efforts to study chromosome evolution, resolve subgenomic structure, and develop a single coherent linkage group nomenclature. Here, we show that octoploid *Fragaria* species harbor millions of subgenome-specific DNA variants. Their diversity was sufficient to distinguish duplicated (homoeologous and paralogous) DNA sequences and develop 50K and 850K SNP genotyping arrays populated with co-dominant, disomic SNP markers distributed throughout the octoploid genome. Whole-genome shotgun genotyping of an interspecific segregating population yielded 1.9M genetically mapped subgenome variants in 5,521 haploblocks spanning 3,394 cM in *F. chiloensis* subsp. *lucida*, and 1.6M genetically mapped subgenome variants in 3,179 haploblocks spanning 2,017 cM in *F.* × *ananassa*. These studies provide a dense genomic framework of subgenome-specific DNA markers for seamlessly cross-referencing genetic and physical mapping information and unifying existing chromosome nomenclatures. Using comparative genomics, we show that geographically diverse wild octoploids are effectively diploidized, nearly completely collinear, and retain strong macro-synteny with diploid progenitor species. The preservation of genome structure among allo-octoploid taxa is a critical factor in the unique history of garden strawberry, where unimpeded gene flow supported its origin and domestication through repeated cycles of interspecific hybridization.

Keywords: *Fragaria*, strawberry, polyploidy, genome evolution, domestication

INTRODUCTION

Interspecific homoploid hybridization and polyploidy-inducing hybrid events have been creative forces in plant genome evolution and speciation, acting as catalysts for *de novo* reorganization of chromosome structure (Jiao et al., 2011; Yakimowski and Rieseberg, 2014; Soltis et al., 2014b; Soltis et al., 2014a; Vallejo-Marín et al., 2015; McKain et al., 2016; Soltis et al., 2016; Wendel et al., 2016; Alix et al., 2017; Mandáková et al., 2019). The cultivated strawberry (*Fragaria* × *ananassa* Duchesne ex Rozier) is unique among domesticated crop species because it arose through both processes. The chromosomes of octoploid garden strawberry (2n = 8x = 56) evolved through a combination of ancient polyploidy, and repeated homoploid hybridization in the last three centuries (Duchesne, 1766; Darrow, 1966). The presence of duplicated (homoeologous) chromosomes in plants frequently leads to meiotic anomalies and associated chromosomal rearrangements, e.g., translocations and inversions, that reduce or eliminate gene flow between the donors and their polyploid offspring (Soltis et al., 2014a; Soltis et al., 2014b; Alix et al., 2017; Latta et al., 2019; Mandáková et al., 2019). Similarly, meiotic mispairing in interspecific homoploid hybrids can lead to rearranged offspring chromosomes that differ from the chromosomes of one or both parents, resulting in reproductive barriers and hybrid speciation, as has been widely documented in sunflower (*Helianthus*) and other plants (Rieseberg, 1997; Burke et al., 2004; Abbott et al., 2010; Barb et al., 2014; Yakimowski and Rieseberg, 2014). However, reproductive barriers among octoploid *Fragaria* taxa remain essentially nonexistent, fueling the recurrence of interspecific homoploid hybridization in the origin, domestication, and modern-day breeding of *F.* × *ananassa*.

The modern *F.* × *ananassa* lineage traces its origin to extinct cultivars developed in western Europe in the 1700s. These cultivars were interspecific hybrids of non-sympatric wild octoploids from the New World: *F. chiloensis* subsp. *chiloensis* from South America and *F. virginiana* subsp. *virginiana* from North America (Darrow, 1966). Repeated introgression of alleles from diverse subspecific ecotypes of *F. virginiana* and *F. chiloensis* defined the later generations, coinciding artificial selection of horticulturally important traits among hybrid descendants in Europe and North America. Modern cultivars have emerged from 250 years of global migration and breeding within this admixed population (Darrow, 1966; Hardigan et al., 2018). Because alleles have been introgressed from up to eight octoploid subspecies, the genomes of modern *F.* × *ananassa* individuals are mosaics of their wild ancestors (Liston et al., 2014; Hardigan et al., 2018). Since the discovery of artificial hybrids at the Gardens of Versaille (Duchesne, 1766), natural interspecific hybrids (*F.* × *ananassa* subsp. *cuneifolia*) were discovered in zones of sympatry between *F. chiloensis* subsp. *pacifica* and *F. virginiana* subsp. *platypetala* in western North America (Hancock and Bringhurst, 1979; Luby et al., 1992; Staudt, 1999; Salamone et al., 2013). Neither cultivated *F.* × *ananassa* or wild *F.* × *ananassa* subsp. *cuneifolia* is reproductively isolated from their octoploid progenitors. Thus, genes appear to flow freely between the wild octoploid progenitors, and between the hybrids and their progenitors. While some genomic rearrangements have been identified between homoeologous chromosomes, and relative to the diploid species (Tennessen et al., 2014; van Dijk et al., 2014), the apparent absence of reproductive isolation implies that homoploid and polyploid hybridization events have not produced significant chromosome rearrangements among octoploid taxa. We hypothesized that the octoploids carry nearly collinear chromosomes tracing to the most recent common ancestor, despite one million years of evolution which produced multiple recognized species and subspecies.

The octoploid strawberry genome has been described as "notoriously complex" and an "extreme example of difficulty" for study (Folta and Davis, 2006; Hirsch and Buell, 2013; Hirakawa et al., 2014; Koskela et al., 2016). While GBS, GWAS, and NGS-reliant applications are relatively straightforward in organisms with well-characterized reference genomes, such approaches were previously difficult or intractable in octoploid strawberry (Liston et al., 2014; Tennessen et al., 2014; Bassil et al., 2015; Vining et al., 2017). Genetic studies in octoploid strawberry previously relied on the genome of woodland strawberry (*F. vesca*), an extant relative to one of four diploid subgenomes contained in *F.* × *ananassa* (Shulaev et al., 2011; Edger et al., 2017). For nearly a decade, the *F. vesca* genome was the only framework available for DNA variant discovery, gene discovery, genetic mapping, and genome-wide association studies in octoploid strawberry (Tennessen et al., 2014; Bassil et al., 2015; Davik et al., 2015; Vining et al., 2017; Pincot et al., 2018). The development of a chromosome-scale reference genome for *F.* × *ananassa* (Edger et al., 2019b) provided the physical framework needed to overcome previous barriers, and explore the organization and evolution of its progenitor genomes.

Here, we report the first study of chromosome evolution and genome structure in octoploid *Fragaria* using an octoploid genome-guided approach to DNA variant discovery and comparative mapping. We demonstrate the ability to differentiate duplicated (homoeologous) octoploid sequences using both NGS and array-based genotyping technologies when applied in conjunction with an octoploid reference genome. In doing so, we overcome a long-standing technical hurdle that has impeded efforts to study strawberry subgenome diversity and chromosome evolution. We estimated strawberry subgenomic diversity by whole-genome shotgun (WGS) sequencing of 93 genealogically and phylogenetically diverse *F.* × *ananassa*, *F. chiloensis*, and *F. virginiana* individuals. The frequency of unique WGS sequence alignment to the octoploid strawberry genome was characteristic of many diploid plant species (Hamilton and Buell, 2012; Lee and Schatz, 2012; Schatz et al., 2012; Treangen and Salzberg, 2012), and permitted the identification of millions of subgenome-specific DNA variants, effectively distinguishing homologous and homoeologous DNA sequences on every chromosome. Using the genetic diversity of *F.* × *ananassa*, we developed publicly available 50K and 850K SNP arrays populated with subgenome anchored marker probes for octoploid genetic mapping and forward genetic studies. We then performed high-density

genetic mapping of five octoploids representing *F.* × *ananassa* and both its octoploid progenitor species using a combination of WGS-based and array-based genotyping. Telomere-to-telomere genetic mapping of nearly every chromosome was enabled by the conserved disomic segregation observed in populations derived from wild species and *F.* × *ananassa*, underscoring the effective diploidization and meiotic stability of octoploid *Fragaria*. Comparative mapping of *F.* × *ananassa* and multiple subspecies of *F. chiloensis* and *F. virginiana* revealed the macro-syntenic structure of the cultivated reference genotype (Camarosa) and its progenitor species were nearly identical.

The collinear and diploidized genomes of *F.* × *ananassa* and its progenitors support octoploid *Fragaria* as an evolutionary clade which achieved a relatively high degree of genome stability prior to the speciation and sub-speciation of *F. chiloensis* and *F. virginiana*. The interspecific origin of *F.* × *ananassa* followed by successive hybridization throughout domestication is an unusual improvement pathway that frequently contributes to reproductive incompatibility or sterility in wide species hybrids (Ladizinsky, 1985; Hughes et al., 2007; Miller and Gross, 2011; Meyer and Purugganan, 2013). The preservation of genome structure among diverse octoploid *Fragaria* species and subspecies was likely essential to the unique history of the *F.* × *ananassa* lineage, which has undergone repeated cycles of homoploid hybridization without the formation of reproductive barriers or loss of fertility.

RESULTS AND DISCUSSION

Subgenomic Diversity of Octoploid Fragaria

We performed the first deep exploration of the homoeologous sequence diversity of octoploid *Fragaria* using the Camarosa v1.0 octoploid reference genome (Edger et al., 2019b) and a diversity panel of 93 strawberry individuals, including 47 F. × *ananassa*, 24 F. *chiloensis*, and 22 F. *virginiana* individuals (**Table S1**). By incorporating subgenome specificity at the assembly level, previous barriers to copy-specific sequence alignment caused by the octoploid ancestral homology of strawberry posed a less significant obstacle than repetitive DNA elements in diploid genomes such as maize (Hamilton and Buell, 2012; Lee and Schatz, 2012; Schatz et al., 2012; Treangen and Salzberg, 2012). The fraction of uniquely aligning (MapQ > 0) PE150 sequences averaged 83.2% (**Figure S1**), and 90.9% of Camarosa PE250 sequences aligned uniquely, allowing comprehensive coverage and analysis of subgenomic diversity. Using genotype calling software FreeBayes and a series of hard-filters targeting unique sequence alignments, we identified 41.8M subgenomic SNP and INDEL mutations in *F.* × *ananassa* and its wild progenitors.

F. × *ananassa* has been described as "genetically narrow" due to the small number of founders in the pedigrees of modern cultivars (Sjulin and Dale, 1987; Dale and Sjulin, 1990; Stegmeir et al., 2010). Despite a small effective population size, our analyses show that massive genetic diversity has been preserved in *F.* × *ananassa*, with negligible difference between wild species and domesticated germplasm. The subgenome nucleotide diversity (π) of *F.* × *ananassa* ($\pi = 5.857 \times 10^{-3}$) was equivalent to wild progenitors *F. chiloensis* ($\pi = 5.854 \times 10^{-3}$) and *F. virginiana* ($\pi = 5.954 \times 10^{-3}$), and comparable to the sequence diversity of *Zea mays* landraces ($\pi = 4.9 \times 10^{-3}$) and wild *Zea mays* spp. *parviglumis* progenitors ($\pi = 5.9 \times 10^{-3}$) (Hufford et al., 2012). Correlations of *F.* × *ananassa*, *F. chiloensis*, and F. *virginiana* diversity across the 28 octoploid chromosomes ranged from 0.93–0.97, showing that the magnitude and distribution of genomic diversity are broadly conserved among octoploid taxa. This suggested that *F.* × *ananassa* was not strongly bottlenecked by domestication, or that its domestication bottleneck was mitigated by continued introgression of allelic diversity from wild subspecies (Darrow, 1966). We found that variance in the distribution of octoploid nucleotide diversity was influenced more significantly by subgenome ancestry than domestication and breeding (**Table S2**). The diploid *F. vesca* subgenome, dominant with respect to gene abundance and expression (Edger et al., 2019b), contained the least diverse homoeolog of every ancestral chromosome, while subgenomes derived from the ancestors of the extant Asian species (*F. iinumae* and *F. nipponica*) contained greater genetic diversity (**Table S2**). Polyploid genome dominance contributes to differences in expression, gene loss, and purifying selection across subgenomes on the path to diploidy (Schnable et al., 2011; Grover et al., 2012; Parkin et al., 2014; Bird et al., 2018). Reduced nucleotide diversity on the dominant *F. vesca* subgenome supported distinct levels of purifying selection for the ancestral chromosomes of octoploid strawberry, with stronger selection in *F. vesca* ancestral sequences.

Because *F.* × *ananassa* was domesticated as an interspecific hybrid, individual performance was hypothesized to have been improved by allelic diversity between *F. chiloensis* and *F. virginiana*. To support comparisons of strawberry heterozygosity with previously studied polyploid species, we estimated individual heterozygosity based on the genomic frequency of heterozygous nucleotides (nts), a metric previously used in potato (Hardigan et al., 2017), and the frequency of heterozygosity at GBS-derived polymorphic sites, metrics previously used in blueberry and cotton (Page et al., 2013; de Bem Oliveira et al., 2019). Strawberry genomic heterozygosity ranged from 0.02-0.80% and averaged 0.46% genomic nts. This translated to an average of 11.1% heterozygosity at polymorphic marker sites. The most heterozygous octoploid individuals were early interspecific hybrids: White Carolina (PI551681; 0.80% nts), Peruvian Ambato (PI551736; 0.72% nts), Jucunda (PI551623; 0.70% nts), and Ettersberg 121 (PI551904; 0.66% nts). The average subgenomic heterozygosity of octoploid strawberry (0.46% nts) was below diploid potato (1.05% nts) and tetraploid potato (2.73% nts) (Hardigan et al., 2017). The average heterozygosity of octoploid strawberry at GBS-derived polymorphic sites was below the allo-tetraploid cotton A-genome (13% marker sites), above the cotton D-genome (< 1% marker sites) (Page et al., 2013), and below autotetraploid blueberry (32.4% marker sites) (de Bem Oliveira et al., 2019). Due to the presence of four ancestral homoeologs, traditional models of "fixed heterozygosity" applied to allopolyploid

species (Comai, 2005; Obbard et al., 2006) assume an octoploid functional heterozygosity four-fold greater than subgenomic estimates. Under this model, recent *F. virginiana* x *F. chiloensis* hybrids such as White Carolina and Jucunda could be regarded as similarly heterozygous to autopolyploid species such as potato. However, assembly of the allo-octoploid strawberry genome uncovered rampant gene silencing, gene loss, and homoeologous exchanges relative to diploid ancestors (Edger et al., 2019b), eroding the conservation of ancestral allele function. The frequency of unique sequence alignment (**Figure S1**) and unbroken distribution of subgenomic variant detection (**Figure S2**) in our analyses underscore the extensive divergence of the four subgenomes. Thus, traditional polyploid allele dosage models assuming genome-wide fixed heterozygosity may be of limited usefulness for strawberry.

Recombination Breakpoint Mapping of Octoploid Strawberry

We used WGS sequence analysis and recombination breakpoint mapping of an octoploid strawberry population to explore the breadth of disomic variation as an indicator of bivalent pairing during meiosis. Several cytogenetic and DNA marker studies have proposed the occurrence of polysomy in strawberry (Fedorova, 1946; Senanayake and Bringhurst, 1967; Lerceteau-Köhler et al., 2003), while others suggest that octoploids are mainly disomic (Byrne and Jelenkovic, 1976; Arulsekar and Bringhurst, 1981; Bringhurst, 1990; Kunihisa et al., 2005). We performed low-coverage (4-8x) sequencing and subgenomic DNA variant calling in a population (n = 182) derived from a cross between the *F.* × *ananassa* cultivar 'Camarosa' and the beach strawberry (*F. chiloensis* subsp. *lucida*) ecotype 'Del Norte'. These parents were selected to provide a dense comparison of profiles of mappable disomic polymorphisms in wild and domesticated octoploid individuals. Variant calling against the Camarosa v1.0 genome identified 3.7M subgenomic SNPs and INDELs inherited from 1.6M Camarosa heterozygous sites (AB × AA), 1.9M *F. chiloensis* subsp. *lucida* heterozygous sites (AA × AB), and 0.2M co-heterozygous sites (AB × AB). We used the high-density DNA variant data to perform haplotype mapping based on recombination breakpoint prediction and evaluated segregation ratios of parental alleles across the 28 octoploid chromosomes.

We bypassed the computational demand of analyzing pairwise linkage across millions of DNA variants with missing data and genotyping errors by implementing the haplotype calling approach proposed by Huang et al. (2009) and Marand et al. (2017). Our approach performed a sliding-window analysis to predict crossover events, then estimated the consensus of co-segregating DNA variation between recombination breakpoints to reconstruct the representative genotypes of each haploblock. The haploblocks were mapped as unique genetic markers. Using this approach, we mapped 1.9M *F. chiloensis* subsp. *lucida* DNA variants in 5,521 haploblocks spanning 3,393.86 cM, and 1.6M Camarosa DNA variants in 3,179 haploblocks spanning 2,016.95 cM (**Dataset S1**). The paternal *F. chiloensis* subsp. *lucida* map produced telomere-to-telomere coverage of the 28 octoploid chromosomes (**Figure 1**), providing the most comprehensive genetic map of an octoploid *Fragaria* genome to-date. The complete mapping of the extant homoeologs for all seven ancestral *Fragaria* chromosomes in the paternal genome and analysis of chromosome-wide segregation distortion (**Figure 2**) showed that disomic recombination is ubiquitous in the genome of *F. chiloensis*. By contrast, less than 50% of the Camarosa genome could be mapped on chromosomes 1-1, 1-2, 2-4, 3-3, 5-2, 6-2, 6-3, 6-4, and 7-3 (**Figure S3**). We analyzed Camarosa heterozygosity and segregation distortion to determine whether the inability to map large segments of the genome was the result of polysomic recombination in *F.* × *ananassa*. This uncovered a near total loss of polymorphism in the unmapped regions of Camarosa (**Figure 2**), showing that incomplete mapping of *F.* × *ananassa* resulted from depletion of heterozygosity in the hybrid genome, not polysomy. Sargent et al. (2012) previously reported extensive regions of homozygosity that affected mapping of *F.* × *ananassa*. Artificial selection pressure in commercially bred hybrids almost certainly accounts for the lower subgenomic heterozygosity of Camarosa relative to *F. chiloensis* subsp. *lucida*, which does not support a critical role for genome-wide interspecific heterozygosity in driving cultivar performance.

850K Octoploid Screening Array

We designed Affymetrix SNP genotyping arrays populated with subgenome-specific marker probes to enable genetic mapping, genome-wide association studies (GWAS), and genomic prediction in octoploid strawberry. DNA variants were selected for array design from the subgenomic diversity identified in the WGS panel (**Figure 3**). From 90M total unfiltered variant sites, we extracted 45M unfiltered variants that segregated in *F.* × *ananassa*. To identify candidate DNA variants for marker design, we selected only biallelic SNPs above a low-diversity threshold ($\pi \geq 0.05$), excluded rare alleles (MAF ≥ 0.05), required a VCF quality score > 20, and excluded sites with > 15% missing data in the diversity panel. These filters yielded 8M subgenomic SNPs segregating within the *F.* × *ananassa* subset of the diversity panel. We obtained 71-nt marker probes by extracting 35-nt sequences flanking each SNP site from the Camarosa v1.0 genome assembly. Marker probes for the 8M high-confidence SNP sites were then filtered to remove candidates that were problematic for array tiling. These included duplicate or near-duplicate probe sequences and probes that inherited ambiguous reference sequences (Ns), required double-tiling (A/T or C/G alleles), or that Affymetrix scored as having low buildability. We retained 6.6M probes that targeted high-confidence *F.* × *ananassa* variants and were acceptable for array tiling.

We applied three selection criteria for determining a subset of 850K marker probes for tiling: likelihood of probe binding interference by off-target variants, likelihood of off-target (non-single copy) probe binding, and physical genome distribution. The likelihood of probe binding interference was scored as the sum of non-reference allele frequencies for off-target variants in the 35-nt binding region adjacent to the target SNP. The likelihood of off-target probe binding was scored by performing BLAST alignment of the 71-nt probe sequences to the Camarosa v1.0 genome and quantifying the number of off-

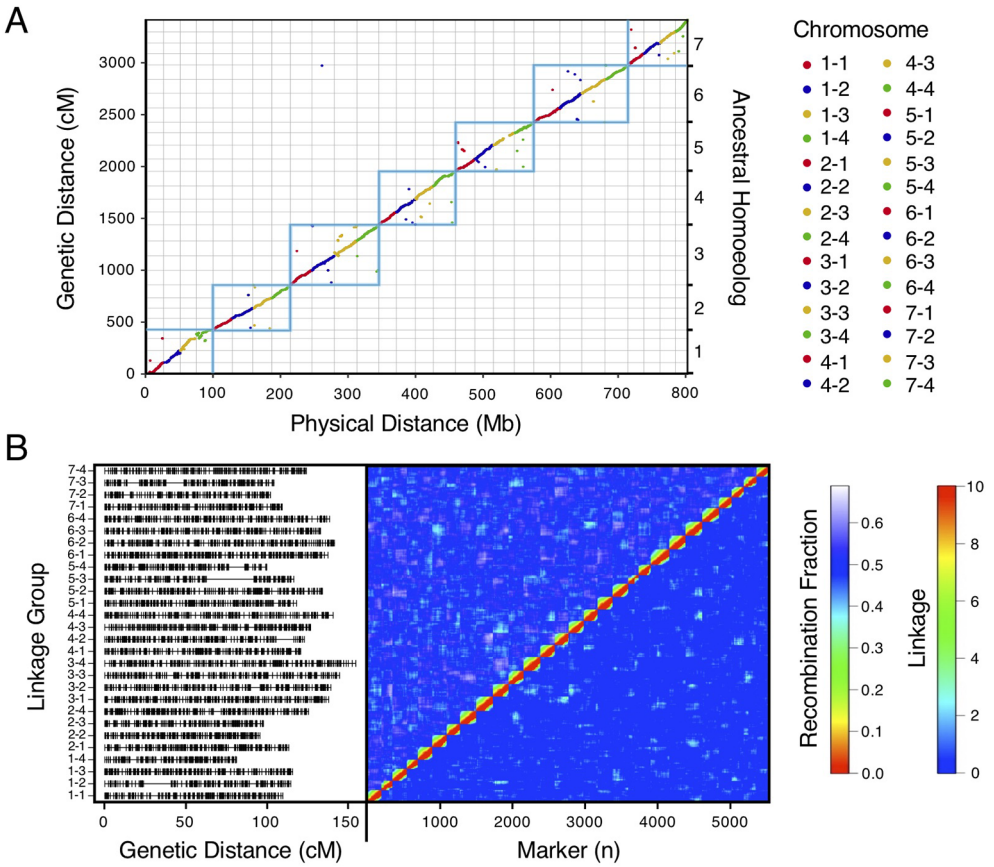

FIGURE 1 | High-density haplotype map of a California beach strawberry (*F. chiloensis* subsp. *lucida*) genome. **(A)** Del Norte genetic map distances plotted against the Camarosa v1.0 physical genome. Box outlines indicate groups of ancestral chromosome homoeologs. **(B)** Del Norte linkage groups plotted with corresponding chromosome heatmap of pairwise recombination fractions (upper diagonal) and pairwise linkage (lower diagonal).

target alignments with query coverage above 90% and sequence identity above 90%. We then iteratively parsed the Camarosa v1.0 genome using 10 kilobase (kb) non-overlapping physical windows, extracting the best available marker from each window based on probe binding interference and off-target binding likelihoods, until reaching an 850K probe threshold. We reserved 16K positions for legacy markers from the iStraw SNP array (Bassil et al., 2015; Verma et al., 2016) that were polymorphic in a previous strawberry diversity study (Hardigan et al., 2018). The set of 850K probe sequences was submitted to Affymetrix for constructing a screening array.

We genotyped a genetically diverse sample of 384 octoploid strawberry accessions to validate subgenome-specific marker performance on the 850K screening array (**Dataset S2**). The sample fluorescence files were analyzed with the Axiom Suite in polyploid mode to generate marker clusters. Collectively, 446,644 of 850,000 marker probes produced QC-passing polymorphic SNP genotype clusters showing disomic (allopolyploid) segregation. Among these, 78.3% were classified as "PolyHighResolution" in the Axiom terminology, producing diploid co-dominant genotype clusters (AA, AB, and BB) without detecting off-target allelic variation. Similarly, 18.8% of markers classified as "NoMinorHomozygote" in the Axiom terminology produced dominant genotype clusters in which heterozygotes (AB) clustered with one of the homozygous genotype classes (AA or BB). The remaining 2.9% of the markers detected non-target alleles and were classified as "OffTargetVariant" markers in the Axiom terminology. The HomRO statistic generated by the Axiom Suite estimates genotype cluster separation, and has been used as a metric to infer octoploid single-copy (i.e. subgenome or paralog specific) probe binding when values exceed 0.3 (Bassil et al., 2015). Based on this threshold, 74% of the QC-passing marker probes on the 850K screening array exhibited single-copy binding, in addition to measuring subgenome-specific DNA variation (**Figure S4**). The complete set of 446,644 validated probes is made available for public use (**Dataset S3**).

50K Octoploid Production Array

We selected 49,483 polymorphic marker probes from the 850K validated probe set to build a 50K production array (**Dataset S4**). We included 5,809 LD-pruned ($r^2 < 0.50$) marker probes from

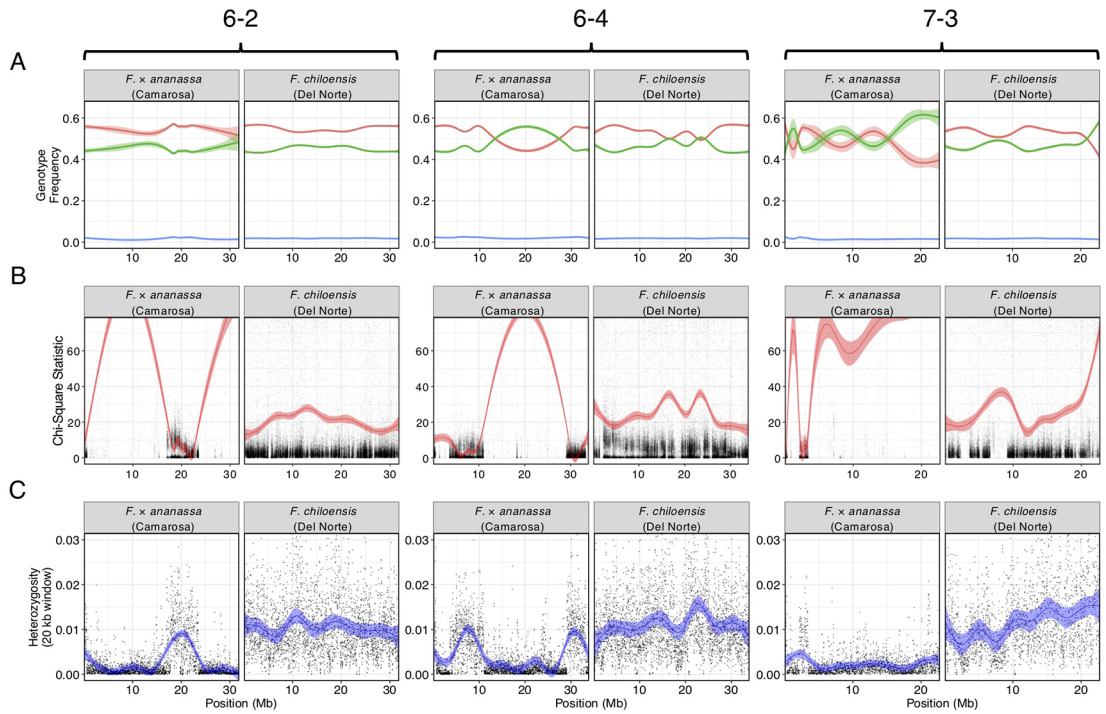

FIGURE 2 | Comparison of genomic heterozygosity and diploid segregation distortion in *F.* × *ananassa* (Camarosa) and *F. chiloensis* (Del Norte) on three chromosomes (6-2, 6-4, 7-3). **(A)** Frequency of AA (green), AB (red), and off-target (BB; blue) genotypes for polymorphic markers in the mapping population. **(B)** Chi-square statistic estimating segregation distortion of polymorphic markers in the mapping population. **(C)** Heterozygous nucleotide frequency of parent genotypes in 20 kb physical windows.

the iStraw35 SNP array to support cross-referencing with previous genetic and QTL mapping studies. We targeted 2,878 genes based on Camarosa v1.0 functional annotations that indicated R-gene affiliated protein domains (Edger et al., 2019b) or homology to *Fragaria vesca* genes involved in flowering and fruit development expression networks (Kang et al., 2013; Hollender et al., 2014). Candidate genes were pre-allocated up to two markers (within 1 kb) from the screening panel. We next selected a set of the most commonly segregating markers to support genetic mapping. We identified this set by selecting the marker with the highest pairwise diversity (π) in *F.* × *ananassa* across non-overlapping 50 kb physical genome windows. The remainder of the 50K array was populated by iteratively parsing the genome with 50 kb physical windows and selecting random QC-passing markers to provide an unbiased genome distribution. Both the 850K and 50K probe sets provide unbroken, telomere-to-telomere physical coverage of the 28 octoploid strawberry chromosomes (**Figure S5**). Within the 50K probe set, 53% of the probes were located within genes, and 79% were located within 1 kb of a gene. The 50K probe set was provided to Affymetrix for building the production array.

We screened 1,421 octoploid samples from multiple bi-parental populations and a large diversity panel on the 50K production array. Collectively, 42,081 markers (85%) successfully replicated QC-passing polymorphic genotype clusters when screened in the larger sample group. Of the 7,402 non-replicated markers, only 1% were excluded due to becoming monomorphic ("MonoHighResolution" Axiom class) or from increased missing data ("CallRateBelowThreshold" Axiom class). Sub-clustering and increased dispersion within the AA or BB genotype clusters ("AAvarianceX", "AAvarianceY", "BBvarianceX", "BBvarianceY" Axiom classes) accounted for 20% on non-replicated markers. The Axiom software provided no specific cause for failure for the remaining 79% of non-replicated markers. These results suggest that increasing the size and diversity of a genotyping population may affect the reproducibility of a fraction (15%) of markers on the 50K array, while a majority (85%) are highly reproducible. The fraction of polymorphic co-dominant ("PolyHighResolution") markers increased from 78% on the 850K screening array to 86% among reproducible markers 50K production array. The average correlation between genotypes predicted for the Camarosa × Del Norte bi-parental population based on low-coverage sequencing and 50K array probes at polymorphic segregating sites was 0.93.

Genetic Mapping of Wild Octoploid Ecotypes

We demonstrated that the 50K SNP array allows dense genetic mapping of heterozygous regions on all 28 chromosomes of *F.* × *ananassa*, *F. chiloensis*, and *F. virginiana*. Genetic mapping of *F. chiloensis* and *F. virginiana* provided telomere-to-telomere physical representation of the 28 octoploid chromosomes and near-complete representation within individual wild maps (**Dataset S1**). We

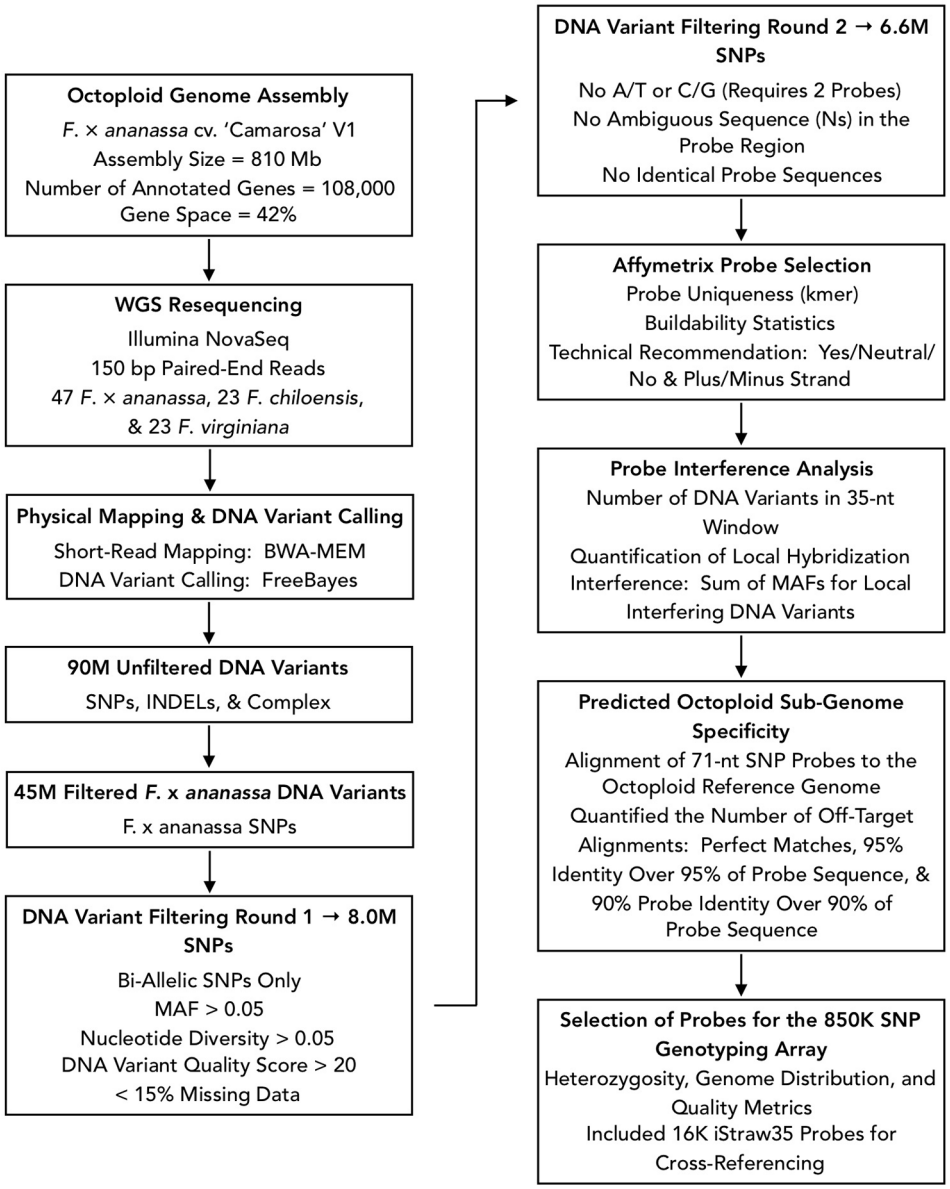

FIGURE 3 | Flowchart of bioinformatic protocols used to select genome-wide sequence variants for design of the 850K SNP screening array.

selected four octoploid parents from two outcrossing bi-parental populations genotyped on the 50K array for mapping. The first was the Camarosa × Del Norte population used for WGS recombination breakpoint mapping (n = 182). The second population was derived from a cross between *F. virginiana* subsp. *virginiana* accession PI552277 (female parent) and *F. virginiana* subsp. *virginiana* accession PI612493 (male parent) (n = 96). The number of SNPs segregating in the octoploid parent genotypes varied considerably (**Table S3**). Camarosa contained the most segregating markers (9,062), followed by PI552277 (5,575), PI612493 (5,464), and Del Norte (2,368). The larger number of markers segregating in Camarosa relative to the wild parent genotypes reflected the array design strategy, which targeted *F.* × *ananassa* diversity. The unbalanced representation of *F. virginiana* and *F. chiloensis* diversity was not expected, because genome-wide variant calls showed similar profiles of heterozygosity and nucleotide diversity in the wild progenitors, and showed that *F. chiloensis* subsp. *lucida* was more heterozygous than Camarosa. The higher level of ascertainment bias against *F. chiloensis* diversity that resulted from probing *F.* × *ananassa* alleles supports previous findings that *F. virginiana* diversity is more prevalent in cultivated hybrids (Hardigan et al., 2018). We mapped heterozygous variant sites of the four parent octoploids using software ONEMAP (Margarido et al., 2007) to generate initial linkage groups and markers orders

and BatchMap (Schiffthaler et al., 2017) for marker re-ordering and genetic distance estimation. Despite the ascertainment bias for domesticated allelic diversity on the 50K array, the relatively unbiased distribution of genomic heterozygosity in wild genotypes (**Table S3**) provided a more complete representation of the wild octoploid genomes than *F.* × *ananassa* (Camarosa) (**Figure 4**). Large homozygous regions that produced breaks in the Camarosa WGS haplotype map and 50K array map (chromosomes 1-1, 1-2, 2-4, 3-3, 5-2, 6-2, 6-3, 6-4, and 7-3) were clearly featured in the wild genetic maps (**Figure 4**). Camarosa contained an average of 11.7 (± 6.8) SNPs/megabase (Mb) across 28 chromosomes, with as many as 25.2 SNPs/Mb (1-4) and as few as 0.8 SNPs/Mb (1-3), underscoring the scattered distribution of mappable subgenomic diversity in the cultivated strawberry hybrid.

The wild octoploid maps revealed large (Mb+) chromosomal rearrangements relative to the Camarosa v1.0 physical genome on chromosomes 1-2, 1-4, 2-1, 2-3, 6-2, and 6-4. These rearrangements were conserved across the wild species genomes, and supported by corresponding regions represented in the Camarosa genetic map (1-2, 1-4, 2-1) (**Figure S3**), indicating intra-chromosomal scaffolding errors in the physical reference genome. The fraction of SNPs that genetically mapped to non-reference chromosomes ranged from 1.5%–1.9% in the four parents, with the highest fraction observed in Camarosa. This indicated minimal inter-chromosomal errors in the physical genome and minimal inter-chromosomal marker discordance between octoploid progenitor species. Thus, the octoploid genetic maps provided no evidence of chromosome rearrangement among the wild and cultivated octoploid species.

Genome Structure of Ancestral Species

Previous studies have reported octoploid chromosome rearrangements relative to diploid *Fragaria*, potentially contributing to sex determination (Spigler et al., 2008; Spigler et al., 2010; Tennessen et al., 2014; Govindarajulu et al., 2015). Moreover, there is phylogenetic evidence of chromosome exchanges among the four ancestral subgenomes (Liston et al., 2014; Edger et al., 2019b). However, there is no evidence for chromosome-scale structural variation between octoploid taxa. It remains unclear to what extent the structural variation of octoploid *Fragaria* reflects initial polyploid 'genome shock' occurring in the common ancestor, as opposed to ongoing mutations contributing to octoploid species diversification. Through comparative mapping, we show that the genomes of diverse octoploid ecotypes contributing to the homoploid hybrid lineage of *F.* × *ananassa* are nearly completely collinear. We constructed genetic maps for two additional wild genotypes, *F. chiloensis* subsp. *pacifica* (SAL3) and *F. virginiana* subsp. *platypetala* (KB3) using publicly available DNA capture libraries to obtain a more diverse set of octoploid subspecific taxa. We aligned capture sequences from an *F. chiloensis* subsp. *chiloensis* × *F. chiloensis* subsp. *pacifica* population (GP33 × SAL3, n = 46; Tennessen et al., 2014) and an *F. virginiana* subsp. *platypetala* × *F. virginiana* subsp. *platypetala* population (KB3 × KB11, n = 46; Tennessen et al., 2018) to the Camarosa v1.0 genome assembly and predicted subgenome DNA variant genotypes using FreeBayes. Genetic mapping of the DNA capture markers followed the protocol used for the 50K SNP datasets. Using the 50K array linkage groups and DNA capture linkage groups (**Dataset S1**), we performed comparative mapping of four octoploid subspecies—*F. chiloensis* subsp. *lucida* (Del Norte), *F. chiloensis* subsp. *pacifica* (SAL3), *F. virginiana* subsp. *platypetala* (KB3), and *F. virginiana* subsp. *virginiana* (PI552277)—in ALLMAPS using genetically mapped DNA variant sites anchored to 50-kb physical genome windows in Camarosa v1.0. The chromosomes of the octoploid progenitor subspecies were completely syntenic (**Figure 5**, **Figure S6**). Based on these results, large-scale chromosome rearrangements in octoploid *Fragaria* relative to the diploid ancestral genomes would have occurred before the speciation of *F. chiloensis* and *F. virginiana*.

Comparative genomic analysis of the four octoploid strawberry subgenomes, in addition to diploid strawberry species *F. vesca and F. iinumae* and black raspberry (*Rubus occidentalis*), revealed a remarkable degree of karyotypic stability not only within octoploid strawberry, but within related species that have evolved independently for millions of years (**Figure 6**). While several rearrangements and inversions were detected between ancestral diploid chromosomes, they primarily consisted of large, syntenic blocks with high degrees of collinearity (**Figure 6**). These results demonstrate that the structural conservation retained between octoploid species extends to broader *Fragaria* taxa, including strawberry species that are estimated to have diverged ~7.5 million years ago and beyond that to a distantly related Rosaceous species estimated to have diverged 33 M years ago (Njuguna et al., 2013; Qiao et al., 2016). This is unique compared to extensive karyotype evolution documented in polyploid species involving numerous chromosomal fusion and rearrangement events, e.g. *Camelina sativa* (Mandáková et al., 2019).

Unification of Octoploid Chromosome Nomenclatures

Previous octoploid genetic mapping studies relied on a variety of DNA marker technologies including early PCR-based assays (e.g., simple sequence repeats and amplified fragment length polymorphisms) and technical ploidy reduction of sequence variants called against the diploid *F. vesca* genome (Rousseau-Gueutin et al., 2008; Spigler et al., 2008; Spigler et al., 2010; Tennessen et al., 2014; Bassil et al., 2015; Vining et al., 2017; Tennessen et al., 2018). This diversity of DNA marker genotyping strategies without information linking to the *F.* × *ananassa* physical genome has caused a proliferation of disconnected strawberry chromosome nomenclatures that may not accurately reflect the phylogenetic origins of its respective subgenomes. The Camarosa v1.0 reference genome provides an anchoring point for unifying the existing octoploid nomenclatures. We aligned all historic *Fragaria* microsatellite markers in the Rosacea Genomic Database (GDR) to the Camarosa v1.0 genome and anchored the Spigler et al. (2010) nomenclature to the Camarosa physical genome, which provided the corresponding linkage groups for anchoring the Tennessen

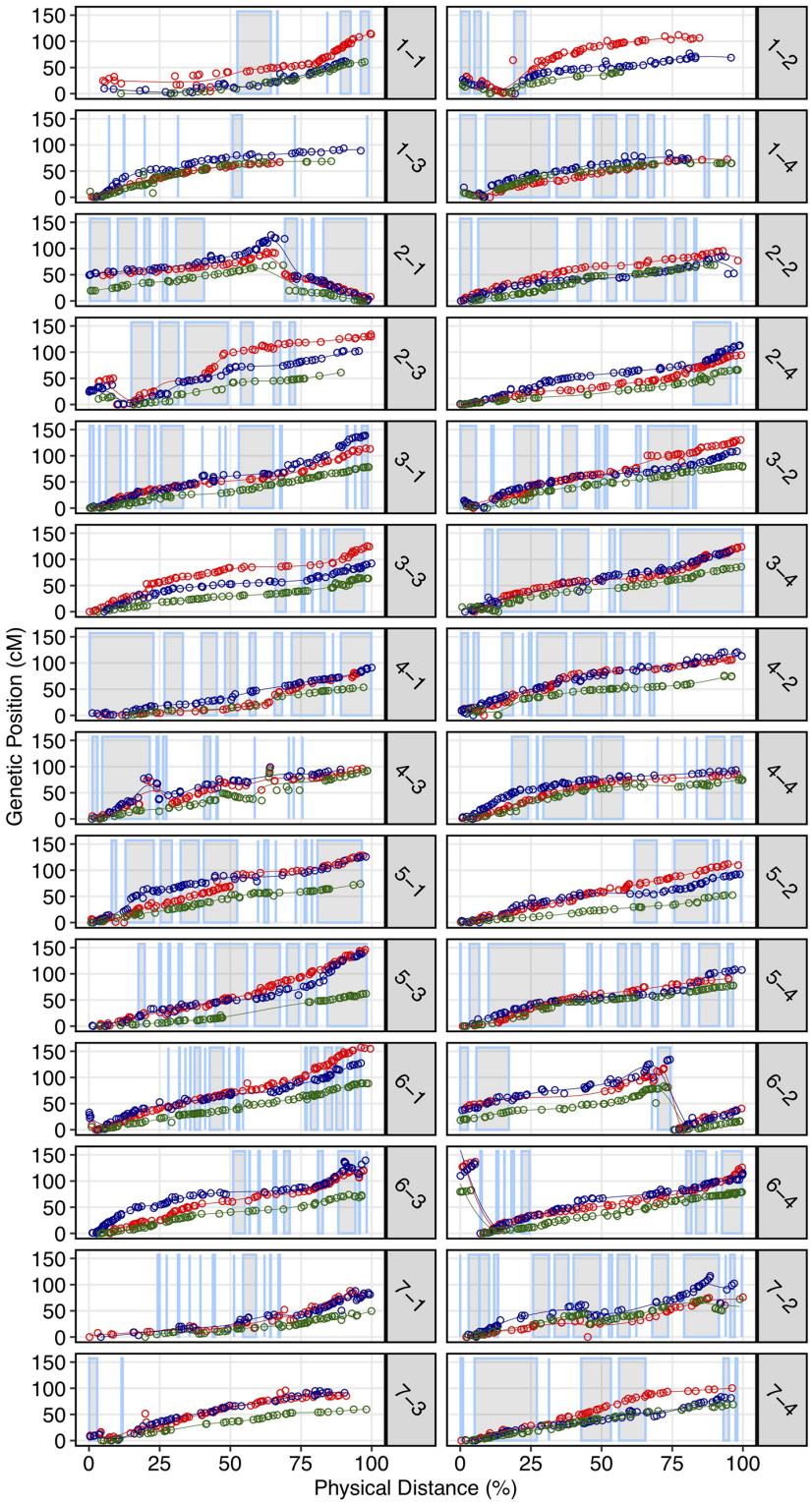

FIGURE 4 | Genetic maps of three wild octoploid strawberry genotypes (PI552277–red; PI612493–blue; Del Norte–green) based on 50K SNP array genotypes plotted against the Camarosa v1.0 physical genome. Grey highlighted chromosome segments indicate contiguous (up to 500 kb) regions of the physical genome represented by the Camarosa 50K SNP array map.

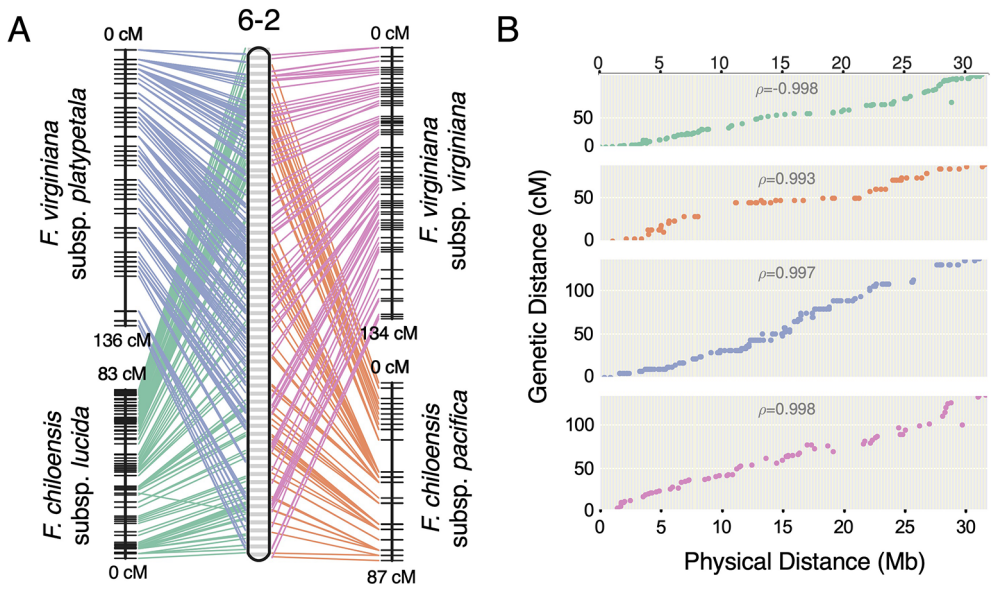

FIGURE 5 | Comparative mapping of four wild octoploid *Fragaria* subspecies (*F. chiloensis* subsp. *lucida*, *F. chiloensis* subsp. *pacifica*, *F. virginiana* subsp. *platypetala*, *F. virginiana* subsp. *virginiana*) on chromosome 6-2. **(A)** Marker collinearity between individual species maps and the consensus ordering of 50-kb physical windows from the Camarosa v1.0 physical genome. **(B)** Marker genetic distances plotted against the consensus ordering of 50-kb physical windows.

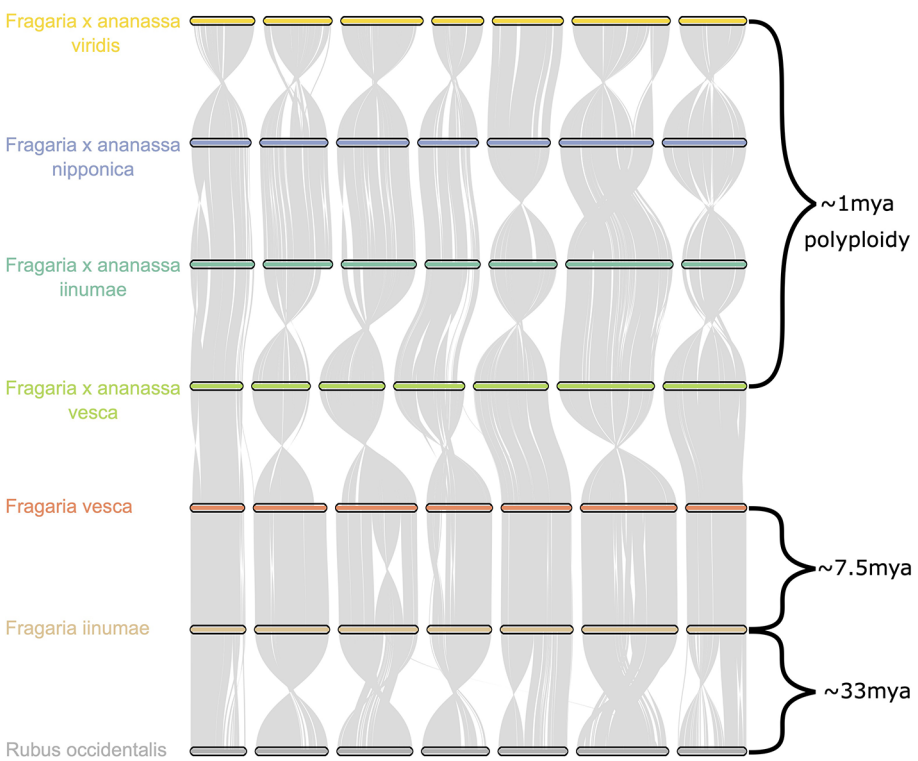

FIGURE 6 | Chromosome scale collinearity of the *F.* × *ananassa* subgenomes, *F. vesca*, *F. iinumae, and Rubus occidentalis*. Large inversions reflect the orientations of genome assemblies, not whole-chromosome inversions. Divergence estimates for *F. vesca* and *F. iinumae*, and *Fragaria and Rubus* are from Qiao et al. (2016) and estimate for date of *F.* × *ananassa* polyploid are from Njuguna et al. (2013).

et al. (2014) nomenclature. We then utilized legacy iStraw probes retained on the 50K array to link the Sargent and van Dijk chromosome nomenclatures (van Dijk et al., 2014; Sargent et al., 2016) to the Camarosa v1.0 genome, which was scaffolded using the map published by Davik et al. (2015). In total, five of the most widely cited octoploid strawberry chromosome nomenclatures were unified in relation to the physical genome (**Table 1**). The existing octoploid nomenclatures each contained subgenome assignments that were incongruent with ancestral chromosomal origins determined by phylogenetic analysis of the physical genome (Edger et al., 2019b; Edger et al., 2019a), particularly with respect to the non-*F. vesca* and non-*F. iinumae* subgenomes. The unmasking of octoploid homoeologous chromosome lineages (Edger et al., 2019b; Edger et al., 2019a) and construction of genetic maps showing complete collinearity among ancestral species provide the foundation for a unified octoploid nomenclature reflecting the phylogenetic origins of its subgenomes.

CONCLUSION

Using the first octoploid genome-guided approach to subgenomic (diploid) DNA variant discovery, we have demonstrated that the genomes of the octoploid progenitors of *F.* × *ananassa* are highly collinear and diploidized (**Figure 5**, **Figure S6**). Octoploid *Fragaria* taxa do not follow the common polyploid rule book for chromosome rearrangement (Ramsey and Schemske, 2002; Gaeta et al., 2007; Leitch and Leitch, 2008; Cifuentes et al., 2010; Gaeta and Pires, 2010; Chester et al., 2012; Renny-Byfield and Wendel, 2014; Wendel et al., 2016; Alix et al., 2017), but instead exhibit incredible karyotypic stability across biogeographically diverse subspecies. Strikingly, we did not observe any large-scale (Mb+) structural rearrangements (e.g., translocations or inversions) in the genomes of *F. chiloensis*, *F. virginiana*, or *F.* × *ananassa* (**Figure 4**). The broad conservation of chromosome structure across diverse progenitor taxa partly explains the absence of reproductive barriers and ease of gene flow between wild octoploid species and the domesticated hybrid lineage. In this regard, octoploid *Fragaria* species appear part of a minority among polyploid plants, though similar examples of karyotypic stability have been described in monocots and dicots (Sun et al., 2017; VanBuren et al., 2019). Because of the ubiquity of polyploidy in angiosperms and the diversity of chromosome-restructuring outcomes along the pathway to diploidization, universal rules do not necessarily apply (Cifuentes et al., 2010; Le Comber et al., 2010; Renny-Byfield and Wendel, 2014; Wendel et al., 2016). The remarkable karyotypic stability and absence of chromosome rearrangements among octoploid *Fragaria* taxa are indicative of regular diploid meiotic behavior and suggest that homoeologous recombination has failed to disrupt the ancestral octoploid karyotype, which has been preserved over the past 0.4-2.1 million years (Tennessen et al., 2014). The unique history of strawberry as a crop lineage, including its origin as an interspecific hybrid and frequent use of interspecific hybridization throughout domestication, was

TABLE 1 | Published octoploid strawberry linkage group nomenclatures anchored to corresponding physical chromosomes in the Camarosa v1.0 reference genome.

Octoploid Chromosome	Proposed Origin (Edger et al., 2019b)	Spigler et al. (2010)	Tennessen et al. (2014)	van Dijk et al. (2014)	Davik et al. (2015)	Sargent et al. (2016)
1-2	*F. iinumae*	I.D	I-Bi	1C	1-B	1b
2-4	*F. iinumae*	II.B	II-Bi	2C	2-D	2b
3-2	*F. iinumae*	III.A	III-Bi	3D	3-B	3b
4-4	*F. iinumae*	IV.C	IV-Bi	4A	4-D	4b
5-3	*F. iinumae*	V.D	V-Bi	5B	5-C	5b
6-3	*F. iinumae*	VI.D	VI-Bi	6B	6-C	6X2
7-3	*F. iinumae*	VII.A	VII-Bi	7D	7-C	7b
1-3	*F. nipponica*	I.C	I-B2	1B	1-C	1X2
2-1	*F. nipponica*	II.D	II-B1	2D	2-A	2X2
3-3	*F. nipponica*	III.C	III-B1	3B	3-C	3X2
4-2	*F. nipponica*	IV.D	IV-B2	4D	4-B	4X2
5-4	*F. nipponica*	V.B	V-B2	5D	5-D	5X1
6-2	*F. nipponica*	VI.B	VI-B1	6D	6-B	6b
7-1	*F. nipponica*	VII.D	VII-B2	7C	7-A	7X2
1-4	*F. vesca*	I.A	I-Av	1A	1-D	1A
2-2	*F. vesca*	II.A	II-Av	2A	2-B	2A
3-4	*F. vesca*	III.D	III-Av	3A	3-D	3A
4-3	*F. vesca*	IV.A	IV-Av	4B	4-C	4A
5-1	*F. vesca*	V.C	V-Av	5A	5-A	5A
6-1	*F. vesca*	VI.A	VI-Av	6A	6-A	6A
7-2	*F. vesca*	VII.B	VII-Av	7A	7-B	7A
1-1	*F. viridis*	I.B	I-B1	1D	1-A	1X1
2-3	*F. viridis*	II.C	II-B2	2B	2-C	2X1
3-1	*F. viridis*	III.B	III-B2	3C	3-A	3X1
4-1	*F. viridis*	IV.B	IV-B1	4C	4-A	4X1
5-2	*F. viridis*	V.A	V-B1	5C	5-B	5X2
6-4	*F. viridis*	VI.C	VI-B2	6C	6-D	6X1
7-4	*F. viridis*	VII.C	VII-B1	7B	7-D	7X1

almost certainly supported by the uncommon stability of its progenitor genomes, allowing unrestricted gene flow across octoploid genetic backgrounds. Comparative genomic analysis of octoploid strawberry with related diploid species and *Rubus occidentalis* revealed this stability extends back over 30 million years. Thus, the high degree of structural conservation among genomes of diverse ancestral *Fragaria* species may factor into the absence of any chromosomal fusions and/or rearrangements arising as a result of polyploid or homoploid hybridization.

The global importance and rapid commercial success of *F.* × *ananassa* over the last 250 years has been attributed to the interspecific homoploid hybrid component of heterosis (Spangelo et al., 1971; Shaw, 1995; Stegmeir et al., 2010; Rho et al., 2012). Quantitative evidence for heterosis in *F.* × *ananassa* is limited (Spangelo et al., 1971; Shaw, 1995; Stegmeir et al., 2010; Rho et al., 2012). Heterosis is an often-cited advantage of polyploidy, where genomic heterozygosity preserved by subgenomic recombination maintains "fixed heterosis" in allopolyploids (Comai, 2005). The *F.* × *ananassa* genome is riddled with ancient homoeologous exchanges (Edger et al., 2019b), a hallmark of inter-subgenomic recombination in early generations. Neverthless, the genomes of present-day octoploid taxa appear to be highly diploidized. We observed disomic inheritance of DNA variants across the genomes of the octoploids in the present study, and similar ranges of subgenomic heterozygosity for wild individuals and commercial hybrids. The success of *F.* × *ananassa* should not be solely attributed to "fixed heterosis" because neither octoploid progenitor species, which share the effects of fixed heterosis and show similar subgenomic heterozygosity, was commercially successful before the hybrid (Darrow, 1966; Finn et al., 2013). We hypothesize that interspecific complementation, a broader pool of potentially adaptive alleles, and masking of deleterious mutations could be more important than fixed heterosis in *F.* × *ananassa* (Comai, 2005; Alix et al., 2017).

We have shown that the inherent complexity and previous intractability of octoploid strawberry genomics were largely associated with the technical challenge of distinguishing subgenome level variation from the broader pool of ancestral sequence homology. The use of an allo-octoploid reference genome addressed this problem by allowing variant calling based on unique sequence alignments to the respective subgenomes. While local subgenome homology could remain an issue, we identified a nearly continuous distribution of subgenome-specific variation spanning the octoploid genome by traditional short-read sequencing. With the design of the 850K and 50K arrays, facile high-density genotyping of octoploid-anchored disomic SNP markers has been further enabled, reducing the bioinformatic requirements for octoploid strawberry research. In addition to expanding and validating the current molecular toolset, we have demonstrated that allopolyploid reference genomes facilitate the use of straightforward diploid approaches for genomic analysis and quantitative genetics of octoploid strawberry. In doing so, the results of this study help pave the way for molecular breeding of a historically difficult plant genome.

MATERIALS AND METHODS

WGS Sequence Datasets

We generated Illumina sequencing libraries for a diversity panel of 84 wild and domesticated octoploid genotypes (PE150), and the Camarosa reference genotype (PE250). Eight sequenced octoploid libraries (PE100) were obtained from the NCBI sequence read archive (SRA) (SRR1513906, SRR1513893, SRR1513905, SRR1513903, SRR1513892, SRR1513904, SRR1513867, SRR1513873), providing a total of 93 sample libraries in the diversity panel. We generated additional Illumina libraries (PE150) for 182 progeny of an *F.* × *ananassa* x *F. chiloensis* subsp. *lucida* mapping population (Camarosa x Del Norte). Genomic DNA was extracted from immature leaf tissue using the E-Z 96 Plant DNA kit (Omega Bio-Tek, Norcross, GA, USA) with Proteinase K was added to the initial buffer, and RNase treatment delayed until lysate was removed from the cellular debris. An additional spin was added, and incubation steps were heated to 65 C during elution. Libraries were prepared using the KAPA Hyper Plus kit using BIOO Nextflex adapters. DNA was sheared using the Covaris E220 and size selected for an average insert size of 300-nt using magnetic beads (Mag-Bind® RxnPure Plus, Omega Bio-tek). Libraries were pooled and sequenced on the Novaseq at UCSF for average 8x genome coverage in the diversity panel, and 4-8x coverage in the mapping population. DNA capture-based Illumina sequences for the *F. chiloensis* subsp. *lucida* x *F. chiloensis* subsp. *pacifica* population (GP33 x SAL3: n = 46) described in Tennessen et al. (2014), and the *F. virginiana* subsp. *platypetala* x *F. virginiana* subsp. *platypetala* population (KB3 x KB11: n = 46) described in Tennessen et al. (2018) were downloaded from the NCBI SRA.

Subgenomic WGS Variant Calling

We predicted SNP and INDEL variants in the Camarosa v1.0 subgenomes using sequenced Illumina short-read libraries for the octoploid diversity panel, Camarosa x Del Norte bi-parental population, and DNA capture sequences downloaded from the SRA. Short-read sequences were quality-trimmed with CutAdapt v1.8 using default parameters and a minimum Phred score of 25. Trimmed short-reads were aligned to the Camarosa v1.0 genome assembly (Edger et al., 2019b) using BWA-mem v0.7.16, processed to mark optical and PCR duplicates using Picard Tools v2.18, and INDEL-realigned using GATK v3.8. Genomic variants were predicted based on uniquely mapped reads (MapQ > 20) using FreeBayes v1.2 and filtered with vcflib. For analysis of subgenomic variation, a set of hard-filters was applied to remove variants with low site quality (vcflib: QUAL > 40), low contribution of allele observations to site quality (vcflib: QUAL/AO > 10), low read coverage (vcflib: DP > 500), strand bias (vcflib: SAF > 0 & SAR > 0), read-placement bias (RPR > 1 & RPL > 1), unbalanced mapping quality of reference and alternate alleles (vcflib: $0.4 \leq$ [MQM/MQMR] ≤ 2.5), unbalanced allele frequencies at heterozygous sites (vcflib: $0.2 \leq$ AB ≤ 0.8), low end-placement probability score (EPP ≥ 3), and low strand-bias probability score (vcflib: SRP ≥ 3 & SAP ≥ 3). Sample genotypes

were required to have individual read coverage ≥ 4, and at least two reads and a minimum of 0.20 read observations supporting each allele.

Octoploid Genomic Diversity

We estimated octoploid genetic diversity metrics from a VCF file containing genotype calls for 45M filtered subgenomic SNPs and INDELs. We calculated the population-level diversity (π) and per-sample heterozygosity of sequence variants using a custom perl script. Chromosomal and genome-wide population nucleotide diversity estimates were calculated as the sum of pairwise diversity for all variant sites divided by total non-gap (N) genomic nucleotides. Sample heterozygosity was calculated as the sum of heterozygous variant sites divided by total non-gap (N) genomic nucleotides, and the fraction of total variant sites that were heterozygous.

Array Design and Genotyping

Unfiltered genomic variants were filtered to retain sites segregating in *F.* × *ananassa* cultivars. Cultivar variants were filtered to retain biallelic SNP sites with minor allele frequency ≥ 0.05, marker diversity ≥ 0.05, variant QUAL score > 20, and missing data < 15%. Variants requiring 2-probe assays (A/T or C/G) were excluded. 71-nt marker probe sequences were obtained by retrieving 35-nt SNP flanking sequences from the Camarosa v1.0 assembly. Markers containing ambiguous sequences (Ns), or identical probes were excluded. A set of 6.6M probes was submitted to Affymetrix for scoring and recommendation based on strand, kmer uniqueness, and buildability. Probes were scored for likelihood of binding interference or non-specific binding based on off-target variant counts in the binding region, the sum of minor allele frequencies of interfering variants, and counting off-target BLAST alignments (> 90% id, > 90% query length) in the genome. A final screening panel of 850,000 markers, including 16,000 iStraw probes (Bassil et al., 2015; Verma et al., 2016), was submitted to Affymetrix for constructing the 850K screening array. A panel of 384 octoploid strawberry genotypes was screened on the 850K array. Marker genotype clusters were scored using the Axiom Analysis Suite. Clustering was performed in "polyploid" mode with a marker call-rate threshold of 0.89. Samples were filtered with a dQC threshold of 0.82 and QC CR threshold of 93. A subset of 49,483 probes was selected from polymorphic, QC-passing markers ("PolyHighResolution", "NoMinorHomozygote", "OffTargetVariant") on the 850K screening array to populate the 50K production array. 5,809 LD-pruned ($r^2 < 0.50$) probes were pre-selected from the iStraw design, in addition to 47 probes associated with QTL for *Fusarium oxysporum* resistance and the Wasatch day neutral flowering locus (unpublished data). We assigned two markers per gene to a set of 2,878 genes located in expression networks related to flowering and fruit development (Kang et al., 2013; Hollender et al., 2014), or associated with R-gene domains. Non-overlapping 50 kb physical windows were parsed to select single markers containing the highest pairwise diversity in *F.* × *ananassa* genotypes. The remainder of the 50K array was populated by iteratively parsing 50 kb physical windows to select random QC-passing markers for uniform genomic distribution. 1,421 octoploid samples, including the Camarosa x Del Norte mapping population (n = 182), and PI552277 x PI612493 mapping population (n = 96), were genotyped on the 50K array and processed using the Axiom Analysis Suite using the same settings as the 850K dataset.

WGS Haplotype Linkage Mapping

We used 1.6M female parent informative variant calls (AB x AA), and 1.9M male parent informative variant calls (AA x AB) to generate haploblock markers for mapping genome-wide variant calls in the Camarosa x Del Norte bi-parental population. Camarosa-informative and Del Norte-informative variant calls were divided into parent-specific marker sets, then split by chromosome. For each chromosome, we performed pairwise linkage disequilibrium (LD)-clustering of markers (LD ≥ 0.96) in an initial seed region containing the first 250 chromosome variants, to identify marker groups called in the same phase relative to the unphased Camarosa v1.0 genome. The genotype phase of the LD cluster containing the largest number of markers were selected as the "seed phase". A 50 kb sliding window was initiated in the seed region and moved across the chromosome, identifying downstream markers in negative LD with the seed phase and reversing the repulsion genotype calls, in order to phase the chromosome into an artificial backcross. If a phasing window skipped a physical region larger than 100 kb without markers, reached a window with fewer than 25 markers, or the average downstream marker LD fell below 0.90, the chromosome was then fragmented at the breakpoint, seed-phase clustering repeated, and the sliding window reset for the subsequent downstream region. We used the software PhaseLD (Marand et al., 2017) with a 50-marker window and 1-marker step size to predict crossover events in the backcross-phased chromosome blocks. Window-specific variant calls lying between the predicted recombination breakpoints were used to generate consensus genotypes representing the haploblock. We mapped the haploblock markers using software ONEMAP (Margarido et al., 2007) and BatchMap (Schiffthaler et al., 2017) in outcross mode. ONEMAP was used to bin co-segregating markers, calculate pairwise recombination fractions, determine optimal LOD thresholds, then cluster markers into linkage groups based on a LOD threshold of 8, and maximum recombination fraction of 0.22. Marker orders and genetic distances were re-estimated in parallel with BatchMap using a window of 50 markers, window overlap of 30, and ripple window of 6 markers.

Array and DNA Capture Linkage Mapping

We performed single-marker linkage mapping of populations genotyped using the 50K array or DNA capture sequences because each contained fewer than 10,000 segregating markers per parent. Individual parent genotypes were mapped separately

using their respective informative marker subsets. We filtered markers based on a chi-square test for segregation distortion (p-value < 0.10), and excluded markers with >5% missing data. ONEMAP was used to bin co-segregating markers, calculate pairwise recombination fractions, determine optimal LOD thresholds, and cluster markers into linkage groups based on a LOD threshold of 8, and maximum recombination fraction of 0.22. Marker orders and genetic distances were re-estimated in parallel with BatchMap using a window of 20 markers, window overlap of 15, and ripple window of 6 markers.

GDR Microsatellite Alignment

We obtained the complete set of microsatellite primers developed in *Fragaria* species from the Rosaceae Genomics Database (GDR). Primers pairs were aligned to the Camarosa v1.0 genome in an orientation-aware manner using IPCRESS with a maximum amplicon fragment size of 500 bp and allowing 1 mismatch per primer.

Comparative Genomic Analysis

The subgenomes of *F.* × *ananassa* cultivar Camarosa and the genomes of *F. vesca* (Edger et al., 2017), *F. iinumae* (Edger et al., 2019a), and *Rubus occidentalis* (VanBuren et al., 2018) were compared with the MCScan Toolkit v1.1 (Wang et al., 2012) Syntenic gene pairs were identified using all-vs-all BLAST search, filtered to remove pairs with cscores below 0.7, and clustered into syntenic blocks in MCScan. Macrosynteny plots were constructed using the python version of MCScan (https://github.com/tanghaibao/jcvi/wiki/MCscan) (Python-version).

DATA AVAILABILITY STATEMENT

Sequence datasets used in this study were deposited in the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra), BioProject ID PRJNA578384. Supplemental data files, genetic maps, and Affymetrix SNP genotypes were deposited in the DRYAD database (DOI https://doi.org/10.25338/B8R31Q).

AUTHOR CONTRIBUTIONS

MH, PE, and SK contributed conception and design of the study. MH, AL, KB, and MF performed the statistical analysis. RF, CA, and GC provided genetic material and collected, submitted DNA samples. MH and SK wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This research is supported by grants to SK from the United States Department of Agriculture (http://dx.doi.org/10.13039/100000199) National Institute of Food and Agriculture (NIFA) Specialty Crops Research Initiative (#2017-51181-26833), and a United States Department of Agriculture NIFA postdoctoral fellowship (#2018-67012-27980).

ACKNOWLEDGMENTS

We thank our collaborators at Affymetrix for constructing the 50K and 850K octoploid strawberry SNP genotyping arrays.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01789/full#supplementary-material

REFERENCES

Abbott, R. J., Hegarty, M. J., Hiscock, S. J., and Brennan, A. C. (2010). Homoploid hybrid speciation in action. *Taxon* 59, 1375–1386. doi: 10.1002/tax.595005

Alix, K., Gérard, P. R., Schwarzacher, T., and Heslop-Harrison, J. S. (2017). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann. Bot.* 120, 183–194. doi: 10.1093/aob/mcx079

Arulsekar, S., and Bringhurst, R. S. (1981). Genetic model for the enzyme marker PGI in diploid California Fragaria vesca L: its variability and use in elucidating the mating system. *J. Hered.* 72, 117–120. doi: 10.1093/oxfordjournals.jhered.a109438

Barb, J. G., Bowers, J. E., Renaut, S., Rey, J. I., Knapp, S. J., Rieseberg, L. H., et al. (2014). Chromosomal evolution and patterns of introgression in Helianthus. *Genetics* 197, 969–979. doi: 10.1534/genetics.114.165548

Bassil, N. V., Davis, T. M., Zhang, H., Ficklin, S., Mittmann, M., Webster, T., et al. (2015). Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria× ananassa. *BMC Genom.* 16, 155. doi: 10.1186/s12864-015-1310-1

Bird, K. A., VanBuren, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi: 10.1111/nph.15256

Bringhurst, R. S. (1990). Cytogenetics and evolution in American Fragaria. *HortScience* 25, 879–881. doi: 10.21273/HORTSCI.25.8.879

Burke, J. M., Lai, Z., Salmaso, M., Nakazato, T., Tang, S., Heesacker, A., et al. (2004). Comparative mapping and rapid karyotypic evolution in the genus Helianthus. *Genetics* 167, 449–457. doi: 10.1534/genetics.167.1.449

Byrne, D., and Jelenkovic, G. (1976). Cytological diploidization in the cultivated octoploid strawberry Fragaria× ananassa. *Can. J. Genet. Cytol.* 18, 653–659. doi: 10.1139/g76-076

Chester, M., Gallagher, J. P., Symonds, V. V., da Silva, A. V. C., Mavrodiev, E. V., Leitch, A. R., et al. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). *Proc. Natl. Acad. Sci.* 109, 1176–1181. doi: 10.1073/pnas.1112041109

Cifuentes, M., Grandont, L., Moore, G., Chèvre, A. M., and Jenczewski, E. (2010). Genetic regulation of meiosis in polyploid species: new insights into an old question. *New Phytol.* 186, 29–36. doi: 10.1111/j.1469-8137.2009.03084.x

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836. doi: 10.1038/nrg1711

Dale, A., and Sjulin, T. M. (1990). Few cytoplasm contribute to North American strawberry cultivars. *HortScience* 25, 1341–1342. doi: 10.21273/HORTSCI.25.11.1341

Darrow, G. M. (1966). *The strawberry. History, breeding and physiology* (New York: Holt, Rinehart & Winston).

Davik, J., Sargent, D. J., Brurberg, M. B., Lien, S., Kent, M., and Alsheikh, M. (2015). A ddRAD based linkage map of the cultivated strawberry, Fragaria xananassa. *PLoS One* 10, e0137746. doi: 10.1371/journal.pone.0137746

de Bem Oliveira, I., Resende, M. F. R., Ferrão, L. F. V., Amadeu, R. R., Endelman, J. B., Kirst, M., et al. (2019). Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. *Genes Genom. Genet.* 9, 1189–1198. doi: 10.1534/g3.119.400059

Duchesne, A.-N. (1766). *Histoire naturelle des fraisiers contenant les vues d'économie réunies à la botanique, et suivie de remarques particulières sur plusieurs points qui ont rapport à l'histoire naturelle générale, par M. Duchesne fils*. chez Didot le jeune.

Edger, P. P., VanBuren, R., Colle, M., Poorten, T. J., Wai, C. M., Niederhuth, C. E., et al. (2017). Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (Fragaria vesca) with chromosome-scale contiguity. *Gigascience* 7, 1–7. gix124. doi: 10.1093/gigascience/gix124

Edger, P. P., McKain, M. R., Yocca, A. E., Knapp, S. J., Qiao, Q., and Zhang, T. (2019a). Phylogenetic analyses and Fragaria iinumae genome support octopoid strawberry origin. *Nat. Genet.*

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019b). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51, 541–547. doi: 10.1038/s41588-019-0356-4

Fedorova, N. J. (1946). Crossability and phylogenetic relations in the main European species of Fragaria. *Compil. Natl. Acad. Sci. USSR* 52, 545–547.

Finn, C. E., Retamales, J. B., Lobos, G. A., and Hancock, J. F. (2013). The Chilean strawberry (Fragaria chiloensis): over 1000 years of domestication. *HortScience* 48, 418–421. doi: 10.21273/HORTSCI.48.4.418

Folta, K. M., and Davis, T. M. (2006). Strawberry genes and genomics. *CRC Crit. Rev. Plant Sci.* 25, 399–415. doi: 10.1080/07352680600824831

Gaeta, R. T., and Pires, J. C. (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol.* 186, 18–28. doi: 10.1111/j.1469-8137.2009.03089.x

Gaeta, R. T., Pires, J. C., Iniguez-Luy, F., Leon, E., and Osborn, T. C. (2007). Genomic changes in resynthesized Brassica napus and their effect on gene expression and phenotype. *Plant Cell* 19, 3403–3417. doi: 10.1105/tpc.107.054346

Govindarajulu, R., Parks, M., Tennessen, J. A., Liston, A., and Ashman, T. (2015). Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *Am. J. Bot.* 102, 544–554. doi: 10.3732/ajb.1500026

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196, 966–971. doi: 10.1111/j.1469-8137.2012.04365.x

Hamilton, J. P., and Buell, C. R. (2012). Advances in plant genome sequencing. *Plant J.* 70, 177–190. doi: 10.1111/j.1365-313X.2012.04894.x

Hancock, J. F.Jr., and Bringhurst, R. S. (1979). Ecological differentiation in perennial, octoploid species of Fragaria. *Am. J. Bot.* 66, 367–375. doi: 10.1002/j.1537-2197.1979.tb06237.x

Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., et al. (2017). Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci.* 114, E9999–E10008. doi: 10.1073/pnas.1714380114

Hardigan, M. A., Poorten, T. J., Acharya, C. B., Cole, G. S., Hummer, K. E., Bassil, N., et al. (2018). Domestication of temperate and coastal hybrids with distinct ancestral gene selection in octoploid strawberry. *Plant Genome* 11. doi: 10.3835/plantgenome2018.07.0049

Hirakawa, H., Shirasawa, K., Kosugi, S., Tashiro, K., Nakayama, S., Yamada, M., et al. (2014). Dissection of the octoploid strawberry genome by deep sequencing of the genomes of Fragaria species. *DNA Res.* 21, 169–181. doi: 10.1093/dnares/dst049

Hirsch, C. N., and Buell, C. R. (2013). Tapping the promise of genomics in species with complex, nonmodel genomes. *Annu. Rev. Plant Biol.* 64, 89–110. doi: 10.1146/annurev-arplant-050312-120237

Hollender, C. A., Kang, C., Darwish, O., Geretz, A., Matthews, B. F., Slovin, J., et al. (2014). Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. *Plant Physiol.* 165, 1062–1075. doi: 10.1104/pp.114.237529

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res.* 19, 1068–1076. doi: 10.1101/gr.089516.108

Hufford, M. B., Xu, X., Van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R. A., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808. doi: 10.1038/ng.2309

Hughes, C. E., Govindarajulu, R., Robertson, A., Filer, D. L., Harris, S. A., and Bailey, C. D. (2007). Serendipitous backyard hybridization and the origin of crops. *Proc. Natl. Acad. Sci.* 104, 14389–14394. doi: 10.1073/pnas.0702193104

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97. doi: 10.1038/nature09916

Kang, C., Darwish, O., Geretz, A., Shahan, R., Alkharouf, N., and Liu, Z. (2013). Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry Fragaria vesca. *Plant Cell* 25, 1960–1978. doi: 10.1105/tpc.113.111732

Koskela, E. A., Sønsteby, A., Flachowsky, H., Heide, O. M., Hanke, M., Elomaa, P., et al. (2016). TERMINAL FLOWER 1 is a breeding target for a novel everbearing trait and tailored flowering responses in cultivated strawberry (Fragaria× ananassa Duch.). *Plant Biotechnol. J.* 14, 1852–1861. doi: 10.1111/pbi.12545

Kunihisa, M., Fukino, N., and Matsumoto, S. (2005). CAPS markers improved by cluster-specific amplification for identification of octoploid strawberry (Fragaria× ananassa Duch.) cultivars, and their disomic inheritance. *Theor. Appl. Genet.* 110, 1410–1418. doi: 10.1007/s00122-005-1956-1

Ladizinsky, G. (1985). Founder effect in crop-plant evolution. *Econ. Bot.* 39, 191–199. doi: 10.1007/BF02907844

Latta, R. G., Bekele, W. A., Wight, C. P., and Tinker, N. A. (2019). Comparative linkage mapping of diploid, tetraploid, and hexaploid Avena species suggests extensive chromosome rearrangement in ancestral diploids. *Sci. Rep.* 9, 1–12. doi: 10.1038/s41598-019-48639-7

Le Comber, S. C., Ainouche, M. L., Kovarik, A., and Leitch, A. R. (2010). Making a functional diploid: from polysomic to disomic inheritance. *New Phytol.* 186, 113–122. doi: 10.1111/j.1469-8137.2009.03117.x

Lee, H., and Schatz, M. C. (2012). Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score. *Bioinformatics* 28, 2097–2105. doi: 10.1093/bioinformatics/bts330

Leitch, A. R., and Leitch, I. J. (2008). Genomic plasticity and the diversity of polyploid plants. *Science* 320, 481–483. doi: 10.1126/science.1153585

Lerceteau-Köhler, E., Guerin, G., Laigret, F., and Denoyes-Rothan, B. (2003). Characterization of mixed disomic and polysomic inheritance in the octoploid strawberry (Fragaria× ananassa) using AFLP mapping. *Theor. Appl. Genet.* 107, 619–628. doi: 10.1007/s00122-003-1300-6

Liston, A., Cronn, R., and Ashman, T. (2014). Fragaria: a genus with deep historical roots and ripe for evolutionary and ecological insights. *Am. J. Bot.* 101, 1686–1699. doi: 10.3732/ajb.1400140

Luby, J. J., Hancock, J. F., and Ballington, J. R. (1992). Collection of native strawberry germplasm in the Pacific Northwest and Northern Rocky Mountains of the United States. *HortScience* 27, 12–17. doi: 10.21273/HORTSCI.27.1.12

Mandáková, T., Pouch, M., Brock, J. R., Al-Shehbaz, I. A., and Lysak, M. A. (2019). Origin and evolution of diploid and allopolyploid camelina genomes was accompanied by chromosome shattering. *Plant Cell.* 31, 2596-2612 tpc-00366. doi: 10.1105/tpc.19.00366

Marand, A. P., Jansky, S. H., Zhao, H., Leisner, C. P., Zhu, X., Zeng, Z., et al. (2017). Meiotic crossovers are associated with open chromatin and enriched with Stowaway transposons in potato. *Genome Biol.* 18, 203. doi: 10.1186/s13059-017-1326-8

Margarido, G. R. A., Souza, A. P., and Garcia, A. A. F. (2007). OneMap: software for genetic mapping in outcrossing species. *Hereditas* 144, 78–79. doi: 10.1111/j.2007.0018-0661.02000.x

McKain, M. R., Tang, H., McNeal, J. R., Ayyampalayam, S., Davis, J. I., dePamphilis, C. W., et al. (2016). A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biol. Evol.* 8, 1150–1164. doi: 10.1093/gbe/evw060

Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840. doi: 10.1038/nrg3605

Miller, A. J., and Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *Am. J. Bot.* 98, 1389–1414. doi: 10.3732/ajb.1000522

Njuguna, W., Liston, A., Cronn, R., Ashman, T.-L., and Bassil, N. (2013). Insights into phylogeny, sex function and age of Fragaria based on whole chloroplast genome sequencing. *Mol. Phylogenet. Evol.* 66, 17–29. doi: 10.1016/j.ympev.2012.08.026

Obbard, D. J., Harris, S. A., and Pannell, J. R. (2006). Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. *Hered. (Edinb)* 97, 296. doi: 10.1038/sj.hdy.6800862

Page, J. T., Huynh, M. D., Liechty, Z. S., Grupp, K., Stelly, D., Hulse, A. M., et al. (2013). Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *Genes Genom. Genet.* 3, 1809–1818. doi: 10.1534/g3.113.007229

Parkin, I. A. P., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., et al. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol.* 15, R77. doi: 10.1186/gb-2014-15-6-r77

Pincot, D. D. A., Poorten, T. J., Hardigan, M. A., Harshman, J. M., Acharya, C. B., Cole, G. S., et al. (2018). Genome-wide association mapping uncovers Fw1, a dominant gene conferring resistance to Fusarium wilt in strawberry. *Genes Genom. Genet.* 8, 1817–1828. doi: 10.1534/g3.118.200129

Qiao, Q., Xue, L., Wang, Q., Sun, H., Zhong, Y., Huang, J., et al. (2016). Comparative transcriptomics of strawberries (Fragaria spp.) provides insights into evolutionary patterns. *Front. Plant Sci.* 7, 1839. doi: 10.3389/fpls.2016.01839

Ramsey, J., and Schemske, D. W. (2002). Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* 33, 589–639. doi: 10.1146/annurev.ecolsys.33.010802.150437

Renny-Byfield, S., and Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop plants. *Am. J. Bot.* 101, 1711–1725. doi: 10.3732/ajb.1400119

Rho, I. R., Woo, J. G., Jeong, H. J., Jeon, H. Y., and Lee, C. (2012). Characteristics of F1 hybrids and inbred lines in octoploid strawberry (Fragaria× ananassa Duchesne). *Plant Breed.* 131, 550–554. doi: 10.1111/j.1439-0523.2012.01958.x

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annu. Rev. Ecol. Syst.* 28, 359–389. doi: 10.1146/annurev.ecolsys.28.1.359

Rousseau-Gueutin, M., Lerceteau-Köhler, E., Barrot, L., Sargent, D. J., Monfort, A., Simpson, D., et al. (2008). Comparative genetic mapping between octoploid and diploid Fragaria species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry. *Genetics* 179, 2045–2060. doi: 10.1534/genetics.107.083840

Salamone, I., Govindarajulu, R., Falk, S., Parks, M., Liston, A., and Ashman, T. (2013). Bioclimatic, ecological, and phenotypic intermediacy and high genetic admixture in a natural hybrid of octoploid strawberries. *Am. J. Bot.* 100, 939–950. doi: 10.3732/ajb.1200624

Sargent, D. J., Passey, T., Šurbanovski, N., Girona, E. L., Kuchta, P., Davik, J., et al. (2012). A microsatellite linkage map for the cultivated strawberry (Fragaria× ananassa) suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection. *Theor. Appl. Genet.* 124, 1229–1240. doi: 10.1007/s00122-011-1782-6

Sargent, D. J., Yang, Y., Šurbanovski, N., Bianco, L., Buti, M., Velasco, R., et al. (2016). HaploSNP affinities and linkage map positions illuminate subgenome composition in the octoploid, cultivated strawberry (Fragaria× ananassa). *Plant Sci.* 242, 140–150. doi: 10.1016/j.plantsci.2015.07.004

Schatz, M. C., Witkowski, J., and McCombie, W. R. (2012). Current challenges in *de novo* plant genome sequencing and assembly. *Genome Biol.* 13, 243. doi: 10.1186/gb-2012-13-4-243

Schiffthaler, B., Bernhardsson, C., Ingvarsson, P. K., and Street, N. R. (2017). BatchMap: a parallel implementation of the OneMap R package for fast computation of F1 linkage maps in outcrossing species. *PLoS One* 12, e0189256. doi: 10.1371/journal.pone.0189256

Schnable, J. C., Springer, N. M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci.* 108, 4069–4074. doi: 10.1073/pnas.1101368108

Senanayake, Y. D. A., and Bringhurst, R. S. (1967). Origin of Fragaria polyploids. I. Cytological analysis. *Am. J. Bot.* 54, 221–228. doi: 10.1002/j.1537-2197.1967.tb06912.x

Shaw, D. V. (1995). Comparison of ancestral and current-generation inbreeding in an experimental strawberry breeding population. *Theor. Appl. Genet.* 90, 237–241. doi: 10.1007/BF00222207

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., et al. (2011). The genome of woodland strawberry (Fragaria vesca). *Nat. Genet.* 43, 109. doi: 10.1038/ng.740

Sjulin, T. M., and Dale, A. (1987). Genetic diversity of North American strawberry cultivars. *J. Hortic. Sci. Biotechnol.* 112, 375–385.

Soltis, D. E., Segovia-Salcedo, M. C., Jordon-Thaden, I., Majure, L., Miles, N. M., Mavrodiev, E. V., et al. (2014a). Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytol.* 202, 1105–1117. doi: 10.1111/nph.12756

Soltis, P. S., Liu, X., Marchant, D. B., Visger, C. J., and Soltis, D. E. (2014b). Polyploidy and novelty: gottlieb's legacy. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130351. doi: 10.1098/rstb.2013.0351

Soltis, D. E., Visger, C. J., Marchant, D. B., and Soltis, P. S. (2016). Polyploidy: pitfalls and paths to a paradigm. *Am. J. Bot.* 103, 1146–1166. doi: 10.3732/ajb.1500501

Spangelo, L. P. S., Hsu, C. S., Fejer, S. O., and Watkins, R. (1971). Inbred line× tester analysis and the potential of inbreeding in strawberry breeding. *Can. J. Genet. Cytol.* 13, 460–469. doi: 10.1139/g71-070

Spigler, R. B., Lewers, K. S., Main, D. S., and Ashman, T. L. (2008). Genetic mapping of sex determination in a wild strawberry, Fragaria virginiana, reveals earliest form of sex chromosome. *Hered. (Edinb)* 101, 507. doi: 10.1038/hdy.2008.100

Spigler, R. B., Lewers, K. S., Johnson, A. L., and Ashman, T.-L. (2010). Comparative mapping revesals autosomal origin of sex chromosome in octoploid Fragaria virginiana. *J. Hered.* 101, S107–S117. doi: 10.1093/jhered/esq001

Staudt, G. (1999). *Systematics and geographic distribution of the American strawberry species: Taxonomic studies in the genus Fragaria (Rosaceae: Potentilleae)* (Berkeley, CA, USA: Univ of California Press).

Stegmeir, T. L., Finn, C. E., Warner, R. M., and Hancock, J. F. (2010). Performance of an elite strawberry population derived from wild germplasm of Fragaria chiloensis, and F. virginiana. *HortScience* 45, 1140–1145. doi: 10.21273/HORTSCI.45.8.1140

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid Cucurbita genomes. *Mol. Plant* 10, 1293–1306. doi: 10.1016/j.molp.2017.09.003

Tennessen, J. A., Govindarajulu, R., Ashman, T.-L., and Liston, A. (2014). Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biol. Evol.* 6, 3295–3313. doi: 10.1093/gbe/evu261

Tennessen, J. A., Wei, N., Straub, S. C. K., Govindarajulu, R., Liston, A., and Ashman, T.-L. (2018). Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biol.* 16, e2006062. doi: 10.1371/journal.pbio.2006062

Treangen, T. J., and Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13, 36. doi: 10.1038/nrg3117

Vallejo-Marín, M., Buggs, R. J. A., Cooley, A. M., and Puzey, J. R. (2015). Speciation by genome duplication: repeated origins and genomic composition of the recently formed allopolyploid species Mimulus peregrinus. *Evol. (N. Y.)* 69, 1487–1500. doi: 10.1111/evo.12678

van Dijk, T., Pagliarani, G., Pikunova, A., Noordijk, Y., Yilmaz-Temel, H., Meulenbroek, B., et al. (2014). Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map. *BMC Plant Biol.* 14, 55. doi: 10.1186/1471-2229-14-55

VanBuren, R., Wai, C. M., Colle, M., Wang, J., Sullivan, S., Bushakra, J. M., et al. (2018). A near complete, chromosome-scale assembly of the black raspberry (Rubus occidentalis) genome. *Gigascience* 7, giy094. doi: 10.1093/gigascience/giy094

VanBuren, R., Wai, C. M., Pardo, J., Yocca, A. E., Wang, X., Wang, H., et al. (2019). Exceptional subgenome stability and functional divergence in allotetraploid teff, the primary cereal crop in Ethiopia. *bioRxiv* 580720. doi: 10.1101/580720

Verma, S., Bassil, N. V., van de Weg, E., Harrison, R. J., Monfort, A., Hidalgo, J. M., et al. (2016). "Development and evaluation of the Axiom® IStraw35 384HT array for the allo-octoploid cultivated strawberry Fragaria× ananassa," in *VIII International Strawberry Symposium* (Belgium: ISHS Acta Horticulturae) 1156 75–82. doi: 10.17660/ActaHortic.2017.1156.10

Vining, K. J., Salinas, N., Tennessen, J. A., Zurn, J. D., Sargent, D. J., Hancock, J., et al. (2017). Genotyping-by-sequencing enables linkage mapping in three octoploid cultivated strawberry families. *PeerJ* 5, e3731. doi: 10.7717/peerj.3731

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–e49. doi: 10.1093/nar/gkr1293

Wendel, J. F., Jackson, S. A., Meyers, B. C., and Wing, R. A. (2016). Evolution of plant genome architecture. *Genome Biol.* 17, 37. doi: 10.1186/s13059-016-0908-1

Yakimowski, S. B., and Rieseberg, L. H. (2014). The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *Am. J. Bot.* 101, 1247–1258. doi: 10.3732/ajb.1400201

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 February 2020
doi: 10.3389/fpls.2020.00041

The Utility of Graph Clustering of 5S Ribosomal DNA Homoeologs in Plant Allopolyploids, Homoploid Hybrids, and Cryptic Introgressants

Sònia Garcia[1,2], *Jonathan F. Wendel*[3], *Natalia Borowska-Zuchowska*[4], *Malika Aïnouche*[5], *Alena Kuderova*[2] *and Ales Kovarik*[2]*

[1] Institut Botànic de Barcelona (IBB, CSIC - Ajuntament de Barcelona), Barcelona, Spain, [2] Department of Molecular Epigenetics, Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czechia, [3] Department of Ecology, Evolution & Organismal Biology, Iowa State University, Ames, IA, United States, [4] Faculty of Natural Sciences, Institute of Biology, Biotechnology and Environmental Protection, University of Silesia in Katowice, Katowice, Poland, [5] UMR CNRS 6553 ECOBIO, Université de Rennes 1, Rennes, France

OPEN ACCESS

Edited by:
Hanna Weiss-Schneeweiss,
University of Vienna, Austria

Reviewed by:
Tony Heitkam,
Dresden University of Technology, Germany
Jelena Mlinarec,
University of Zagreb, Croatia

***Correspondence:**
Ales Kovarik
kovarik@ibp.cz

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

Received: 17 October 2019
Accepted: 13 January 2020
Published: 10 February 2020

Citation:
Garcia S, Wendel JF, Borowska-Zuchowska N, Aïnouche M, Kuderova A and Kovarik A (2020) The Utility of Graph Clustering of 5S Ribosomal DNA Homoeologs in Plant Allopolyploids, Homoploid Hybrids, and Cryptic Introgressants. Front. Plant Sci. 11:41. doi: 10.3389/fpls.2020.00041

Introduction: Ribosomal DNA (rDNA) loci have been widely used for identification of allopolyploids and hybrids, although few of these studies employed high-throughput sequencing data. Here we use graph clustering implemented in the RepeatExplorer (RE) pipeline to analyze homoeologous 5S rDNA arrays at the genomic level searching for hybridogenic origin of species. Data were obtained from more than 80 plant species, including several well-defined allopolyploids and homoploid hybrids of different evolutionary ages and from widely dispersed taxonomic groups.

Results: (i) Diploids show simple circular-shaped graphs of their 5S rDNA clusters. In contrast, most allopolyploids and other interspecific hybrids exhibit more complex graphs composed of two or more interconnected loops representing intergenic spacers (IGS). (ii) There was a relationship between graph complexity and locus numbers. (iii) The sequences and lengths of the 5S rDNA units reconstituted *in silico* from k-mers were congruent with those experimentally determined. (iv) Three-genomic comparative cluster analysis of reads from allopolyploids and progenitor diploids allowed identification of homoeologous 5S rRNA gene families even in relatively ancient (c. 1 Myr) *Gossypium* and *Brachypodium* allopolyploids which already exhibit uniparental partial loss of rDNA repeats. (v) Finally, species harboring introgressed genomes exhibit exceptionally complex graph structures.

Conclusion: We found that the cluster graph shapes and graph parameters (k-mer coverage scores and connected component index) well-reflect the organization and intragenomic homogeneity of 5S rDNA repeats. We propose that the analysis of 5S rDNA cluster graphs computed by the RE pipeline together with the cytogenetic analysis might be a reliable approach for the determination of the hybrid or allopolyploid plant species parentage and may also be useful for detecting historical introgression events.

Keywords: 5S rRNA genes, allopolyploidy, hybridization, evolution, graph structure clustering, high-throughput sequencing, repeatome

INTRODUCTION

It is well-established that all modern plant species have experienced at least one whole genome duplication and that many also have interspecific hybridization and recurrent introgression in their recent history (Wendel, 2015; Alix et al., 2017; Nieto Feliner et al., 2017; Van De Peer et al., 2017). Documenting recent allopolyploidy is relatively straightforward using cytogenetic analysis and genome size measurements, since allopolyploids have twice as many chromosomes (or more) as the parental species. Identification of homoploid hybrids is more difficult since the chromosome number and genome size are often similar to that of the parental species (Nieto Feliner et al., 2017). Evolutionary young allopolyploids and other hybrids tend to retain fixed polymorphisms at protein-coding and non-coding loci. These duplicated loci are called homoeologs (Glover et al., 2016) and are useful for documenting parentage as well as understanding the dynamics of polyploid genomes (Yoo et al., 2014; Wendel, 2015; Bourke et al., 2018). Older allopolyploids can have experienced episodes of intergenomic translocation, dysploidy, gene conversion, localized deletions, and other genetic events, leading eventually to diploidization of the genome (Wendel, 2015; Wendel et al., 2018).

Ribosomal RNA genes encoding 5S, 5.8S, 18S, and 26S ribosomal RNA are ubiquitous in plants and are organized into arrays containing hundreds to thousands of tandem repeats at one or more genomic loci (Hemleben and Zentgraf, 1994; Nieto Feliner and Rosselló, 2007; Roa and Guerra, 2012; Garcia et al., 2017). Due to their rapidly diverging intergenic (IGS) and internally transcribed spacers (ITS), rDNA loci have become popular taxonomic markers revealing allopolyploidy and other interspecific hybridization in many plant and animal systems (Alvarez and Wendel, 2003; Poczai and Hyvonen, 2010; Nieto Feliner and Rossello, 2012). The internet searches using *ITS* and *allopolyploidy* resulted in more than 650 hits in Web of Science for just 2019. Most studies have used classical single clone sequencing approaches whereas high-throughput data have only rarely been employed and are limited to the 35S (45S) rDNA (Matyasek et al., 2012; West et al., 2014; Boutte et al., 2016). The analysis of 5S rDNA is also informative and has been successfully used in many phylogenetic studies (Cronn et al., 1996; Fulnecek et al., 2002; Baum et al., 2004; Besendorfer et al., 2005; Volkov et al., 2007; Baum et al., 2012; Jang et al., 2016). Its analysis is complementary to that of 35S since both loci usually occur separately on chromosomes (Roa and Guerra, 2012; Garcia et al., 2017). The 5S rDNA is usually located in one chromosome pair in most angiosperms and can occupy variable chromosome positions. It is organized in tandemly arranged units comprising hundreds to tens of thousands of copies. Each unit is composed of a conserved c. 120 bp coding region separated by a variable intergenic spacer (Sastri et al., 1992). Similar to 35S loci, 5S rDNA loci undergo concerted evolution, a process maintaining high homogeneity within and often between arrays (Dover, 1982; Elder and Turner, 1995; Parks et al., 2019). Such a process may rapidly homogenize rDNA sequences and induce copy number variation (Bughio and Maggert, 2019) blurring their hybridogenic signatures in allopolyploids (Wendel et al., 1995a; Volkov et al., 1999; Muir et al., 2001; Matyasek et al., 2003). In contrast to 35S rDNA, the 5S rDNA loci appear to be less sensitive to homogenization in some allopolyploids (Fulnecek et al., 2002; Pedrosa-Harand et al., 2006; Weiss-Schneeweiss et al., 2008; Garcia et al., 2017), retaining diagnostic capacity with respect to their parental origin.

The clustering algorithm employed by RepeatExplorer (RE) (Novak et al., 2010; Novak et al., 2013) has become a tool of choice for the analysis of chromosome composition and genome evolution (Renny-Byfield et al., 2012; Weiss-Schneeweiss et al., 2015; Ribeiro et al., 2017; Mlinarec et al., 2019; Peska et al., 2019). The phylogenetic signal of the repeatome has also been exploited in phylogenetic studies (Dodsworth et al., 2015; Dodsworth et al., 2016; Grover et al., 2019; Vitales et al., 2019). The analysis of genomes by RE is based on an all-to-all comparison of sequence reads revealing their similarities. Subsequently, the data are used to build clusters of overlapping reads representing different repetitive elements. The TAREAN tool, recently introduced into the RepeatExplorer2 pipeline, allows repeat identification and reconstruction of tandem repeats solely from sequence reads (Novak et al., 2017). Graph theory and connected component methods lying in the heart of the computation algorithm produce graph structures reflecting genomic organization of repeats. Typically, tandem repeats exhibit circular (ring) shape topologies are characterized by high values of circularity parameters. Although the RepeatExplorer2/TAREAN tool was initially developed for identification of non-coding satellites, 5S rDNA can also be analyzed with the program. This is because 5S rDNA shows many features of satellite repeats: (I) its highly homogeneous units are tandemly arranged in a head to tail orientation, (II) it appears in high copy number, allowing analyses even at low coverage, and (III) the size of 5S rDNA monomers (c. 200–1,000 bp) (Sastri et al., 1992; Fulnecek et al., 2006) falls within the range defined for satellite DNA, allowing circularization of chains of overlapping reads.

In this study we investigated the 5S rDNA genomic organization and homogeneity in more than 80 plant diploids and polyploids, exploiting high-throughput reads available from read archives in public genomic databases and also *de novo* sequenced by us. Particular attention was paid on hybrid systems with well-defined evolutionary histories, both eudicots and monocots: (i) *Brachypodium hybridum* (Poaceae), *Brassica carinata* (Ethiopian mustard, Brassicaceae), *Chenopodium quinoa* (quinoa, Amaranthaceae), *Gossypium hirsutum* (cotton, Malvaceae), and *Nicotiana rustica* (Aztec tobacco, Solanaceae) allotetraploids. (ii) *Spartina* × *townsendii* (cordgrass, Poaceae) homoploid hybrid. (iii) Species with frequent introgression events included *Gossypium gossypioides* and *Thinopyrum intermedium* (intermediate wheatgrass, Poaceae). We used the RepeatExplorer2/TAREAN clustering pipeline and cluster graph computation methods to address the following questions: (1) What is the relationship between graph complexity and intragenomic heterogeneity of 5S rDNA repeats? (2) Can the full-length 5S rDNA units be assembled from short sequence reads? (3) Can allopolyploids and other interspecific hybrids be distinguished from their progenitors based on cluster graph

topologies? We show that cluster graphs may represent a convenient and simple-use approach for identification of interspecific hybrids from high-throughput sequencing data.

RESULTS

Relationship Between Cluster Graph Topology and Intragenomic Diversity of 5S rDNA

All of the 5S RNA gene families analyzed share a conserved c. 120 bp coding region while they differ in their intergenic spacers. We tested the hypothesis that the graph topologies of 5S rDNA clusters reflect the divergence and number of homoeologous gene families in allopolyploid genomes. Under this hypothesis, diploid species with a single gene family (and a locus) would generate a simple circular graph while allopolyploid and other hybrid genomes with multiple gene families (and loci) would display more complex graphs. To test this hypothesis we examined 5S rDNA cluster graph topologies in 84 plant species (**Supplementary Table S1**). Examples of cluster graph analyses in *Gossypium, Brachypodium,* and *Spartina* hybrid systems are shown in **Figure 1**. We have chosen these species because the parental genome donors, number of 5S rDNA loci, and approximate ages are known (**Table 1**). In the graphs, each vertex represents a sequence read and nodes connecting vertices depict sequence similarity between the reads. Simple circular 5S graphs with no or little deviation from regular circularity (referred as type 1 graphs) were observed in *G. arboreum, G. raimondii, B. distachyon,* and *B. stacei* diploid species. Except for the hexaploid *S. alterniflora*, the 5S graph topologies were more complex in polyploids (**Figure 1**). Specifically, two or more loops (rings) interconnected by a junction region (composed of 5S coding sequences) could be recognized. These complex structures are referred as type 2 graphs. Both loops were composed of vertices depicted in grey in **Figure 1** representing variable IGS regions. The total *k-mer coverage* scores (mean of cluster homogeneity) were high in diploid species while they were lower in the allotetraploids (**Table 1**). The *connected component index* C (mean of graph circularity) was uniformly high across the species. The read richness varied between the loops. For

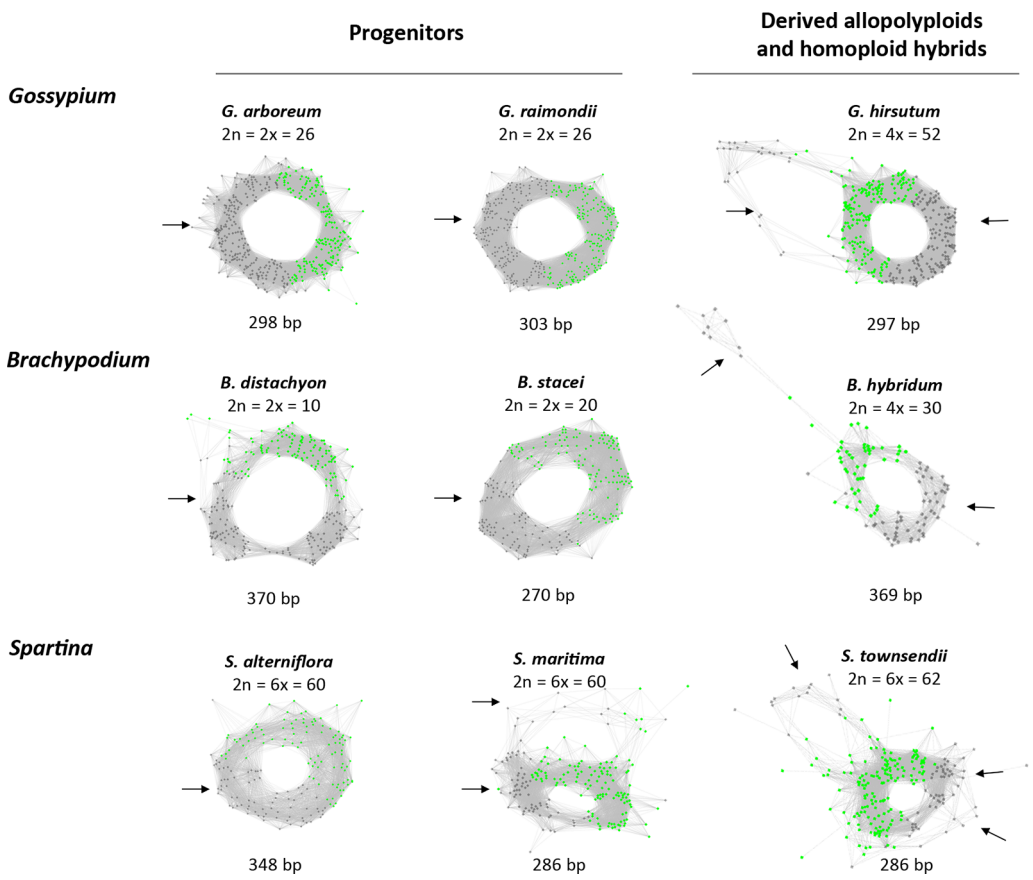

FIGURE 1 | The 5S rDNA sequence reads organized in graph structures from the RepeatExplorer2 graphical output. Single reads are represented by vertices (nodes) and their sequence overlaps by edges. The 5S coding sequences and intergenic spacers are highlighted in green and grey vertices, respectively. Note, regular circular structures (referred to type 1) in most progenitors, and complex structures (referred to type 2) in derived allopolyploid and homoploid hybrids. Arrows indicate one or several intergenic spacers.

TABLE 1 | Cytogenetic characteristics of 5S rDNA loci and cluster graph parameters in allopolyploid and homoploid species and their progenitors.

	Ploidy level	N. loci/ 1C	N. reads in the cluster	Genome proportion (%)	Repeat size (bp)	k-mer coverage	Connected component index C	Graph shape (type)
G. hirsutum[1]	4x	2	351	0.170	297	0.660	0.997	2
G. arboreum	2x	1	418	0.210	303	1.000	1.000	1
G. raimondii	2x	1	386	0.190	298	0.974	0.974	1
B. hybridum[2]	4x	2	109	0.054	303	0.680	0.982	2
B. distachyon	2x	1	266	0.130	370	0.810	0.981	1
B. stacei	2x	1	239	0.120	270	0.950	1.000	1
S. × townsendii[3]	6x	n.d.	225	0.044	286	0.593	0.947	2
S. alterniflora	6x	n.d.	123	0.031	348	0.871	0.976	1
S. maritima	6x	n.d.	210	0.053	286	0.628	0.943	2

[1]G. hirsutum *(2n = 4x = 52, AADD genome composition) is thought to originate from hybridization of species similar to modern* G. raimondii *(2n = 2x = 26, D genome donor) and* G. arboreum *(2n = 2x = 26, A genome donor).*

[2]B. hybridum *(2n = 4x = 30) is a natural allotetraploid with divergent subgenomes derived from diploid species similar to modern* B. distachyon *(2n = 2x = 10) and* B. stacei *(2n = 2x = 20).*

[3]S. × townsendii *(2n = 6x = 62) is a natural homoploid hybrid derived from* S. alterniflora *(2n = 6x = 62) and* S. maritima *(2n = 6x = 60).*

example, in *Brachypodium hybridum*, the right loop contained far more reads than the left loop.

In the whole dataset (**Supplementary Table S1**) typical circular graph shapes of 5S rDNA clusters (type 1) were obtained in 81 (96%) species. The *connected component index* C parameter values (reported by a TAREAN) were high, ranging from 0.684 to 1.00 (average 0.959, s.d. 0.0599). In three species (4%), no circularization of 5S graphs clusters was obtained. **Figure 2A** shows the frequency of individual cluster types in allopolyploid and diploid species. The majority (87%) of diploid species showed type 1 structures while most (79%) allopolyploids displayed type 2 graphs (one-way ANOVA, $F = 75.507$, $p < 0.001$, **Supplementary Table S2**). About 95% single locus karyotypes displayed type 1 graphs while most (94%) karyotypes with two or more loci had type 2 structures (**Figure 2B**). There was a relationship between locus number and graph complexity (one-way ANOVA, $F = 24.259$, $p < 0.001$) with type 1 structures showing significantly lower locus numbers. The sequence homogeneity within each 5S rDNA cluster was estimated based on *total k-mer coverage* score reported by TAREAN and ranged 0.416–1.00 (average 0.760, s.d. 0.1392). The k-mer coverage values were significantly higher (one-way ANOVA, $F = 200.363$, $p < 0.001$, **Supplementary Table S2**) in the diploids compared to the allotetraploids and other hybrids (**Supplementary Figure S1**).

Tracking the Origin of 5S rDNA Families in *Gossypium*, *Brachypodium*, and *Spartina* Allopolyploids and Homoploid Hybrids by Comparative Cluster Analysis

Next, we investigated whether the homoeologous 5S genes can be visualized in cluster graph layouts and whether homoeologous gene families occur in assembled contigs. To address these questions we carried out a comparative three-genomic analysis (**Figure 3**) where inputs for clustering include reads from hybrids (allopolyploids) and their putative progenitor species. The overall cluster shapes were similar as in single genome analyses (**Figure 1**) indicating that the progenitor 5S rDNA sequences overlap with those of the derived hybrids and allopolyploids confirming, thus, their putative origin. Reads derived from 5S coding sequences (in green) were found in the junction region connecting both loops (**Figures 3A, D, G**). In **Figures 3B, E, H** reads originating from each progenitor (red and yellow) and hybrid (blue) genomes are labeled by different colors showing the 5S rDNA variants origin:

1. **Figure 3B** shows three-genomic cluster graph structures in *Gossypium hirsutum* and its progenitors. *G. hirsutum* is a 1–2 M years-old allotetraploid composed of subgenomes close to *G. arboreum* (A genome species) and *G. raimondii* (D genome species) (Wendel, 1989). The blue colored reads from the *G. hirsutum* allopolyploid were found in all parts of the graph—both in the junction region and loops; the red color reads from the *G. arboreum* diploid progenitor were located in the right loop and the junction region; the yellow colored reads originating from the other parental species, G. *raimondii*, were located in the left loop and the junction region. Similar cluster graph shapes were observed in remaining four *Gossypium* allotetraploids (*G. barbandense*, G. *mustelinum*, *G. darwinii*, and *G. tomentosum*, see **Supplementary Figure S2**), all having a similar AADD composition of the genome. In order to determine the identity of loop structures in the graphs we carried out a phylogenetic analysis of assembled 5S rDNA contigs (**Figure 3C** and **Supplementary Figure S2**). On the trees, sequences of both progenitors were well resolved forming separate branches, consistent with sequence divergence. The contigs from the *G. hirsutum* cluster grouped within the *G. arboreum* and *G. raimondii* branches, respectively.
2. **Figure 3E** shows three-genomic cluster graph structures in *Brachypodium hybridum* and its progenitors. *B. hybridum* is a c. 1 M years-old allotetraploid composed of subgenomes close to *B. distachyon* and *B. stacei* (Catalán et al., 2012). The comparative cluster graph displayed two loops composed of reads either from the *B. distachyon* (in red) or *B. stacei* (yellow). In contrast, reads from *B. hybridum* (blue) were shared between both loops. However, there were much less *B. hybridum* reads in the *B. stacei* loop compared to that of the

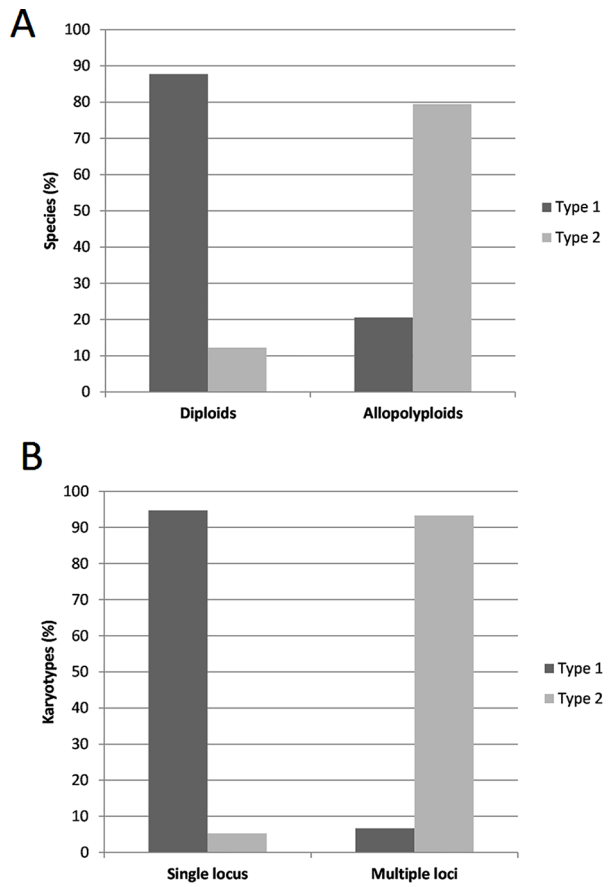

FIGURE 2 | (A) Distribution of 5S rDNA type 1 and type 2 cluster graph structures between diploid (N = 50) and allopolyploid (including homoploid hybrids) (N = 32) species from the 84 species data set in this study. **(B)** Occurrence of type 1 and type 2 cluster graph structures in karyotypes with single (N = 38) and multiple (N = 30) 5S rDNA loci (data on 5S rDNA loci number only available for 68 species analyzed). The source data and the basic statistics are in **Supplementary Table S2**.

B. distachyon. No B. *stacei* homoeologs were found among the contigs (**Figure 3F**).

3. **Figure 3H** shows three-genomic cluster graph structures in *Spartina* × *townsendii* and its progenitors. *Spartina* × *townsendii* is a less than 150 years-old homoploid hybrid composed of subgenomes inherited from *S. alterniflora* and *S. martima* hexaploids (Ainouche et al., 2004). The bottom read-rich circle contained sequences from the *S. maritima* (yellow) parent and *S.* × *townsendii* (blue); the upper read-poor circle was mostly formed by reads from *S. alterniflora* (red) parent and a few reads from *S.* × *townsendii* (blue). The junction region contained 5S genic sequences plus part of the IGS (grey) indicating short conserved sequences flanking the genic region. Another region of homoeologous genes similarity seems to exist in the middle of IGS indicated by interconnected reads from all three genomes (arrow). On the tree (**Figure 3I**), sequences from both progenitors were well resolved forming separate branches. However, all the assembled contigs grouped exclusively with the *S. maritima* branch.

Additional examples of 5S rDNA cluster analyses are shown in the **Supplementary Figure S3** comprising the well-known allopolyploids, *Brassica carinata* (4x), *Chenopodium quinoa* (4x), and *Nicotiana rustica* (4x). All these species harbored complex type 2 graphs, in which at least one (*Chenopodium*) or both progenitors (*Brassica* and *Nicotiana*) could be identified.

Comparative Analysis Reveals Genetic Complexity in Species With Cryptic Introgression Histories

1. *Gossypium gossypioides* is a new world (D-genome) diploid species known to have experienced several rounds of introgressive hybridization from old world species (A genome) (Cronn et al., 2003). Its comparative cluster 5S rDNA graph of the three *Gossypium* species analyzed (**Figures 4A, B**) showed three loops where the *G. gossipioides* reads formed a unique loop (blue) that did not overlap with either A (red) or D (yellow) genome loops. Except for the genic junction region no significant interconnecting edges between the three genomes were visualized.
2. The intermediate wheatgrass *Thinopyrum intermedium* (Poaceae) is a hexaploid species experiencing multiple introgression events, potentially including genome parts from several species. We therefore included candidate *Aegilops tauschii* and *Hordeum vulgare* progenitor species in our comparative analysis of 5S rDNA (**Figures 4C, D**) which had been suggested as potential genome contributors for *T. intermedium* (Tang et al., 2000; Mahelka et al., 2011). At least four loops could be recognized on the cluster graph. Two loops contained shared reads from *T. intermedium* (blue) and *A. tauschii* (yellow). In addition, there was a prominent read-rich *Thinopyrum*-specific loop (**Figure 4D**, all-blue loop) that did not contain reads from other genomes and may originate from *Dasypirium* (Mahelka et al., 2013) for which read archives were unavailable. No *T. intermedium* reads were present in the *H. vulgare* loop (red).

Quantification of 5S and 35S rDNA Homoeologs in *Gossypium* and *Brachypodium* Allotetraploids From High-Throughput Sequencing Data

In the cluster graphs of the hybrid species (**Figures 1** and **3**) we often observed differences in read-richness between both loops suggesting a skewed representation of homoeologous 5S rDNA variants. To validate this assumption we quantified homoeologous 5S rDNAs by mapping of Illumina reads to the reference sequences of 5S rDNA units (**Supplementary Figure S4**). In five *Gossypium* allopolyploids analyzed (**Supplementary Figure S4A**), the 5S rDNA homoeologs were slightly skewed toward the A genome units. This

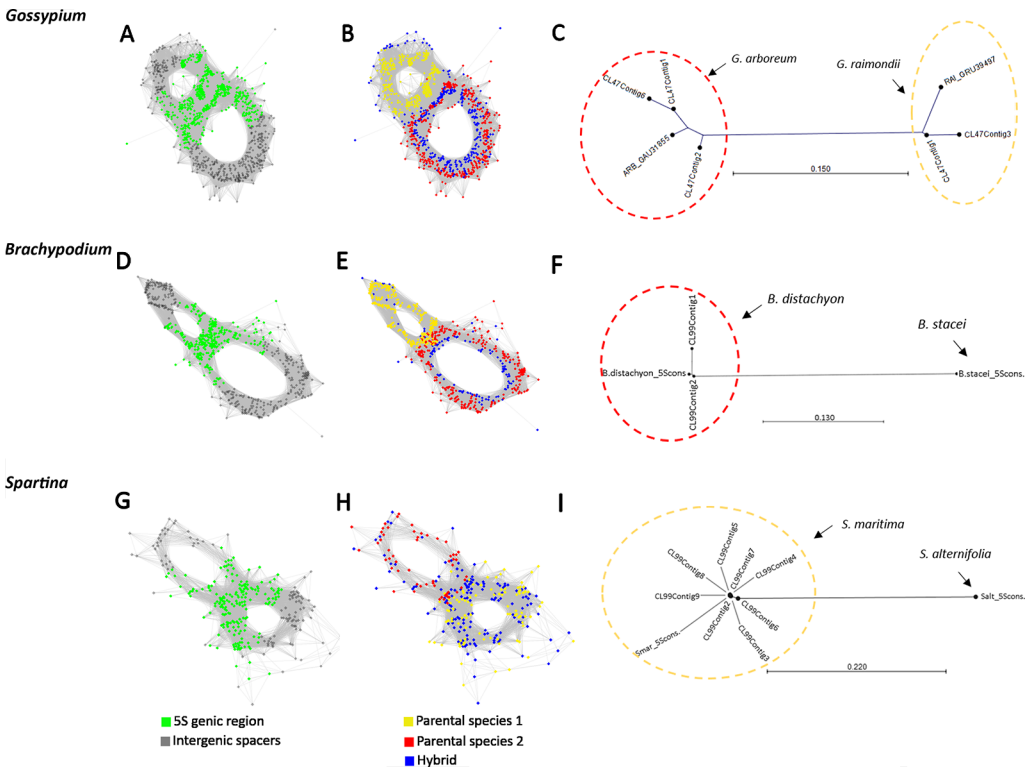

FIGURE 3 | RepeatExplorer2 graphical output of the three-genomic comparative 5S rDNA analyses involving progenitor species and derived hybrids. **(A, D, G)** Graphs with highlighted 5S rDNA genic regions in green. **(B, E, H)** Cluster graphs with annotated reads origin: yellow vertices represent reads of one of the parental species (*G. raimondii* in **B**, *B. stacei* in **E**, and *S. maritima* in **H**); red vertices represent reads of the other putative parental species (*G. arboreum* in **B**, *B. distachyon* in **E**, and *S. alterniflora* in **H**); blue vertices represent reads from the hybrids and allotetraploids (*G. hirsutum* in **B**, *B. hybridum* in **E**, and *S.* × *townsendii* in **H**). **(C, F, I)** Neighbor joining phylogenetic trees constructed from aligned contigs from hybrids and progenitor 5S rDNA sequences.

clearly contrasts with 35S rDNA (**Supplementary Figure S4B**), where all but one allotetraploids contained far fewer A-genome than D-genome ITS1 types, except of *Gossypium mustelinum* in which the homoelogous ratio was inversed. A similar analysis carried out in *Brachypodium hybridum* showed a higher representation of *B. distachyon* homoeologs (in this case, both 35S and 5S rDNA) than those of B. *stacei* (**Supplementary Figure S4C**).

Reconstitution of 5S rDNA Units and Gene Copy Number in *Gossypium* Allotetraploids

The unit length is an important characteristic of rDNA arrays. We compared the lengths of *in silico* assembled *Gossypium* units with those previously determined by conventional cloning and Sanger sequencing (**Table 2**). RepeatExplorer2/TAREAN generates consensus sequences of 5S rDNA units from the decomposition of read sequences into k-mers (Novak et al., 2017). The lengths of 5S rDNA units determined by cloning ranged from 295–303 bp while those calculated from sequence data by bioinformatics tools ranged from 265–303 bp. In general, there was congruence between both methods. In contrast, the copy number variation between species was extremely high (up to 10 fold), confirming previous findings (Cronn et al., 1996). In some cases, copy numbers determined by computation methods differed by more than five-fold from those of slot blot hybridization experiments (Cronn et al., 1996).

Southern Blot Hybridization Analysis of 5S rDNA in *Spartina*

Previous *in silico* analyses showed highly skewed 5S rDNA homoeologs in *Spartina* × *townsendii* hexaploid toward the *S. maritima* genome. In order to confirm this result, we carried out southern blot hybridization using genomic DNA from *S.* × *townsendii* (*6x*), the derived *S. anglica* allododecaploid (12x), and the progenitors of both species, *S. maritima* and *S. alterniflora* (**Figure 5**). Genomic DNA was digested with *Bam*HI which has a conserved site in the angiosperm 5S rDNA units (Röser et al., 2001) (**Figure 5A**). The 5S rDNA probe generated ladders of bands, expected from a tandemly arranged sequence as the 5S rRNA genes. The probe hybridized strongly to the *S. maritima* DNA while the hybridization to *S. alterniflora* was relatively weak (**Figure 5B**). The *S. maritima* oligomers were slightly shorter than those of *S. alterniflora* consistent with shorter length of the *S. maritima* units (**Figure 1**). In both *S.* × *townsendii* and

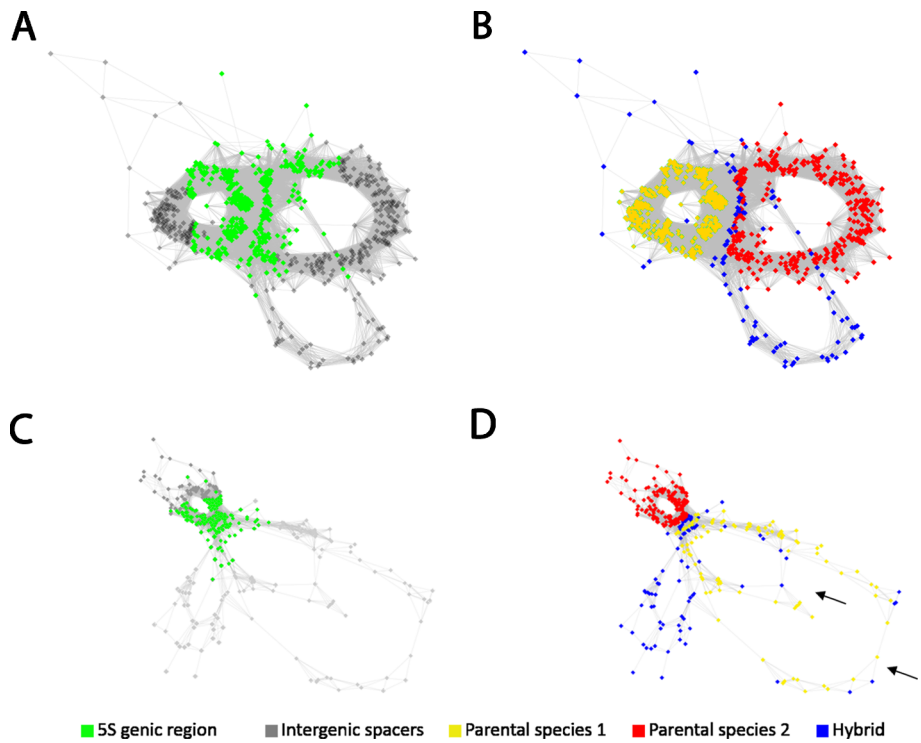

FIGURE 4 | RepeatExplorer2 graphical output of the comparative analysis of the 5S rDNA clustering in species of hybrid origin and with complex evolutionary histories, in which introgressive hybridization may have been involved. Cluster graphs with highlighted 5S rDNA genic region in green **(A, C)** and annotated reads origin in **(B, D)** in *Gossypium gossypioides* and *Thinopyrum intermedium*, respectively. The comparative analysis in **(B)** mixing reads of the putative parental genome donors, *G. arboreum* as the **A** genome donor (red), *G. raimondii* as the **D** genome donor (yellow), and *G. gossypioides* as the hybrid (blue) shows that only few reads of the putative hybrid are placed in the **A** or **D** genome loops. The comparative analysis in **(D)** shows that the reads corresponding to the hexaploid *Thinopyrum intermedium* (blue) are partially shared with the reads of one of the putative parental genome donors, *Aegilops tauschii* (yellow) while there are no coincidences with the reads of the other putative parental genome donor, *Hordeum vulgare* (red). Arrows indicate two spacers of different sizes stemming from the *Aegilops* parental genome donor.

S. anglica oligomeric bands derived from both parents were visible indicating additivity.

DISCUSSION

Here we studied the organization and evolution of 5S rDNA in 84 plant genomes. We found that the clustering analysis of high-throughput Illumina reads by RepeatExplorer2/TAREAN provides a comprehensive view about 5S rDNA origin and organization, corroborating classical cytogenetic and molecular studies. The 5S rDNA cluster graphs are typically circular, showing high values of circularity parameters underlining regular tandem arrangement of these genes. Graphs displayed no or little node irregularities (discussed further below) consistent with a high homogeneity of arrays and confirming the model of concerted evolution (Dover, 1982; Kellogg and Appels, 1995). The sequence and size of *in silico* reconstituted 5S rDNA units were congruent with those obtained from cloning and Sanger sequencing. Below we discuss a value-added information obtained from cluster analyses that is not obtainable (or with difficulty) by conventional molecular and cytogenetic analyses.

Dynamism of rDNA Loci in Hybrid Genomes

We investigated rDNAs in two relatively ancient (c.1 Myr) allopolyploid systems (*Gossypium* and *Brachypodium*) which already show substantial loss of 35S rDNA homoeologous units (Wendel et al., 1995a; Borowska-Zuchowska and Hasterok, 2017). *Gossypium* allotetraploids were represented by five species with a typical AADD genome composition originating from common diploid ancestors closely related to *G. arboreum* (A genome) and *G. raimondii* (D genome). In these allotetraploids, previous cloning analyses identified both homoeologous 5S rDNA sequences in *G. hirsutum* and *G. mustelinum* but not in *G. barbadense* and *G. tomentosum,* where only the A genome sequences were recovered (Cronn et al., 1996). However, comparative graph clustering of the 5S rDNA revealed both A and D genome homoeologs in these species, with a dominance of the A-genome units. Also, with respect to the 35S rDNA, the D-genome type of ITS in

TABLE 2 | 5S rDNA unit lengths and copy number in *Gossypium* allotetraploids and diploids.

	SRA Identification	Unit length[1]		Copy number[2]	
		High-throughput data[4]	Cloning	High-throughput data[5]	Slot blot hybridization
Gossypium mustelinum	SRR769542	265	301–303	14,015	21,845
Gossypium hirsutum[3]	SRR768357	297	295–279	18,412	11,190
	ERR1449079	265	295–279	14,903	11,190
Gossypium barbadense	SRR8624709	265	296–298	18,157	23,515
Gosspium raimondii	ERR1449077	303	301–303	11,061	4,730
Gossypium arboreum	SRR1216970	298	297–298	23,691	7,550
Gossypium thurberi	SRR8076131	302	301–302	10,607	2,070
Gossypium darwinii	SRX5347640	273	n.d.[6]	24,276	n.d.
Gossypium tomentosum	SRR8815512	259	296–297	38,691	22,290
Gossypium gossypioidies	SRR8136267	297	301–303	3,292	1,145
Gossypium herbaceum	SRR617255	265	297–298	7,819	3,415
Gossypium davidsonii	SRR8136261	302	301–303	19,909	10,280

[1]*Genic and intergenic region (bp). Data from sequencing of clones are from Cronn et al. (1996).*
[2]*Copy number in the somatic cell genome (2C). Slot blot hybridization results are from Cronn et al. (1996).*
[3]*Data are from two different accessions.*
[4]*K-mer assembly.*
[5]*Calculated from the genome proportion.*
[6]*n.d.—not determined.*

G. mustelinum sequence was barely detectable using Southern blot hybridization (Wendel et al., 1995a), but high-throughput sequencing recovers similar sequences at a frequency of about 15% (**Supplementary Figure S4B**). A similar example of skewed homoelog ratios is represented by *B. hybridum* and *S.* × *townsendii*, where one loop contained far more reads than the other in the 5S rDNA cluster graphs. Indeed, Southern blot analysis confirmed one strong and one weak 5S rDNA family in the *S.* × *townsendii* homoploid hybrid and *S. anglica*. Skewed gene ratios exist already in the progenitor genomes based on read abundance in the graph loops and differential intensity of Southern hybridization signals. A strong repeat-rich locus likely occurs in the *S. maritima* parent while a weaker locus may be present in the *S. alterniflora* parent. These examples demonstrate a higher sensitivity of a cluster graph-based approach over the *de novo* assembly or PCR-cloning approaches, where various technical biases may occur (Lunerova et al., 2017).

Evaluation of the Graph-Based Method for Identification of Allopolyploids and Hybrids

Diploid genomes show simple circular structures of 5S rDNA cluster graphs referred to type 1 (**Figure 2A**). In contrast, allopolyploids and homoploid hybrids display more complex graph structures (type 2) in which divergent gene families are visualized as distinct loops. In addition, there was a good correlation between cluster graph complexity and number of 5S rDNA loci (**Figure 2B**). These observations are consistent with a general view that most diploids carry a single 5S rDNA locus (and a single gene family) while allopolyploids tend to maintain multiple loci (and multiple 5S rDNA families) (**Table 1**) and (Roa and Guerra, 2015; Garcia et al., 2017). Thus, a simple visual inspection of the 5S rDNA cluster topology appears to be informative with respect to the putative hybridogenic origin of a species. Among the computation parameters, the k-mer coverage seems to reflect the intragenomic homogeneity—low k-mer scores associate with complex graph shapes and multiple gene families, while high k-mer scores associate with simple circular structures and single gene families. Thus, the k-mer coverage may be taken as a semi-quantitative parameter of 5S rDNA intragenomic homogeneity, although more studies are needed to validate the relationship.

One of the advantages of the clustering-based method is that it may provide initial information about the 5S rDNA homoeologs without prior knowledge of progenitor genomes based on the graph complexity. Certainly, the origin of 5S rDNA families in a hybrid is indicated by comparative clustering requiring sequences from candidate progenitor genomes. We observed similar graph complexities for the 35S rDNA encoding 18S-5.8S-26 rRNA genes (**Supplementary Figure S5**) suggesting that these clusters (particularly, 3' 26S region and the IGS) may be equally informative as that of 5S rDNA. These cases pose additional opportunities for studying the recombination dynamics of dispersed 5S and ITS arrays, which may be subject to complex and incomplete concerted evolutionary forces. Despite the apparent good correlation between cluster complexity and a hybrid character of the genome there were several notable exceptions from the rule:

1. About 21% of allopolyploids and homoploids showed simple type 1 graphs. These simple graphs can be explained by high similarity of progenitor units, preventing separation of reads. However, it can also be explained by locus loss and/or homogenization of 5S rDNA in allopolyploids over longer evolutionary times. Indeed, ancient (c.5 Myr) *Nicotiana* allotetraploids from section Repandae showed simple circular type 1 graphs (not shown) and a diploid character of 5S rDNA loci (Lim et al., 2007).

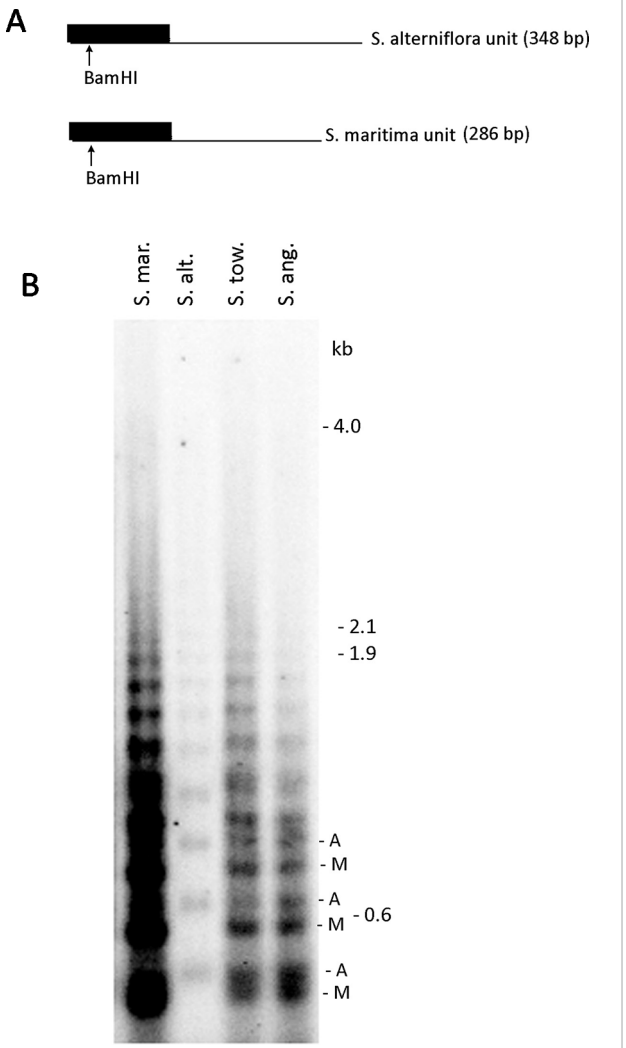

FIGURE 5 | Southern blot analysis of 5S rDNA in *Spartina* hybrids and their progenitors. **(A)** Restriction maps of progenitor *S. maritima* and *S. alterniflora* of 5S rDNA. **(B)** The 5S rDNA probe hybridization to *S. maritima* (S. mar.), *S. alterniflora* (S. alt.), *S.* × *townsendii* homoploid hybrid (S. tow.), and *S. anglica* dodecaploid (S. ang.) DNAs. In *S.* × *townsendii* and *S. anglica*, hybridization fragments inherited from progenitors are marked as "A" (from *S. alterniflora*) and "M" (from *S. maritima*), respectively.

Interestingly, *Triticum turgidum* (0.5 Myr) and *Spartina alterniflora* (3 Myr) (both Poaceae) polyploids also show simple cluster graphs despite their relatively young age, suggesting that the process of rDNA homogenization and diploidization may proceed at different rates in different systems. Frequent losses of 5S rDNA loci in *Triticum* (Baum et al., 2008), rice (Zhu et al., 2008) and *Spartina* (this work) polyploids may also suggest certain instability of 5S rDNA in Poaceae. Nevertheless, the *Thinopyrum* hybrid (Poaceae) displayed a highly complex cluster graph (**Figures 4C, D**) consistent with the retention of progenitor 5S rDNA families (Mahelka et al., 2013) and therefore arguing against generalization of these observations.

2. About 13% of diploids showed complex type 2 graphs, indicating intragenomic heterogeneity of 5S rDNA loci in these genomes. In at least some cases, the intragenomic heterogeneity of 5S rDNA in these diploids can be explained by homoploid hybridization and introgression events. This is probably the explanation for the complex graphs in *Gossypium gossypioides* (**Figure 4A**) which has a complex evolutionary history entailing at least two temporally widely separated divergence events (Wendel et al., 1995b; Cronn et al., 2003). Although introgression and hybridization is also relatively frequent in the banana genus (Němečková et al., 2018) a more likely explanation for complex graph structures in *Musa acuminata* (**Figure 6C**) is an exceptionally high number of 5S rDNA loci in this species (six per diploid genome) (Valarik et al., 2002; Garcia et al., 2012a) and probably inefficient interlocus recombination (Schlotterer and Tautz, 1994) leading to poor homogenization. Actually, the mechanisms of amplification of 5S rDNA loci across the chromosomes are still poorly understood (Schubert and Wobus, 1985; Symonova et al., 2017; Joachimiak et al., 2018; Souza et al., 2019).
3. The occurrence of non-rDNA sequences within the 5S rDNA clusters may potentially distort graph shapes. In *Tragopogon porrifolius* and *Senecio campestris* (both Asteraceae) the 5S rDNA clusters apparently contain traces of Cassandra transposable elements. These LTR elements are widespread in angiosperm genomes and carry a 5S rDNA related sequence (Kalendar et al., 2008). In cluster graphs the Cassandra element can be identified by divergent reads connected by only a few nodes to the 5S rDNA genic region (**Figures 6A, B**). The known high mobility of 5S rDNA in the *Musa* genus (Valarik et al., 2002) could be related to the activity of transposable elements whose remnants (TY1 copia/Tork family) are apparently found in some *M. acuminata* 5S rDNA units (**Figure 6C**). In general, the frequency of non-rDNA sequences was low (<4% analyses) in the major 5S rDNA clusters and likely does not represent significant source of artefacts.

Concluding Remarks

To summarize, we infer that the visual inspection of rDNA cluster graph topologies coupled with calculation of graph parameters is highly informative for the assessment of rDNA genomic organization, number of rRNA gene families, and loci. The method may provide clues for testable hypotheses about evolutionary histories of interspecific hybrids and allopolyploids, especially in biological systems with unknown or not well defined genome donors (Mahelka et al., 2011; Kaplan et al., 2013; Fredotovic et al., 2014; Belyayev et al., 2018). It is necessary to stress that a robust evaluation of hybridization and polyploidy cannot be solely based on read clustering, but should involve a combination of various cytogenetic, molecular and genomic methods.

MATERIALS AND METHODS

DNA Isolation, High-Throughput Sequencing, and Read Archive Accessions

Most sequences used in this study were downloaded from sequence read archives at the EBI server (**Supplementary Table S1**). Six genomes were sequenced *de novo* as follows: genomic DNA from leaf tissue was isolated by a modified CTAB method and sequenced by Illumina technology at BGI. The *Spartina* DNAs originated from natural samples collected in Southampton area, UK: *S. maritima* (Isle of Wight), *S.*× *townsendii* (Hythe), *S. alterniflora*, and *S. anglica* (both from Eling Marchwood) (Huska et al., 2016); *Cardamine* × *insueta* and *C. amara* were from natural populations in Urnerboden, Switzerland (Zozomova-Lihova et al., 2014); *C. flexuosa* was from Zelezne, Slovakia, and *C. hirsuta* from Gehausen, Germany (Mandakova et al., 2014). Details of sequencing are provided in **Supplementary Table S3**.

In Silico Identification of 5S rDNA Repeats

The fastq reads were initially filtered for quality and trimmed to a uniform length by pre-processing and QC tools by RepeatExplorer2 (Novak et al., 2013). The pipeline is implemented in the Galaxy environment (https://galaxy-elixir.cerit-sc.cz/). For computation, resources at the international ELIXIR infrastructure (European research infrastructure for biological information) were used. Read length ranged 100–150 bp depending on sequencing library and platform. After the fastq > fasta conversion reads were analyzed with RepeatExplorer2 using default parameters. The RepeatExplorer2 pipeline runs a graph-based clustering algorithm (Novak et al., 2013) to assemble the groups of frequently overlapping reads into clusters of reads, representing a repetitive element, or part of a repetitive element with a higher order genome structure. It uses a BLAST threshold of 90% similarity across 55% of the read to identify reads to each cluster by default (minimum overlap = 55, cluster threshold = 0.01%, minimum overlap for assembly = 40), and the clusters are identified based on a principle of maximum modularity. Typically, 200,000 of pair-end reads were used as input for clustering. This number typically yields a cluster comprising several hundreds of 5S rDNA specific reads. Analysis of larger (>2 Gb/1C) genomes requires an increase of the number of input reads (up to 2 million), as the 5S rDNA coverage decreases. Although this prolongs computation times (typically 5–6 hours on the MetaCentrum ELIXIR computer clusters) we were able to reconstruct 5S rDNA units in the large (50 GB) *Fritillaria imperialis* genome (Zonneveld, 2010) although, in this case, the number of reads in the cluster was too low, preventing graph analysis. High coverages may also help to reveal rare 5S rDNA variants and pseudogenes that are frequent in gymnosperms (Wang et al., 2016; Wang et al., 2019) while they rarely occur in angiosperms. In interspecific comparisons, the usage of a standard fraction of genome (0.1–1.0%) is recommended to prevent biases in interspecific comparisons.

The 5S rDNA clusters were searched among the *cluster annotation* files using "rDNA" search keyword. Alternatively, 5S rDNA clusters were found in TAREAN tandem reports (a specific tool for the analysis of tandem repeats implemented in RepeatExplorer2). The shapes of cluster graphs were characterized by a *connected component index* parameter (C) which is calculated as the proportion of the largest strongly connected component in graph composed of oriented reads (Novak et al., 2017). Cluster graph topologies were visually inspected and categorized into two groups (simple, type 1, and complex, type 2, graphs). The k-mer score was calculated by the RepeatExplorer2/TAREAN program as the sum of frequencies of all k-mers used for consensus sequence reconstruction.

Identification and Quantification of Homoeologous 5S rRNA Gene Families

5S rDNA homoeologous families were quantified by mapping analysis using CLC Genomics Workbench (QIAGEN), CLC onwards. Trimmed reads (typically >7 million) were mapped to the corresponding reference with following parameters: insertion and deletion costs = 3, lengths fraction = 0.5, similarity fraction = 0.9, deletion cost = 2. As reference sequences we used *Gossypium arboreum* (GenBank no. GAU31855) and *G. raimondii* (GRU39497) clones. Since no

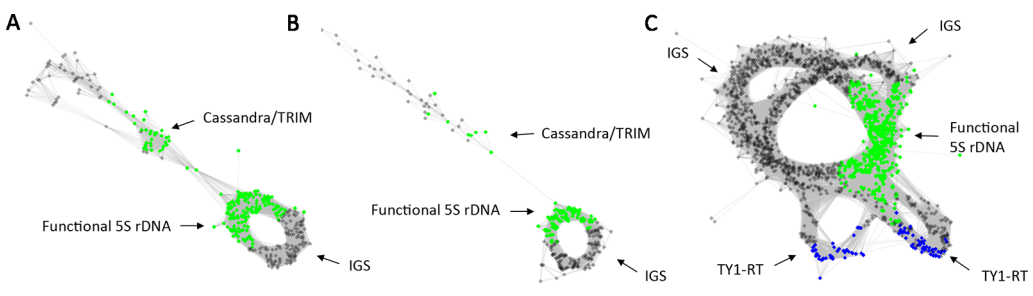

FIGURE 6 | Graph structures of 5S rDNA clusters from the RepeatExplorer2 graphical output containing a significant number of retroelement sequences in *Tragopogon porrifolius* **(A)**, *Senecio vulgaris* **(B)** and *Musa acuminata* **(C)**. In **(A, B)** reads derived from a Cassandra/TRIM element. Note, the Cassandra/TRIM elements were relatively isolated and connected to 5S rDNA with 1–2 reads only. In **(C)** the highly complex structure of 5S rDNA in *M. acuminata* containing multiple IGS and regions of similarity to a reverse transcriptase domain of a TY-1 Copia retroelement.

GenBank 5S rDNA clones were available for *Brachypodium* we used consensus sequences of *B. distachyon* (370 bp) and *B. stacei* (270 bp) generated by RepeatExplorer2 as a reference.

Phylogenetic analysis was carried out using assembled contigs computed by RepeatExplorer2. Briefly, BLAST libraries of contigs from hybrids and allopolyploids were BLASTed against the 5S rDNA sequences: For *Gossypium,* these were GenBank clones (#GAU31855 and #GRU39497); for *Brachypodium* and *Spartina* contigs generated by RepeatExplorer were used. The IGS subregions were extracted from BLAST outputs by a selection command in CLC, grouped and aligned. Alignments were manually edited and neighbor joining phylogeny trees constructed (CLC).

Because homoeologous ITS1 (internally transcribed spacer 1 of 35S rDNA) cannot be quantified by mapping procedures due to their overall similarity, we calculated the ITS1 homoeologous ratios from the number of nodes in genome-specific clades of phylogeny trees: (i) ITS sequences from *Gossypium* and *Brachypodium* allotetraploids were extracted from mapped reads, yielding typically hundreds to thousands of sequences. (ii) Stand-alone BLAST databases were generated from the ITS sequences. The databases were queried with reference sequences derived from variable 50–70 bp central subregions of ITS1. The ITS1 consensus sequences were obtained from the alignment of GenBank clones: *B. stacei* (JX665827-JX665832), *B. hybridum* (JX665718-JX665731), *G. arboreum* (GAU12712), and *G. raimondii* (GTU12711). (iii) Reads extracted from BLAST outputs were trimmed to uniform length, sampled (100–500 reads), and NJ trees constructed using a phylogeny tool of CLC. Homoeologous sequences in distinct clades were extracted, counted, and expressed as a ratio.

Statistical Methods

We analyzed the data by one-way ANOVA statistical test implemented within the MS Office package (XL-Toolbox NG). Box-plots were constructed using an online BoxPlotR tool (www.shiny.chmgid.org/boxplotr/).

Southern Blot Hybridization

Total genomic DNA was extracted from fresh young leaves using a modified CTAB method following procedures described previously (Kovarik et al., 1997). Genomic DNA was digested with the *Bam*HI restriction enzyme and hybridized on blots. The DNAs were digested with *Bam*HI and hybridized with the radioactively labeled ([32P] dCTP, Dekaprimer labeling kit (Thermo Fischer, USA) 5S rDNA probe. The probe was a trimer of the 5S rRNA gene from *Artemisia tridentata* [S4 clone, GenBank # JX101915.1, (Garcia et al., 2012b)]. Hybridization signals were visualized using a PhosphorImager (Fuji, FLA 9000).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GenBank Sequence Read Archives (Bioprojects "Chromosome evolution in invasive *Spartina* plants", ID: PRJNA575642 and "Chromosome evolution in *Cardamine* hybrids and polyploids", ID: PRJNA575831).

AUTHOR CONTRIBUTIONS

Conceived and designed the study: AKo, SG. Performed the experiments and collected material: SG, AKo, and MA. Analyzed the data: SG, MA, JW, NB-Z, AKu, and AKo. Wrote the manuscript: AKo, SG.

FUNDING

The work was supported by the Czech Science Foundation (grant 19-03442S), the Polish OPUS project of the National Science Centre (2018/31/B/NZ3/01761) and by the Dirección General de Investigación Científica y Técnica (CGL2016-75694-P AEI/FEDER, UE) from the government of Spain. SG benefited from a Ramón y Cajal contract (RYC-2014-16608) from the government of Spain.

ACKNOWLEDGMENTS

We thank members of the Malika Ainouche, Andrew R. Leitch and Martin Lysak laboratories for their kind help with *Spartina* and *Cardamine* sampling. We further thank the editor and reviewers for their helpful comments and the organizers of annual RepeatExplorer meetings in Ceske Budejovice (Czech Republic). Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the program Projects of Large Research, Development, and Innovations Infrastructures (CESNET LM2015042) is greatly appreciated.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00041/full#supplementary-material

REFERENCES

Ainouche, M. L., Baumel, A., and Salmon, A. (2004). *Spartina anglica* Schreb.: a natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biol. J. Linn. Soc* 82, 474–475. doi: 10.1111/j.1095-8312.2004.00334.x

Alix, K., Gerard, P. R., Schwarzacher, T., and Heslop-Harrison, J. S. (2017). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann. Bot.* 120, 619–619. doi: 10.1093/aob/mcx096

Alvarez, I., and Wendel, J. W. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434. doi: 10.1016/S1055-7903(03)00208-2

Baum, B. R., Bailey, L. G., Belyayev, A., Raskina, O., and Nevo, E. (2004). The utility of the nontranscribed spacer of 5S rDNA units grouped into unit classes assigned to haplomes - a test on cultivated wheat and wheat progenitors. *Genome* 47, 590–599. doi: 10.1139/g03-146

Baum, B. R., Edwards, T., and Johnson, D. A. (2008). Loss of 5S rDNA units in the evolution of agropyron, pseudoroegneria, and douglasdeweya. *Genome* 51, 589–598. doi: 10.1139/G08-045

Baum, B. R., Edwards, T., Mamuti, M., and Johnson, D. A. (2012). Phylogenetic relationships among the polyploid and diploid Aegilops species inferred from the nuclear 5S rDNA sequences (Poaceae: Triticeae). *Genome* 55, 177–193. doi: 10.1139/g2012-006

Belyayev, A., Pastova, L., Fehrer, J., Josefiova, J., Chrtek, J., and Mraz, P. (2018). Mapping of *Hieracium* (Asteraceae) chromosomes with genus-specific satDNA elements derived from next-generation sequencing data. *Plant Syst. Evol.* 304, 387–396. doi: 10.1007/s00606-017-1483-y

Besendorfer, V., Krajacić-Sokol, I., Jelenić, S., Puizina, J., Mlinarec, J., Sviben, T., et al. (2005). Two classes of 5S rDNA unit arrays of the silver fir, Abies alba Mill.: structure, localization and evolution. *Theor. Appl. Genet.* 110, 730–741. doi: 10.1007/s00122-004-1899-y

Borowska-Zuchowska, N., and Hasterok, R. (2017). Epigenetics of the preferential silencing of Brachypodium stacei-originated 35S rDNA loci in the allotetraploid grass Brachypodium hybridum. *Sci. Rep.* 7, 5260. doi: 10.1038/s41598-017-05413-x

Bourke, P. M., Voorrips, R. E., Visser, R. G. F., and Maliepaard, C. (2018). Tools for genetic studies in experimental populations of polyploids. *Front. Plant Sci.* 9, 513. doi: 10.3389/fpls.2018.00513

Boutte, J., Aliaga, B., Lima, O., de Carvalho, J. F., Ainouche, A., Macas, J., et al. (2016). Haplotype detection from Next-generation sequencing in high-ploidy-level species: 45S rDNA gene copies in the hexaploid Spartina maritima. *Genes Genomes Genet.* 6, 29–40. G3-. doi: 10.1534/g3.115.023242

Bughio, F., and Maggert, K. A. (2019). The peculiar genetics of the ribosomal DNA blurs the boundaries of transgenerational epigenetic inheritance. *Chromosome Res.* 27, 19–30. doi: 10.1007/s10577-018-9591-2

Catalán, P., Müller, J., Hasterok, R., Jenkins, G., Mur, L. A., Langdon, T., et al. (2012). Evolution and taxonomic split of the model grass Brachypodium distachyon. *Ann. Bot.* 109, 385–405. doi: 10.1093/aob/mcr294

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* 42, 685–705. doi: 10.1007/BF02338802

Cronn, R., Small, R. L., Haselkorn, T., and Wendel, J. F. (2003). Cryptic repeated genomic recombination during speciation in Gossypium gossypioides. *Evolution* 57 (11), 2475–2489. doi: 10.1111/j.0014-3820.2003.tb01493.x

Dodsworth, S., Chase, M. W., Kelly, L. J., Leitch, I. J., Macas, J., Novak, P., et al. (2015). Genomic repeat abundances contain phylogenetic signal. *Syst. Biol.* 64, 112–126. doi: 10.1093/sysbio/syu080

Dodsworth, S., Chase, M. W., Sarkinen, T., Knapp, S., and Leitch, A. (2016). Using genomic repeats for phylogenomics: a case study in wild tomatoes (Solanum section Lycopersicon: Solanaceae). *Biol. J. Linn. Soc* 117, 96–105. doi: 10.1111/bij.12612

Dover, G. A. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299, 111–117. doi: 10.1038/299111a0

Elder, J. F., and Turner, B. J. (1995). Concerted evolution of repetitive DNA-sequences in eukaryotes. *Q. Rev. Biol.* 70, 297–320. doi: 10.1086/419073

Fredotovic, Z., Samanic, I., Weiss-Schneeweiss, H., Kamenjarin, J., Jang, T. S., and Puizina, J. (2014). Triparental origin of triploid onion, Allium x cornutum (Clementi ex Visiani, 1842), as evidenced by molecular, phylogenetic and cytogenetic analyses. *BMC Plant Biol.* 14, 14. doi: 10.1186/1471-2229-14-24

Fulnecek, J., Lim, K. Y., Leitch, A. R., Kovarik, A., and Matyasek, R. (2002). Evolution and structure of 5S rDNA loci in allotetraploid Nicotiana tabacum and its putative parental species. *Heredity* 88, 19–25. doi: 10.1038/sj.hdy.6800001

Fulnecek, J., Matyasek, R., and Kovarik, A. (2006). Plant 5S rDNA has multiple alternative nucleosome positions. *Genome* 49, 840–850. doi: 10.1139/g06-039

Garcia, S., Garnatje, T., and Kovarik, A. (2012a). Plant rDNA database: ribosomal DNA loci data including other karyological and cytogenetic information in plants. *Chromosoma* 121, 389–394. doi: 10.1007/s00412-012-0368-7

Garcia, S., Crhak Khaitova, L., and Kovarik, A. (2012b). Expression of 5S rRNA genes linked to 35S rDNA in plants, their epigenetic modification and regulatory element divergence. *BMC Plant Biol.* 12, 95. doi: 10.1186/1471-2229-12-95

Garcia, S., Kovarik, A., Leitch, A. R., and Garnatje, T. (2017). Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. *Plant J.* 89, 1020–1030. doi: 10.1111/tpj.13442

Glover, N. M., Redestig, H., and Dessimoz, C. (2016). Homoeologs: what are they and how do we infer them? *Trends Plant Sci.* 21, 609–621. doi: 10.1016/j.tplants.2016.02.005

Grover, C. E., Arick, M. A., Thrash, A., Conover, J. L., Sanders, W. S., Peterson, D. G., et al. (2019). Insights into the evolution of the New world diploid cottons (Gossypium, subgenus Houzingenia) based on genome sequencing. *Genome Biol. Evol.* 11, 53–71. doi: 10.1093/gbe/evy256

Hemleben, V., and Zentgraf, U. (1994). "Structural organisation and regulation of transcription by RNA polymerase I of plant nuclear ribosomal genes," in *Results and problems in cell differentiation 20: plant promoters and transcription factors*, vol. 3-24 . Ed. N. Lapitan (Berlin/Heidelberg: Springer-Verlag). doi: 10.1007/978-3-540-48037-2_1

Huska, D., Leitch, I. J., de Carvalho, J. F., Leitch, A. R., Salmon, A., Ainouche, M., et al. (2016). Persistence, dispersal and genetic evolution of recently formed Spartina homoploid hybrids and allopolyploids in Southern England. *Biol. Invasions* 18, 2137–2151. doi: 10.1007/s10530-015-0956-6

Jang, T. S., McCann, J., Parker, J. S., Takayama, K., Hong, S. P., Schneeweiss, G. M., et al. (2016). rDNA loci evolution in the genus Glechoma (Lamiaceae). *PloS One* 11, e0167177. doi: 10.1371/journal.pone.0167177

Joachimiak, A. J., Hasterok, R., Sliwinska, E., Musial, K., and Grabowska-Joachimiak, A. (2018). FISH-aimed karyotype analysis in Aconitum subgen. Aconitum reveals excessive rDNA sites in tetraploid taxa. *Protoplasma* 255, 1363–1372. doi: 10.1007/s00709-018-1238-9

Kalendar, R., Tanskanen, J., Chang, W., Antonius, K., Sela, H., Peleg, O., et al. (2008). Cassandra retrotransposons carry independently transcribed 5S RNA. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5833–5838. doi: 10.1073/pnas.0709698105

Kaplan, Z., Jarolimova, V., and Fehrer, J. (2013). Revision of chromosome numbers of Potamogetonaceae: a new basis for taxonomic and evolutionary implications. *Preslia* 85, 421–482. doi: 10.1007/s12228-012-9299-0

Kellogg, E. A., and Appels, R. (1995). Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics* 140 (1), 325–343.

Kovarik, A., Matyasek, R., Leitch, A., Gazdova, B., Fulnecek, J., and Bezdek, M. (1997). Variability in CpNpG methylation in higher plant genomes. *Gene* 204, 25–33. doi: 10.1016/S0378-1119(97)00503-9

Lim, K. Y., Kovarik, A., Matyasek, R., Chase, M. W., Clarkson, J. J., Grandbastien, M. A., et al. (2007). Sequence of events leading to near-complete genome turnover in allopolyploid Nicotiana within five million years. *New Phytol.* 175, 756–763. doi: 10.1111/j.1469-8137.2007.02121.x

Lunerova, J., Renny-Byfield, S., Matyasek, R., Leitch, A., and Kovarik, A. (2017). Concerted evolution rapidly eliminates sequence variation in rDNA coding regions but not in intergenic spacers in Nicotiana tabacum allotetraploid. *Plant Syst. Evol.* 303, 1043–1060. doi: 10.1007/s00606-017-1442-7

Mahelka, V., Kopecky, D., and Pastova, L. (2011). On the genome constitution and evolution of intermediate wheatgrass (Thinopyrum intermedium: Poaceae, Triticeae). *BMC Evol. Biol.* 11, 127. doi: 10.1186/1471-2148-11-127

Mahelka, V., Kopecky, D., and Baum, B. R. (2013). Contrasting Patterns of Evolution of 45S and 5S rDNA families uncover new aspects in the genome constitution of the agronomically important grass Thinopyrum intermedium (Triticeae). *Mol. Biol. Evol.* 30, 2065–2086. doi: 10.1093/molbev/mst106

Mandakova, T., Marhold, K., and Lysak, M. A. (2014). The widespread crucifer species Cardamine flexuosa is an allotetraploid with a conserved subgenomic structure. *New Phytol.* 201, 982–992. doi: 10.1111/nph.12567

Matyasek, R., Lim, K. Y., Kovarik, A., and Leitch, A. R. (2003). Ribosomal DNA evolution and gene conversion in Nicotiana rustica. *Heredity* 91 (3), 268–275. doi: 10.1038/sj.hdy.6800333

Matyasek, R., Renny-Byfield, S., Fulnecek, J., Macas, J., Grandbastien, M. A., Nichols, R., et al. (2012). Next generation sequencing analysis reveals a relationship between rDNA unit diversity and locus number in Nicotiana diploids. *BMC Genomics* 13, 722. doi: 10.1186/1471-2164-13-722

Mlinarec, J., Skuhala, A., Jurković, A., Malenica, N., McCann, J., Weiss-Schneeweiss, H., et al. (2019). The repetitive DNA composition in the natural pesticide producer Tanacetum cinerariifolium: interindividual

variation of subtelomeric tandem repeats. *Front. Plant Sci.* 10, 613. doi: 10.3389/fpls.2019.00613

Muir, G., Fleming, C. C., and Schlotterer, C. (2001). Three divergent rDNA clusters predate the species divergence in Quercus petraea (Matt.) Liebl. and Quercus robur L. *Mol. Biol. Evol.* 18, 112–119. doi: 10.1093/oxfordjournals.molbev.a003785

Němečková, A., Christelová, P., Čížková, J., Nyine, M., Van den Houwe, I., Svačina, R., et al. (2018). Molecular and cytogenetic study of east African highland banana. *Front. Plant Sci.* 9, 1371. doi: 10.3389/fpls.2018.01371

Nieto Feliner, G., and Rosselló, J. A. (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* 44, 911–919. doi: 10.1016/j.ympev.2007.01.013

Nieto Feliner, G., and Rossello, J. A. (2012). "Concerted evolution of multigene families and homeologous recombination," in *Plant Genome Diversity*. Ed. J. F. Wendel (Wien: Springer-Verlag), 171–194. doi: 10.1007/978-3-7091-1130-7_12

Nieto Feliner, N., Alvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., et al. (2017). Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* 118, 513–516. doi: 10.1038/hdy.2017.7

Novak, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinf.* 11, 378. doi: 10.1186/1471-2105-11-378

Novak, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793. doi: 10.1093/bioinformatics/btt054

Novak, P., Robledillo, L. A., Koblizkova, A., Vrbova, I., Neumann, P., and Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucl. Acids Res.* 45, E111. doi: 10.1093/nar/gkx257

Parks, M. M., Kurylo, C. M., Batchelder, J. E., Vincent, C. T., and Blanchard, S. C. (2019). Implications of sequence variation on the evolution of rRNA. *Chromosome Res.* 27, 89–93. doi: 10.1007/s10577-018-09602-w

Pedrosa-Harand, A., de Almeida, C. C. S., Mosiolek, M., Blair, M., Schweizer, D., and Guerra, M. (2006). Extensive ribosomal DNA amplification during Andean common bean (Phaseolus vulgaris L.) evolution. *Theor. Appl. Genet.* 112, 924–933. doi: 10.1007/s00122-005-0196-8

Peska, V., Mandakova, T., Ihradska, V., and Fajkus, J. (2019). Comparative dissection of three giant genomes: *Allium cepa*, *Allium sativum*, and *Allium ursinum*. *Int. J. Mol. Sci.* 20, 25. doi: 10.3390/ijms20030733

Poczai, P., and Hyvonen, J. (2010). Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Mol. Biol. Rep.* 37, 1897–1912. doi: 10.1007/s11033-009-9630-3

Renny-Byfield, S., Kovarik, A., Chester, M., Nichols, R. A., Macas, J., Novak, P., et al. (2012). Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of Nicotiana tabacum. *PloS One* 7, 722. doi: 10.1371/journal.pone.0036963

Ribeiro, T., Marques, A., Novak, P., Schubert, V., Vanzela, A. L., Macas, J., et al. (2017). Centromeric and non-centromeric satellite DNA organisation differs in holocentric Rhynchospora species. *Chromosoma* 126, 325–335. doi: 10.1007/s00412-016-0616-3

Röser, M., Winterfeld, G., Grebenstein, B., and Hemleben, V. (2001). Molecular diversity and physical mapping of 5S rDNA in wild and cultivated oat grasses (Poaceae: Aveneae). *Mol. Phylogenet. Evol.* 2, 198–217. doi: 10.1006/mpev.2001.1003

Roa, F., and Guerra, M. (2012). Distribution of 45S rDNA sites in chromosomes of plants: structural and evolutionary implications. *BMC Evol. Biol.* 12, 225. doi: 10.1186/1471-2148-12-225

Roa, F., and Guerra, M. (2015). Non-random distribution of 5S rDNA sites and its association with 45S rDNA in plant chromosomes. *Cytogenet. Genome Res.* 146, 243–249. doi: 10.1159/000440930

Sastri, D. C., Hilu, K., Appels, R., Lagudah, E. S., Playford, J., and Baum, B. R. (1992). An overview of evolution in plant 5S-DNA. *Plant Syst. Evol.* 183, 169–181. doi: 10.1007/BF00940801

Schlotterer, C., and Tautz, D. (1994). Chromosomal homogeneity of Drosophila ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.* 4, 777–783. doi: 10.1016/S0960-9822(00)00175-5

Schubert, I., and Wobus, U. (1985). In situ hybridization confirms jumping nucleolus organizing regions in Allium. *Chromosoma* 92, 143–148. doi: 10.1007/BF00328466

Souza, G., Marques, A., Ribeiro, T., Dantas, L. G., Speranza, P., Guerra, M., et al. (2019). Allopolyploidy and extensive rDNA site variation underlie rapid karyotype evolution in Nothoscordum section Nothoscordum (Amaryllidaceae). *Bot. J. Linn. Soc* 190, 215–228. doi: 10.1093/botlinnean/boz008

Symonova, R., Ocalewicz, K., Kirtiklis, L., Delmastro, G. B., Pelikanova, S., Garcia, S., et al. (2017). Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (Esox sp.). *BMC Genomics* 18, 391. doi: 10.1186/s12864-017-3774-7

Tang, S., Li, Z., Jia, X., and Larkin, P. J. (2000). Genomic in situ hybridization (GISH) analyses of Thinopyrum intermedium, its partial amphiploid Zhong 5, and disease-resistant derivatives in wheat. *Theor. Appl. Genet.* 100, 344–352. doi: 10.1007/s001220050045

Valarik, M., Simkova, H., Hribova, E., Safar, J., Dolezelova, M., and Dolezel, J. (2002). Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (Musa spp.). *Chromosome Res.* 10, 89–100. doi: 10.1023/A:1014945730035

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

Vitales, D., Garcia, S., and Dodsworth, S. (2019). "Recostructing phylogenetic relatioships based on repeat sequence similarities," in *BioRxiv* (USA: Cold Spring Harbour Laboratory publishers). doi: 10.1101/624064

Volkov, R. A., Borisjuk, N. V., Panchuk, I. I., Schweizer, D., and Hemleben, V. (1999). Elimination and rearrangement of parental rDNA in the allotetraploid Nicotiana tabacum. *Mol. Biol. Evol.* 16, 311–320. doi: 10.1093/oxfordjournals.molbev.a026112

Volkov, R. A., Komarova, N. Y., and Hemleben, V. (2007). Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Syst. Biodivers. (NHM London)* 5, 261–276. doi: 10.1017/S1477200007002447

Wang, W. C., Ma, L., Becher, H., Garcia, S., Kovarikova, A., Leitch, I. J., et al. (2016). Astonishing 35S rDNA diversity in the gymnosperm species Cycas revoluta Thunb. *Chromosoma* 125, 683–699. doi: 10.1007/s00412-015-0556-3

Wang, W., Wan, T., Becher, H., Kuderova, A., Leitch, I. J., Garcia, S., et al. (2019). Remarkable variation of ribosomal DNA organization and copy number in gnetophytes, a distinct lineage of gymnosperms. *Ann. Bot.* 123, 767–781. doi: 10.1093/aob/mcy172

Weiss-Schneeweiss, H., Tremetsberger, K., Schneeweiss, G. M., Parker, J. S., and Stuessy, T. F. (2008). Karyotype diversification and evolution in diploid and polyploid South American Hypochaeris (Asteraceae) inferred from rDNA localization and genetic fingerprint data. *Ann. Bot.* 101, 909–918. doi: 10.1093/aob/mcn023

Weiss-Schneeweiss, H., Leitch, A. R., McCann, J., Jang, T. S., and Macas, J. (2015). "Employing next-generation sequencing to explore the repeat landscape of the plant genome," in *Next generation sequencing in plant systematics*. Eds. E. Hörandl and M. S. Appelhans (Germany: Königstein), 155–179. doi: 10.14630/000006

Wendel, J. F., Schnabel, A., and Seelanan, T. (1995a). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium). *Proc. Natl. Acad. Sci. U.S.A.* 92, 280–284. doi: 10.1073/pnas.92.1.280

Wendel, J. F., Schnabel, A., and Seelanan, T. (1995b). An unusual ribosomal DNA-sequence from Gossypium gossypioides reveals ancient, cryptic, intergenomic introgression. *Mol. Phylogenet. Evol.* 4, 298–313. doi: 10.1006/mpev.1995.1027

Wendel, J. F., Lisch, D., Hu, G. J., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/j.gde.2018.01.004

Wendel, J. F. (1989). New World tetraploid cottons contain Old World cytoplasm. *Proc. Natl. Acad. Sci. U.S.A.* 86, 4132–4136. doi: 10.1073/pnas.86.11.4132

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320

West, C., James, S. A., Davey, R. P., Dicks, J., and Roberts, I. N. (2014). Ribosomal DNA sequence heterogeneity reflects intraspecies phylogenies and predicts genome structure in two contrasting yeast species. *Syst. Biol.* 63, 543–554. doi: 10.1093/sysbio/syu019

Yoo, M. J., Liu, X. X., Pires, J. C., Soltis, P. S., and Soltis, D. E. (2014). Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* 48, 485–517. doi: 10.1146/annurev-genet-120213-092159

Zhu, X. Y., Cai, D. T., and Ding, Y. (2008). Molecular and cytological characterization of 5S rDNA in Oryza species: genomic organization and phylogenetic implications. *Genome* 51, 332–340. doi: 10.1139/G08-016

Zonneveld, B. J. (2010). New record holders for maximum genome size in eudicots and monocots. *J. Bot.* article ID 527357. (Hindawi publishing corporation). doi: 10.1155/2010/527357

Zozomova-Lihova, J., Mandakova, T., Kovarikova, A., Muhlhausen, A., Mummenhoff, K., Lysak, M. A., et al. (2014). When fathers are instant losers: homogenization of rDNA loci in recently formed Cardamine x schulzii trigenomic allopolyploid. *New Phytol.* 203, 1096–1108. doi: 10.1111/nph.12873

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

frontiers in Plant Science

MINI REVIEW
published: 21 February 2020
doi: 10.3389/fpls.2020.00117

Origin, Diversity, and Evolution of Telomere Sequences in Plants

Vratislav Peska[1*] *and Sònia Garcia*[2*]

[1] Department of Cell Biology and Radiobiology, The Czech Academy of Sciences, Institute of Biophysics, Brno, Czechia,
[2] Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Barcelona, Spain

Telomeres are basic structures of eukaryote genomes. They distinguish natural chromosome ends from double-stranded breaks in DNA and protect chromosome ends from degradation or end-to-end fusion with other chromosomes. Telomere sequences are usually tandemly arranged minisatellites, typically following the formula $(T_xA_yG_z)_n$. Although they are well conserved across large groups of organisms, recent findings in plants imply that their diversity has been underestimated. Changes in telomeres are of enormous evolutionary importance as they can affect whole-genome stability. Even a small change in the telomere motif of each repeat unit represents an important interference in the system of sequence-specific telomere binding proteins. Here, we provide an overview of telomere sequences, considering the latest phylogenomic evolutionary framework of plants in the broad sense (Archaeplastida), in which new telomeric sequences have recently been found in diverse and economically important families such as Solanaceae and Amaryllidaceae. In the family Lentibulariaceae and in many groups of green algae, deviations from the typical plant telomeric sequence have also been detected recently. Ancestry and possible homoplasy in telomeric motifs, as well as extant gaps in knowledge are discussed. With the increasing availability of genomic approaches, it is likely that more telomeric diversity will be uncovered in the future. We also discuss basic methods used for telomere identification and we explain the implications of the recent discovery of plant telomerase RNA on further research about the role of telomerase in eukaryogenesis or on the molecular causes and consequences of telomere variability.

Keywords: *Allium*, *Cestrum*, circular chromosomes, *Genlisea*, green algae, linear chromosomes, telomerase, telomeres

OPEN ACCESS

Edited by:
Hanna Weiss-Schneeweiss, University of Vienna, Austria

Reviewed by:
Jasna Puizina, University of Split, Croatia
Predrag Slijepcevic, Brunel University London, United Kingdom

***Correspondence:**
Vratislav Peska
vpeska@ibp.cz
Sònia Garcia
soniagarcia@ibb.csic.es

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

Received: *22 November 2019*
Accepted: *27 January 2020*
Published: *21 February 2020*

Citation:
Peska V and Garcia S (2020) Origin, Diversity, and Evolution of Telomere Sequences in Plants. Front. Plant Sci. 11:117. doi: 10.3389/fpls.2020.00117

INTRODUCTION

Telomeres are nucleoprotein structures at the very ends of linear eukaryotic chromosomes. They solve two major end-problems at the same time. The first is about chromosome end protection. It is estimated that normal human cells must repair at least 50 endogenous double-stranded breaks (DSBs) per cell per cell-cycle (Vilenchik and Knudson, 2003). Telomeres distinguish the natural chromosomal ends from harmful DSBs and prevent their ectopic repair, e.g., by end-to-end fusions of chromosomes (vanSteensel and deLange, 1997). The second is the end-replication problem that

deals with the maintenance of proper telomere lengths. This was recognized independently by two researchers (Watson, 1972; Olovnikov, 1973). Since replicative DNA-dependent DNA polymerases cannot complete DNA synthesis at the very ends of chromosomes, compensation for replicative telomere sequence loss must come from an RNA-dependent DNA polymerase. This enzyme, called telomerase, together with the first telomere minisatellite sequence, was discovered in the ciliate *Tetrahymena* (Blackburn and Gall, 1978; Greider and Blackburn, 1985). However, this is only one aspect of telomere length maintenance. The epigenetic regulation of telomere length homeostasis, including interaction of long noncoding telomeric repeat containing RNA and exonuclease activity pathways, have also been extensively studied due to its therapeutical potential (Wellinger et al., 1996; Polotnianka et al., 1998; Pfeiffer and Lingner, 2012).

Telomerase, the enzyme in charge of adding telomere repeat sequences to the 3' end of telomeres, is a conserved complex enzyme with numerous components [its structure has been recently reviewed by (Wang et al., 2019), and specifically for plants, by (Majerska et al., 2017)]. In principle, only two main components are essential for telomerase enzymatic activity, a catalytically active protein component, called telomere reverse transcriptase (TERT), and a template component, formed by the telomerase RNA subunit (TR). While TERT is evolutionarily quite well conserved, TR is very variable, with lengths ranging from ca. 150 nt (*Tetrahymena*) to more than 2,000 nt (fungi from genus *Neurospora*). Only a short region in the whole TR molecule serves as a template for newly synthesized telomere DNA (Greider and Blackburn, 1985; Qi et al., 2013). This region in TR is usually formed by a complete telomere motif followed by a partial one, the latter serving as an annealing region for the existing telomere DNA. Although, in principle, only a single extra nucleotide is needed (as a partial motif), usually more than one is found. For example, two extra nucleotides form the annealing motif in mice or five in human (Blasco et al., 1995; Feng et al., 1995). In plants, however, the size of the template region is variable, e.g., two in *Arabidopsis thaliana*, seven in *Arabis* sp. or six in *Nicotiana* (Fajkus et al., 2019). The other TR regions have structural, regulatory and protein interactive functions [reviewed in (Podlevsky and Chen, 2016)]. See also a schematic depiction of telomerase and its activity cycle in **Figure 1**.

HOW VARIABLE ARE TELOMERE SEQUENCES?

Telomere sequences are usually short minisatellites tandemly arranged, typically following the formula $(T_xA_yG_z)_n$. The minisatellite arrangement originates from the way in which telomerase synthesizes the DNA, in short, and mostly identical motifs, one by one. Several hypotheses consider that such an arrangement is important because it promotes the recognition of telomere specific proteins by homo- and heterodimers [e.g., (Hofr et al., 2009; Visacka et al., 2012)] and for the potential to form G-quadruplexes that may stabilize chromosome ends or serve as substrates for telomere-specific proteins (Spiegel et al., 2020; Tran et al., 2013). Telomere sequences are well conserved through evolution, and large groups of organisms use the group-typical telomere motif to build their telomere DNA. A gradually increasing number of studies and large screenings have shown that all tested vertebrates and many basal metazoans use TTAGGG (Meyne et al., 1989; Traut et al., 2007) while Euarthropoda (arthropods), including Hexapoda (insects), have TTAGG (Frydrychova et al., 2004; Vitkova et al., 2005). Steadily, numerous exceptions are accumulating over time, e.g., $(A(G)_{1-8})$ in *Dictyostelium* (Emery and Weiner, 1981), TTAGGC in *Ascaris*

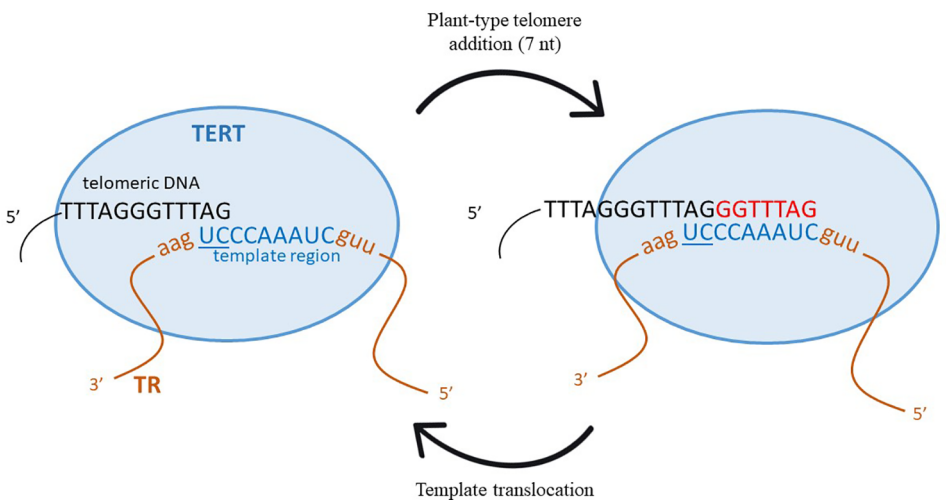

FIGURE 1 | Schematic representation of the telomerase activity cycle with the *Arabidopsis*-type telomere template. TERT, Telomere Reverse Transcriptase; TR, telomerase RNA subunit. Figure based on Sekhri (2014).

lumbricoides (Nematoda) (Muller et al., 1991), TCAGG in Coleoptera (beetles) (Mravinac et al., 2011), TAGGG/TAAGG/TAAGGG in *Giardia* (diplomonads) (Uzlikova et al., 2017), or TTNNNNAGGG in *Yarrowia* clade (yeasts) (Cervenak et al., 2019). Moreover, telomerase-independent systems, in which the minisatellite telomere sequence has been lost and substituted by complex repeats, are represented, for example, by *Diptera* and Chironomidae (reviewed in (Mason et al., 2016)). For a general review on eukaryotic telomere sequence see (Fajkus et al., 2005; Fulneckova et al., 2013).

Telomere composition in plants is even more diverse. Here we use the term "plants" in a broad sense, also known as Archaeplastida or kingdom Plantae *sensu lato*, and comprising Rhodophyta (red algae), Glaucophyta, the Chlorophyte algae grade and the Streptophyte algae grade (altogether known as green algae), and Embryophyta (land plants) (One Thousand Plant Transcriptomes Initiative, 2019). The typical telomere plant sequence is TTTAGGG, also called *Arabidopsis*-type (or simply, plant-type) since it was discovered in *Arabidopsis thaliana* (Richards and Ausubel, 1988) and now in many other species across almost all plant orders. Although TTTAGGG is still the most frequent, there is significant variability in telomere sequences in red and green algal lineages. As for red algae (Rhodophyta), telomere sequence information is mostly missing or fragmentary, although some telomere candidates have been discovered *in silico*, such as AATGGGGGG for *Cyanidioschyzon merolae* (Nozaki et al., 2007), TTATT(T)AGGG for *Galdieria sulphuraria* (Fulneckova et al., 2013); TTAGGG has been found in genomic reads of *Porphyra umbilicalis* (Fulneckova et al., 2013), but more evidence is needed to confirm their terminal position on chromosomes. Telomere diversity in green algae reflects both dynamic changes and its paraphyletic character. Although TTTAGGG prevails in Chlorophyta, such as in genera *Ostreococcus* (Derelle et al., 2006) and *Chlorella* (Higashiyama et al., 1995), many other divergent motifs have been detected there too, such as TTAGGG in genus *Dunaliella* and *Stephanosphaeria* (Fulneckova et al., 2012), and TTTTAGGG in *Chlamydomonas* (Petracek et al., 1990). In basal Streptophyta (Klebsormidiophyceae) progressive changes in motifs from TTTAGGG to TTTTAGGG and TTTTAGG have been described. The presence of TTAGGG in Rhodophyta and Glaucophyta leads to the hypothesis that this is the ancestral motif in plants (Archaeplastida) (Fulneckova et al., 2013).

Concerning land plants, one of the first screenings performed showed that the *Arabidopsis*-type sequence was the most common and was mostly conserved through their phylogeny (Cox et al., 1993; Fuchs et al., 1995), although some of these authors had already detected several exceptions in the family Amaryllidaceae (former Alliaceae), in which the *Arabidopsis*-type sequence was absent in several species. Later, the first telomere sequence unusual for land plants, the vertebrate-type TTAGGG, was characterized in *Aloe* and in some other Asparagales (Weiss and Scherthan, 2002; Puizina et al., 2003; Sykorova et al., 2003c). A hypothesis about repeated losses and recoveries of the TTTAGGG and TTAGGG telomere sequence in Asparagales was formulated (Adams et al., 2001). With the postrefinement of order Asparagales in the APGIII (Angiosperm Phylogeny Group 2009) (Bremer et al., 2009), it was shown that only two major evolutionary switches in telomere sequence composition occurred (rather than several repeated losses and gains), in the following order: the first one in family Iridaceae, in which a shift from the plant-type TTTAGGG to the vertebrate-type TTAGGG happened, followed by families Xeronemataceae, Asphodelaceae and the core Asparagales (including Amarillidaceae s.l and Asparagaceae s.l.); and the second one within subfamily Allioideae (formerly treated as a separate family, Alliaceae) in which a completely new telomere sequence emerged, CTCGGTTATGGG (Fajkus et al., 2016). Outside Asparagales, new telomere sequences have also been detected in land plant groups as disparate as (i) Solanaceae, in which the telomere sequence of *Cestrum elegans* TTTTTTAGGG was described (Sykorova et al., 2003a; Sykorova et al., 2003b; Peska et al., 2008; Peska et al., 2015) and (ii) Lentibulariaceae, where genus *Genlisea* showed a remarkable diversity with some species characterized by the *Arabidopsis*-type telomere repeats while others exhibited intermingled sequence variants TTCAGG and TTTCAGG (Tran et al., 2015).

Despite all the telomere motif exceptions detected, the real diversity in telomeric sequences in land plants is probably greatly underestimated. A recent publication (Vitales et al., 2017), in which a screening of land plant telomere sequences was performed, found that telomere sequences were only known clearly for less than 10% of the species and 40% of the genera contained in the Plant rDNA database (www.plantrdnadatabase.com), a resource providing molecular cytogenetics information on land plants (Garcia et al., 2012). A summary of telomere sequence distribution in plants, following APG IV (The Angiosperm Phylogeny Group, 2016) (Byng et al., 2016), as well as the most recent plant phylogeny (One Thousand Plant Transcriptomes Initiative, 2019) is found in **Figure 2**.

FROM SCREENINGS TO DISCOVERY: HOW TELOMERIC MOTIFS CAN BE IDENTIFIED?

The evidence that a given candidate sequence is a real telomeric one includes several steps that properly declare its localization at all chromosomal termini, and eventually the involvement of telomerase in its synthesis. Molecular cytogenetics (mostly by Fluorescence *in situ* Hybridization, FISH) has become important for visualizing the terminal localization of labeled probes of candidate sequences at all chromosomal termini. However, standalone FISH it is not enough to prove the very terminal position. For example, AcepSAT356 [a 356bp-long satellite from *Allium cepa*, (Peska et al., 2019)] was proposed in onion as the telomere candidate, based on results from FISH analysis (Pich and Schubert, 1998). Nevertheless, its apparent terminal location by *in situ* has never been convincingly linked to telomere function. Actually, the discovery of the *Allium* minisatellite

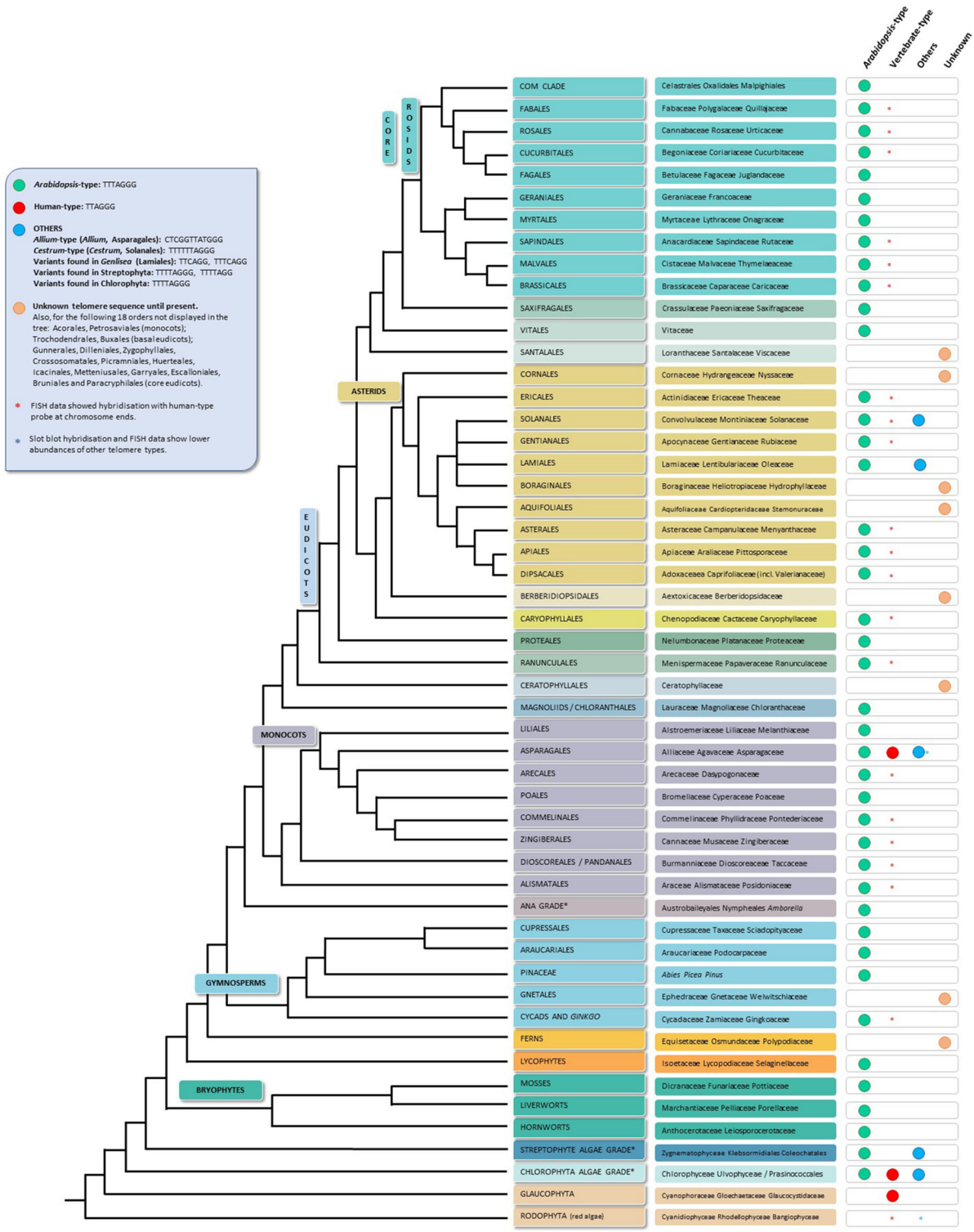

FIGURE 2 | Telomere motifs in Archaeplastida (plants in the broad sense), based on the APG IV (The Angiosperm Phylogeny Group 2016) and on the One Thousand Plant Transcriptomes Initiative (2019). Branch lengths do not express real time scales. For simplicity and to save space, certain polyphyletic "groups" (grades) marked with an asterisk in the tree have been represented by a single branch; for the same reason, several minor orders (listed in the blue square at the left upper side of the figure) are not depicted on the tree. The first tip label usually refers to plant orders and in a few cases, to divisions, grades and even families; the second label displays representative families and in a few cases, representative orders or genera.

telomere sequence CTCGGTTATGGG and telomerase would mean that AcepSAT356 is subterminal (Fajkus et al., 2019). Positive FISH telomeric signals can also mask tiny changes in telomere motifs such as single nucleotide polymorphisms, or false-negative results may result from short telomeres being beneath the detection limit of the technique.

There are two additional approaches that determine the terminal position at greater resolution than FISH; these are based on exonuclease BAL31 activity. The first is the classical Terminal Restriction Fragment (TRF) analysis, in which samples treated by BAL31 show progressive shortening of terminal fragments and a decrease in signal intensity with increasing time of exonuclease treatment. The subsequent analysis of fragment lengths is performed by Southern-blot hybridization (Fojtova et al., 2015). The second is comparative genome skimming (NGS data) of nondigested and BAL31-digested genomic DNA, in parallel. In the BAL31 treated dataset, there is a significant under-representation of telomere sequences, therefore the terminal sequences are identified by comparison with the untreated dataset, using bioinformatics tools RepeatExplorer or Tandem Repeats Finder [a pipeline called BAL31-NGS (Benson, 1999; Novak et al., 2010; Peska et al., 2017)].

The other important test of a given telomere sequence candidate in a species is the demonstration of telomerase activity. In this, a useful experimental approach, developed first for human cells, is the Telomere Repeat Amplification Protocol (TRAP) (Kim et al., 1994), followed by sequencing of the detected products (Peska et al., 2015; Fajkus et al., 2016), which is a little less sensitive to false-positive results than FISH. All these methods, including FISH (Fuchs et al., 1995; Shibata and Hizume, 2011) and others such as slot-blot hybridization (Sykorova et al., 2003c), and TRAP (Fulneckova et al., 2012; Fulneckova et al., 2016), can be used to screen for telomeres across wide groups of complex organisms, including plants. However, only a combination of suitably chosen methods can convincingly lead to a conclusion about the telomere function of a candidate sequence, since results base on a single approach might be misleading. A more complete overview of the strategies for *de novo* telomere candidate sequence identification, including the very first attempt in *Tetrahymena* (Greider and Blackburn, 1985) are summarised in a methodological article, with emphasis on the NGS approach used in plants with extremely large genomes (Peska et al., 2017).

IS THERE HOMOPLASY IN TELOMERE SEQUENCES?

The ancestral telomere sequence is thought to be TTAGGG and is the most commonly found across the tree of life (Fulneckova et al., 2013). Yet, it seems clear that the frequency of homoplasy in telomere motif evolution is relatively high. For example, short, simple motifs like the plant-type TTTAGGG have appeared independently and repeatedly in cryptomonads, oomycete fungi, and alveolates; similarly, the vertebrate-type TTAGGG has emerged secondarily in certain groups of plants (Asparagales, Rodophyta and Chlorophyta algae) (Sykorova et al., 2003c; Fulneckova et al., 2012; Fulneckova et al., 2013; Somanathan and Baysdorfer, 2018). The reason some telomere sequences have emerged more frequently than other, usually more complex sequences is probably related to selection pressures, which would favor accuracy for a particular sequence-specific DNA-protein interaction (Forstemann et al., 2003). If there was a change in each telomere motif, interference in the telomeric nucleoprotein structure would necessarily lead to genome instability. This is the reason telomere sequences are so evolutionary stable, comprising very few novel and successful sequences, a pattern consistent with the idea of repeated losses and the emergence of the typical telomere sequences, as proposed for Asparagales (Adams et al., 2001).

The finding of homoplasy across telomere sequences raises the question, what are the molecular causes and processes taking place during these shifts? A change in telomere sequence, despite seeming trivial in some cases (e.g., one extra T), may cause serious interference with genome integrity, because of a disturbed balance in the telomere DNA-protein interactions. It is also unclear whether a change in telomere sequence may have any evolutionary advantage; in this regard, (Tran et al., 2015) suggested that the appearance of a "methylatable" cytosine in a G-rich telomere strand would raise the possibility of regulation by epigenetic modification.

WHAT ARE THE MOLECULAR REASONS FOR CHANGES IN THE TELOMERE MOTIFS?

To explain telomere sequence change, the first candidate is the template subunit of telomerase, telomerase RNA (TR). The previously identified TR from yeast and vertebrates belongs to a different group of transcripts, whose connecting feature was that they were transcribed by RNA polymerase II (Pol II)—in all but ciliates; this used to be the single exception from Pol II transcripts before publication of the land plant TR identification [reviewed in (Podlevsky and Chen, 2016)]. By using the relatively long telomere motif of *Allium* to look for its TR within the total RNA sequence data pool, Fajkus et al. (2019) showed that a previously characterized noncoding RNA involved in the stress reaction in *A. thaliana,* called AtR8, was indeed the telomerase RNA subunit (Wu et al., 2012; Fajkus et al., 2019). It was a transcript of RNA polymerase III (Pol III) containing the corresponding regulatory elements in its promoter structure. For a long time, researchers expected that plant TR would be so divergent that it would be impossible to identify it based on a homology search (Cifuentes-Rojas et al., 2011). However, a certain degree of similarity was successfully used to identify a common TR in several *Allium* species with comparative Blast. Surprisingly, sequence homology, the presence of the same regulatory elements, and a corresponding template region led to the identification of TRs in *Allium*, *Arabidopsis* and more than 70 other distantly related plants, including those with diverged

telomere motifs like *Genlisea*, *Cestrum*, and *Tulbaghia*. As far as we know, there is still no data on any algal TR, which would elucidate whether Pol III transcription of TR is a general feature for all plants or not. This work (Fajkus et al., 2019), based on CRISPR knock-out and other experiments, also showed that a previously identified telomerase RNA candidate in *A. thaliana* (Cifuentes-Rojas et al., 2011; Beilstein et al., 2012) was not a functional template subunit of telomerase, as was also demonstrated shortly after by (Dew-Budd et al., 2019). Assuming that the Pol II/Pol III dependency for TR transcription is a reliable evolutionary marker, future TR research in other main eukaryotic lineages will probably open new insights into the origin of eukaryotes. Telomerase genes and telomere sequences are unrecognized sources of information in this direction, and the finding of a Pol III dependent TR biogenesis pathway in ciliate and plant lineages may represent the first steps in this direction (Greider and Blackburn, 1989; Fajkus et al., 2019).

HOW DID CHROMOSOMES BECOME LINEAR?

A vast majority of prokaryotes contain circular chromosomes while linear chromosomes are the rule in eukaryotes. Therefore there are two possible scenarios in which either (i) linearization was performed by a primitive telomerase, preceding other processes which led to current linear chromosomal features and functions or (ii) linearization of a pre-eukaryotic circular chromosome was initially telomerase independent, but just before current eukaryotes diverged, a primitive telomerase started to occupy chromosome ends and became essential for the newly formed linear chromosomes (Nosek et al., 2006). Villasante et al. (2007) proposed an evolutionary scenario in which the breakage of the ancestral prokaryotic circular chromosome activated a transposition mechanism at DNA ends, allowing the formation of telomeres by a recombination-dependent replication mechanism: consequences of this hypothesis led to the surprising conclusion that eukaryotic centromeres were derived from telomeres.

Interestingly, the opposite process to linearization, i.e., formation of circular chromosomes (also termed ring chromosomes) has emerged from time to time during the evolution of eukaryotes, although being highly unstable. For example, in the case of *Amaranthus tuberculatus*, ring chromosomes appeared as a stress-induced response, carrying resistance against a herbicide (glyphosate); these extra ring chromosomes did not show hybridization with telomere probes in the karyotype analysis (Koo et al., 2018). The almost universal telomerase system and the exceptionality of circular chromosomes in eukaryotes do not allow us to support one hypothesis over the other. However, the recombinational machinery used in the alternative lengthening of telomeres (ALT), a telomerase-independent pathway, associated with certain human cancers (Zhang et al., 2019), is already present in prokaryotes. In addition, there is evidence of chromosome linearization occurring independently in distinct prokaryote lineages (Ferdows and Barbour, 1989; Nosek et al., 1995; Volff and Altenbuchner, 2000). Therefore, the hypothesis that the first linear eukaryotic chromosome (originating from a prokaryote ancestor) was telomerase-independent seems more likely. There are some examples that show that the telomerase-based system is not essential for telomere maintenance in all eukaryotes: retrotransposons in *Drosophila* telomeres, satellite repeats in *Chironomus*, another insect (Rubin, 1978; Biessmann and Mason, 2003), and ALT in telomerase-negative human cancers (Hu et al., 2016; Zhang et al., 2019). Yet, some of these systems may not be as different, and may perhaps share a common origin: in *Drosophila*, the telomere maintenance, based in retrotransposition, is not too distinct from the telomerase-based mechanism (Danilevskaya et al., 1998), leading to the hypothesis that the telomerase itself may be a former retrotransposon. But certainly, telomerase-negative plant species have not been discovered to date and all exceptions, in which the typical plant-type telomere was absent, were later shown to have different, but still telomerase-synthesized, motifs. Nevertheless, the ALT machinery is present in plants in parallel to the telomerase activity (Watson and Shippen, 2007; Ruckova et al., 2008). Interesting questions about the role of telomerase, telomeres and their maintenance in plant tumors arise from that. An attractive one is about the absence of metastasis in plants, despite the presence of ALT, perhaps related with plant tissue rigidity or different immune systems than in animals (Seyfried and Huysentruyt, 2013).

Although we are gaining increasing knowledge of telomere biology, we are still unable to explain the emergence of telomerase in eukaryotes. Current evidence supports the hypothesis that the emergence of eukaryotes together with their linear chromosomes, telomeres, and telomerase was related to the appearance of spliceosomal introns in archaeal hosts (Koonin, 2006; Fajkus et al., 2019). The similarity between TERT and other retroelements has been discussed for some time (Pardue et al., 1997). Remarkably, a relatively recent study showed that TERT, as a probable member of progeny group II introns, is sequentially close to *Penelope*-like element retrotransposons (Gladyshev and Arkhipova, 2007). But TERT is only one of the two essential telomerase components, and TR is, in its origin, even more enigmatic due to its low sequence conservation across all eukaryotes [see review (Podlevsky and Chen, 2016; Fajkus et al., 2019)].

CONCLUSION

At the beginning of the plant genomics era, the telomere sequence was considered almost changeless. The general conservation of telomeres and the telomerase system suggested that all plants may have the TTTAGGG plant-type telomere. The identification of unusual telomere sequences in complex plant

genomes, in many cases with giant C-values (such as in *Cestrum* and *Allium* sp.), was worth the effort, since the exceptionally long *Allium* telomere motif was the clue in looking for a genuine TR in land plants. The newly described TR in plants and further telomere/telomerase research in basal clades of algae might reveal valuable information about early evolution, therefore plant telomere research can significantly contribute to hypotheses on the emergence of eukaryotes.

AUTHOR CONTRIBUTIONS

VP and SG have contributed equally to the writing, editing, and preparation of this mini-review.

FUNDING

This work was supported by ERDF [project SYMBIT, reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477], EMBO Short-Term Fellowship 7368 to V.P., Spanish [CGL2016-75694-P (AEI/FEDER, UE)] and Catalan [grant number 2017SGR1116] governments. S.G. is the holder of a Ramón y Cajal contract (RYC-2014-16608).

ACKNOWLEDGMENTS

We thank the computational resources of the Virtual Organization Metacentrum (https://metavo.metacentrum.cz/en/index.html).

REFERENCES

Adams, S. P., Hartman, T. P., Lim, K. Y., Chase, M. W., Bennett, M. D., Leitch, I. J., et al. (2001). Loss and recovery of Arabidopsis-type telomere repeat sequences 5'-(TTTAGGG)(n)-3' in the evolution of a major radiation of flowering plants. *Proc. Biol. Sci.* 268 (1476), 1541–1546. doi: 10.1098/rspb.2001.1726

Beilstein, M. A., Brinegar, A. E., and Shippen, D. E. (2012). Evolution of the Arabidopsis telomerase RNA. *Front. Genet.* 3, 188. doi: 10.3389/fgene.2012.00188

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi: 10.1093/nar/27.2.573

Biessmann, H., and Mason, J. M. (2003). Telomerase-independent mechanisms of telomere elongation. *Cell Mol. Life Sci.* 60 (11), 2325–2333. doi: 10.1007/s00018-003-3247-9

Blackburn, E. H., and Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol.* 120 (1), 33–53. doi: 10.1016/0022-2836(78)90294-2

Blasco, M. A., Funk, W., Villeponteau, B., and Greider, C. W. (1995). Functional characterization and developmental regulation of mouse telomerase RNA. *Science* 269 (5228), 1267–1270. doi: 10.1126/science.7544492

Bremer, B., Bremer, K., Chase, M. W., Fay, M. F., Reveal, J. L., Soltis, D. E., et al. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161 (2), 105–121. doi: 10.1111/j.1095-8339.2009.00996.x

Byng, J. W., Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Judd, W. S., Mabberley, D. J., et al. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181 (1), 1–20. doi: 10.1111/boj.12385

Cervenak, F., Jurikova, K., Devillers, H., Kaffe, B., Khatib, A., Bonnell, E., et al. (2019). Identification of telomerase RNAs in species of the Yarrowia clade provides insights into the co-evolution of telomerase, telomeric repeats and telomere-binding proteins. *Sci. Rep.* 9 (1), 13365. doi: 10.1038/s41598-019-49628-6

Cifuentes-Rojas, C., Kannan, K., Tseng, L., and Shippen, D. E. (2011). Two RNA subunits and POT1a are components Arabidopsis telomerase. *Proc. Natl. Acad. Sci. U. S. A.* 108 (1), 73–78. doi: 10.1073/pnas.1013021107

Cox, A. V., Bennett, S. T., Parokonny, A. S., Kenton, A., Callimassia, M. A., and Bennett, M. D. (1993). Comparison of plant telomere locations using a Pcr-generated synthetic probe. *Ann. Bot.* 72 (3), 239–247. doi: 10.1006/anbo.1993.1104

Danilevskaya, O. N., Lowenhaupt, K., and Pardue, M. L. (1998). Conserved subfamilies of the *Drosophila* HeT-A telomere-specific retrotransposon. *Genetics* 148 (1), 233–242. https://www.genetics.org/content/148/1/233.short

Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A. Z., Robbens, S., et al. (2006). Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. *Proc. Natl. Acad. Sci. U. States America* 103 (31), 11647–11652. doi: 10.1073/pnas.0604795103

Dew-Budd, K., Cheung, J., Palos, K., Forsythe, E. S., and Beilstein, M. A. (2019). Evolutionary and biochemical analyses reveal conservation of the Brassicaceae telomerase ribonucleoprotein complex. *BioRxiv* 760785. doi: 10.1101/760785

Emery, H. S., and Weiner, A. M. (1981). An irregular satellite sequence is found at the termini of the linear extrachromosomal rDNA in Dictyostelium discoideum. *Cell* 26 (3 Pt 1), 411–419. doi: 10.1016/0092-8674(81)90210-5

Fajkus, J., Sýkorová, E., and Leitch, A. R. (2005). Telomeres in evolution and evolution of telomeres. *Chromosome Res.* 13 (5), 469–479. doi: 10.1007/s10577-005-0997-2

Fajkus, P., Peska, V., Sitova, Z., Fulneckova, J., Dvorackova, M., Gogela, R., et al. (2016). Allium telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)n is synthesized by telomerase. *Plant J.* 85 (3), 337–347. doi: 10.1111/tpj.13115

Fajkus, P., Peska, V., Zavodnik, M., Fojtova, M., Fulneckova, J., Dobias, S., et al. (2019). Telomerase RNAs in land plants. *Nucleic Acids Res.* 47 (18), 9842–9856. doi: 10.1093/nar/gkz695

Feng, J., Funk, W. D., Wang, S. S., Weinrich, S. L., Avilion, A. A., Chiu, C. P., et al. (1995). The RNA component of human telomerase. *Science* 269 (5228), 1236–1241. doi: 10.1126/science.7544491

Ferdows, M. S., and Barbour, A. G. (1989). Megabase-sized linear DNA in the bacterium Borrelia burgdorferi, the Lyme disease agent. *Proc. Natl. Acad. Sci. U. S. A.* 86 (15), 5969–5973. doi: 10.1073/pnas.86.15.5969

Fojtova, M., Sykorova, E., Najdekrova, L., Polanska, P., Zachova, D., Vagnerova, R., et al. (2015). Telomere dynamics in the lower plant Physcomitrella patens. *Plant Mol. Biol.* 87 (6), 591–601. doi: 10.1007/s11103-015-0299-9

Forstemann, K., Zaug, A. J., Cech, T. R., and Lingner, J. (2003). Yeast telomerase is specialized for C/A-rich RNA templates. *Nucleic Acids Res.* 31 (6), 1646–1655. doi: 10.1093/nar/gkg261

Frydrychova, R., Grossmann, P., Trubac, P., Vitkova, M., and Marec, F. (2004). Phylogenetic distribution of TTAGG telomeric repeats in insects. *Genome* 47 (1), 163–178. doi: 10.1139/g03-100

Fuchs, J., Brandes, A., and Schubert, I. (1995). Telomere sequence localization and karyotype evolution in higher-plants. *Plant Syst. Evol.* 196 (3-4), 227–241. doi: 10.1007/Bf00982962

Fulneckova, J., Hasikova, T., Fajkus, J., Lukesova, A., Elias, M., and Sykorova, E. (2012). Dynamic evolution of telomeric sequences in the green algal order Chlamydomonadales. *Genome Biol. Evol.* 4 (3), 248–264. doi: 10.1093/gbe/evs007

Fulneckova, J., Sevcikova, T., Fajkus, J., Lukesova, A., Lukes, M., Vlcek, C., et al. (2013). A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol. Evol.* 5 (3), 468–483. doi: 10.1093/gbe/evt019

Fulneckova, J., Sevcikova, T., Lukesova, A., and Sykorova, E. (2016). Transitions between the Arabidopsis-type and the human-type telomere sequence in green algae (clade Caudivolvoxa, Chlamydomonadales). *Chromosoma* 125 (3), 437–451. doi: 10.1007/s00412-015-0557-2

Garcia, S., Garnatje, T., and Kovarik, A. (2012). Plant rDNA database: ribosomal DNA loci information goes online. *Chromosoma* 121 (4), 389–394. doi: 10.1007/s00412-012-0368-7

Gladyshev, E. A., and Arkhipova, I. R. (2007). Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 104 (22), 9352–9357. doi: 10.1073/pnas.0702741104

Greider, C. W., and Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* 43 (2 Pt 1), 405–413. doi: 10.1016/0092-8674(85)90170-9

Greider, C. W., and Blackburn, E. H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* 337 (6205), 331–337. doi: 10.1038/337331a0

Higashiyama, T., Noutoshi, Y., Akiba, M., and Yamada, T. (1995). Telomere and LINE-like elements at the termini of the Chlorella chromosome I. *Nucleic Acids Symp. Ser.* (34), 71–72. https://europepmc.org/article/med/8841557

Hofr, C., Sultesova, P., Zimmermann, M., Mozgova, I., Prochazkova Schrumpfova, P., Wimmerova, M., et al. (2009). Single-Myb-histone proteins from Arabidopsis thaliana: a quantitative study of telomere-binding specificity and kinetics. *Biochem. J.* 419 (1), 221–228. doi: 10.1042/BJ20082195

Hu, Y., Shi, G., Zhang, L. C., Li, F., Jiang, Y. L., Jiang, S., et al. (2016). Switch telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX and DAXX. *Sci. Rep.* 6. doi: 10.1038/Srep32280

Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L., et al. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266 (5193), 2011–2015. doi: 10.1126/science.7605428

Koo, D. H., Molin, W. T., Saski, C. A., Jiang, J., Putta, K., Jugulam, M., et al. (2018). Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed Amaranthus palmeri. *Proc. Natl. Acad. Sci. U. S. A.* 115 (13), 3332–3337. doi: 10.1073/pnas.1719354115

Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Direct.* 1, 22. doi: 10.1186/1745-6150-1-22

Majerska, J., Schrumpfova, P. P., Dokladal, L., Schorova, S., Stejskal, K., Oboril, M., et al. (2017). Tandem affinity purification of AtTERT reveals putative interaction partners of plant telomerase *in vivo*. *Protoplasma* 254 (4), 1547–1562. doi: 10.1007/s00709-016-1042-3

Mason, J. M., Randall, T. A., and Capkova Frydrychova, R. (2016). Telomerase lost? *Chromosoma* 125 (1), 65–73. doi: 10.1007/s00412-015-0528-7

Meyne, J., Ratliff, R. L., and Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 86 (18), 7049–7053. doi: 10.1073/pnas.86.18.7049

Mravinac, B., Mestrovic, N., Cavrak, V. V., and Plohl, M. (2011). TCAGG, an alternative telomeric sequence in insects. *Chromosoma* 120 (4), 367–376. doi: 10.1007/s00412-011-0317-x

Muller, F., Wicky, C., Spicher, A., and Tobler, H. (1991). New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in Ascaris lumbricoides. *Cell* 67 (4), 815–822. doi: 10.1016/0092-8674(91)90076-b

Nosek, J., Dinouel, N., Kovac, L., and Fukuhara, H. (1995). Linear mitochondrial DNAs from yeasts: telomeres with large tandem repetitions. *Mol. Gen. Genet.* 247 (1), 61–72. doi: 10.1007/bf00425822

Nosek, J., Kosa, P., and Tomaska, L. (2006). On the origin of telomeres: a glimpse at the pre-telomerase world. *Bioessays* 28 (2), 182–190. doi: 10.1002/bies.20355

Novak, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinf.* 11, 378. doi: 10.1186/1471-2105-11-378

Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., et al. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. *BMC Biol.* 5, 28. doi: 10.1186/1741-7007-5-28

Olovnikov, A. M. (1973). Theory of marginotomy - incomplete copying of template margin in enzymic-synthesis of polynucleotides and biological significance of phenomenon. *J. Theor. Biol.* 41 (1), 181–190. doi: 10.1016/0022-5193(73)90198-7

One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574 (7780), 679–685. doi: 10.1038/s41586-019-1693-2

Pardue, M. L., Danilevskaya, O. N., Traverse, K. L., and Lowenhaupt, K. (1997). Evolutionary links between telomeres and transposable elements. *Genetica* 100 (1–3), 73–84. doi: 10.1007/978-94-011-4898-6_7

Peska, V., Sykorova, E., and Fajkus, J. (2008). Two faces of solanaceae telomeres: a comparison between Nicotiana and Cestrum telomeres and telomere-binding proteins. *Cytogenetic. Genome Res.* 122 (3-4), 380–387. doi: 10.1159/000167826

Peska, V., Fajkus, P., Fojtova, M., Dvorackova, M., Hapala, J., Dvoracek, V., et al. (2015). Characterisation of an unusual telomere motif (TTTTTTAGGG)(n) in the plant Cestrum elegans (Solanaceae), a species with a large genome. *Plant J.* 82 (4), 644–654. doi: 10.1111/tpj.12839

Peska, V., Sitova, Z., Fajkus, P., and Fajkus, J. (2017). BAL31-NGS approach for identification of telomeres *de novo* in large genomes. *Methods* 114, 16–27. doi: 10.1016/j.ymeth.2016.08.017

Peska, V., Mandakova, T., Ihradska, V., and Fajkus, J. (2019). Comparative dissection of three giant genomes: allium cepa, allium sativum, and allium ursinum. *Int. J. Mol. Sci.* 20 (3). doi: 10.3390/ijms20030733

Petracek, M. E., Lefebvre, P. A., Silflow, C. D., and Berman, J. (1990). Chlamydomonas telomere sequences are A+T-rich but contain three consecutive G-C base pairs. *Proc. Natl. Acad. Sci. U. S. A.* 87 (21), 8222–8226. doi: 10.1073/pnas.87.21.8222

Pfeiffer, V., and Lingner, J. (2012). TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. *PLoS Genet.* 8 (6). doi: 10.1371/journal.pgen.1002747

Pich, U., and Schubert, I. (1998). Terminal heterochromatin and alternative telomeric sequences in Allium cepa. *Chromosome Res.* 6 (4), 315–321. doi: 10.1023/a:1009227009121

Podlevsky, J. D., and Chen, J. J. (2016). Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* 13 (8), 720–732. doi: 10.1080/15476286.2016.1205768

Polotnianka, R. M., Li, J., and Lustig, A. J. (1998). The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities. *Curr. Biol.* 8 (14), 831–834. doi: 10.1016/S0960-9822(98)70325-2

Puizina, J., Weiss-Schneeweiss, H., Pedrosa-Harand, A., Kamenjarin, J., Trinajstic, I., Riha, K., and Schweizer, D. (2003). Karyotype analysis in Hyacinthella dalmatica (Hyacinthaceae) reveals vertebrate-type telomere repeats at the chromosome ends. *Genome* 46 (6), 1070–6. doi: 10.1139/g03-078

Qi, X. D., Li, Y., Honda, S., Hoffmann, S., Marz, M., Mosig, A., et al. (2013). The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res.* 41 (1), 450–462. doi: 10.1093/nar/gks980

Richards, E. J., and Ausubel, F. M. (1988). Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. *Cell* 53 (1), 127–136. doi: 10.1016/0092-8674(88)90494-1

Rubin, G. M. (1978). Isolation of a telomeric DNA sequence from Drosophila melanogaster. *Cold Spring Harb. Symp. Quant. Biol.* 42 Pt 2, 1041–1046. doi: 10.1101/sqb.1978.042.01.104

Ruckova, E., Friml, J., Schrumpfova, P. P., and Fajkus, J. (2008). Role of alternative telomere lengthening unmasked in telomerase knock-out mutant plants. *Plant Mol. Biol.* 66 (6), 637–646. doi: 10.1007/s11103-008-9295-7

Sekhri, K. (2014). Telomeres and telomerase: understanding basic structure and potential new therapeutic strategies targeting it in the treatment of cancer. *Postgrad. Med. J.* 60 (3), 303. doi: 10.4103/0022-3859.138797

Seyfried, T. N., and Huysentruyt, L. C. (2013). On the origin of cancer metastasis. *Crit. Rev. Oncog.* 18 (1-2), 43–73. doi: 10.1615/critrevoncog.v18.i1-2.40

Shibata, F., and Hizume, M. (2011). Survey of arabidopsis- and human-type telomere repeats in plants using fluorescence *in situ* Hybridisation. *Cytologia* 76 (3), 353–360. doi: 10.1508/cytologia.76.353

Somanathan, I., and Baysdorfer, C. (2018). A bioinformatics approach to identify telomere sequences. *Biotechniques* 65 (1), 20–25. doi: 10.2144/btn-2018-0057

Spiegel, J., Adhikari, S., and Balasubramanian, S. (2020). The structure and function of DNA G-quadruplexes. *Trends Chem.* 2 (2), 123–136. doi: 10.1016/j.trechm.2019.07.002

Sykorova, E., Lim, K. Y., Fajkus, J., and Leitch, A. R. (2003a). The signature of the Cestrum genome suggests an evolutionary response to the loss of (TTTAGGG) n telomeres. *Chromosoma* 112 (4), 164–172. doi: 10.1007/s00412-003-0256-2

Sykorova, E., Lim, K. Y., Chase, M. W., Knapp, S., Leitch, I. J., Leitch, A. R., et al. (2003b). The absence of Arabidopsis-type telomeres in Cestrum and closely related genera Vestia and Sessea (Solanaceae): first evidence from eudicots. *Plant J.* 34 (3), 283–291. doi: 10.1046/j.1365-313x.2003.01731.x

Sykorova, E., Lim, K. Y., Kunicka, Z., Chase, M. W., Bennett, M. D., Fajkus, J., et al. (2003c). Telomere variability in the monocotyledonous plant order Asparagales. *Proc. Biol. Sci.* 270 (1527), 1893–1904. doi: 10.1098/rspb.2003.2446

Tran, P. L., De Cian, A., Gros, J., Moriyama, R., and Mergny, J. L. (2013). Tetramolecular quadruplex stability and assembly. *Top. Curr. Chem.* 330, 243–273. doi: 10.1007/128_2012_334

Tran, T. D., Cao, H. X., Jovtchev, G., Neumann, P., Novak, P., Fojtova, M., et al. (2015). Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus Genlisea. *Plant J.* 84 (6), 1087–1099. doi: 10.1111/tpj.13058

Traut, W., Szczepanowski, M., Vitkova, M., Opitz, C., Marec, F., and Zrzavy, J. (2007). The telomere repeat motif of basal Metazoa. *Chromosome Res.* 15 (3), 371–382. doi: 10.1007/s10577-007-1132-3

Uzlikova, M., Fulneckova, J., Weisz, F., Sykorova, E., Nohynkova, E., and Tumova, P. (2017). Characterization of telomeres and telomerase from the single-celled eukaryote Giardia intestinalis. *Mol. Biochem. Parasitol.* 211, 31–38. doi: 10.1016/j.molbiopara.2016.09.003

vanSteensel, B., and deLange, T. (1997). Control of telomere length by the human telomeric protein TRF1. *Nature* 385 (6618), 740–743. doi: 10.1038/385740a0

Vilenchik, M. M., and Knudson, A. G. (2003). Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. U. S. A.* 100 (22), 12871–12876. doi: 10.1073/pnas.2135498100

Villasante, A., Abad, J. P., and Mendez-Lago, M. (2007). Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc. Natl. Acad. Sci. U. S. A.* 104 (25), 10542–10547. doi: 10.1073/pnas.0703808104

Visacka, K., Hofr, C., Willcox, S., Necasova, I., Pavlouskova, J., Sepsiova, R., et al. (2012). Synergism of the two Myb domains of Tay1 protein results in high affinity binding to telomeres. *J. Biol. Chem.* 287 (38), 32206–32215. doi: 10.1074/jbc.M112.385591

Vitales, D., D'Ambrosio, U., Galvez, F., Kovarik, A., and Garcia, S. (2017). Third release of the plant rDNA database with updated content and information on telomere composition and sequenced plant genomes. *Plant Syst. Evol.* 303 (8), 1115–1121. doi: 10.1007/s00606-017-1440-9

Vitkova, M., Kral, J., Traut, W., Zrzavy, J., and Marec, F. (2005). The evolutionary origin of insect telomeric repeats, (TTAGG)n. *Chromosome Res.* 13 (2), 145–156. doi: 10.1007/s10577-005-7721-0

Volff, J. N., and Altenbuchner, J. (2000). A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiol. Lett.* 186 (2), 143–150. doi: 10.1111/j.1574-6968.2000.tb09095.x

Wang, Y., Susac, L., and Feigon, J. (2019). Structural biology of telomerase. *Cold Spring Harb. Perspect. Biol.* 11, a032383. doi: 10.1101/cshperspect.a032383

Watson, J. M., and Shippen, D. E. (2007). Telomere rapid deletion regulates telomere length in Arabidopsis thaliana. *Mol. Cell. Biol.* 27 (5), 1706–1715. doi: 10.1128/Mcb.02059-06

Watson, J. D. (1972). Origin of Concatemeric T7 DNA. *Nature-New Biol.* 239 (94), 197–201. doi: 10.1038/newbio239197a0

Weiss, H., and Scherthan, H. (2002). Aloe spp.–plants with vertebrate-like telomeric sequences. *Chromosome Res.* 10 (2), 155–164. doi: 10.1023/a:1014905319557

Wellinger, R. J., Ethier, K., Labrecque, P., and Zakian, V. A. (1996). Evidence for a new step in telomere maintenance. *Cell* 85 (3), 423–433. doi: 10.1016/S0092-8674(00)81120-4

Wu, J., Okada, T., Fukushima, T., Tsudzuki, T., Sugiura, M., and Yukawa, Y. (2012). A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in Arabidopsis. *RNA Biol.* 9 (3), 302–313. doi: 10.4161/rna.19101

Zhang, J. M., Yadav, T., Ouyang, J., Lan, L., and Zou, L. (2019). Alternative lengthening of telomeres through two distinct break-induced replication pathways. *Cell Rep.* 26 (4), 955–95+. doi: 10.1016/j.celrep.2018.12.102

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

REVIEW
published: 25 February 2020
doi: 10.3389/fpls.2020.00200

Competition of Parental Genomes in Plant Hybrids

*Marek Glombik[1], Václav Bačovský[2], Roman Hobza[1,2] and David Kopecký[1]**

[1] *Institute of Experimental Botany, Czech Academy of Sciences, Centre of the Region Hana for Biotechnological and Agricultural Research, Olomouc, Czechia,* [2] *Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czechia*

Interspecific hybridization represents one of the main mechanisms of plant speciation. Merging of two genomes from different subspecies, species, or even genera is frequently accompanied by whole-genome duplication (WGD). Besides its evolutionary role, interspecific hybridization has also been successfully implemented in multiple breeding programs. Interspecific hybrids combine agronomic traits of two crop species or can be used to introgress specific loci of interests, such as those for resistance against abiotic or biotic stresses. The genomes of newly established interspecific hybrids (both allopolyploids and homoploids) undergo dramatic changes, including chromosome rearrangements, amplifications of tandem repeats, activation of mobile repetitive elements, and gene expression modifications. To ensure genome stability and proper transmission of chromosomes from both parental genomes into subsequent generations, allopolyploids often evolve mechanisms regulating chromosome pairing. Such regulatory systems allow only pairing of homologous chromosomes and hamper pairing of homoeologs. Despite such regulatory systems, several hybrid examples with frequent homoeologous chromosome pairing have been reported. These reports open a way for the replacement of one parental genome by the other. In this review, we provide an overview of the current knowledge of genomic changes in interspecific homoploid and allopolyploid hybrids, with strictly homologous pairing and with relaxed pairing of homoeologs.

Keywords: interspecific hybridization, genome stability, whole-genome duplication, allopolyploid, homoeologous recombination, chromosome pairing, fertility

OPEN ACCESS

Edited by:
Hanna Weiss-Schneeweiss, University of Vienna, Austria

Reviewed by:
Tae-Soo Jang, Chungnam National University, South Korea
Ales Kovarik, Academy of Sciences of the Czech Republic (ASCR), Czechia

****Correspondence:***
David Kopecký
kopecky@ueb.cas.cz

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

Received: *08 November 2019*
Accepted: *11 February 2020*
Published: *25 February 2020*

Citation:
Glombik M, Bačovský V, Hobza R and Kopecký D (2020) Competition of Parental Genomes in Plant Hybrids. Front. Plant Sci. 11:200. doi: 10.3389/fpls.2020.00200

INTRODUCTION

Interspecific hybridization merges genomes from two different species or even genera. Compared to animals, interspecific hybridization is much more common in plants and significantly contributes to plant speciation. In fact, many backcross hybrids probably remain undetected as they may be indistinguishable from parental species (Mallet, 2005). In plants, interspecific hybridization is frequently accompanied by whole-genome duplication (WGD), which is only rarely observed in animals. There is evidence that all angiosperms have undergone at least one round of WGD during their evolutionary history (Jiao et al., 2011; Ruprecht et al., 2017), and it is estimated that 30–70% of extant plant species are polyploids (Masterson, 1994). Hybridization is frequently accompanied by enhanced heterozygosity and hybrid vigor (e.g., growth and seed production), while WGD restores the fertility of a newly formed hybrids and contributes to the stabilization of the hybrid genome,

fixing both heterozygosity and new hybrid characters (Chen, 2010). Interspecific hybridization may also lead directly to speciation without polyploidization, but such homoploid hybrids are rare. To date, only a limited number of putative homoploid hybrid speciation events have been documented in flowering plants (Yakimowski and Rieseberg, 2014).

Apart from the evolutionary aspect of interspecific hybridization in plants, many major crops such as wheat, oilseed rape, banana, tobacco, coffee, and cotton also originated from hybridization of two or more species. Moreover, wide hybridization is frequently used in breeding programs to increase the global genetic diversity of the crop gene pool. This can be accomplished either by the creation of a new crop species, such as Triticale (hybrids of wheat and rye) and Festulolium (a hybrid of fescue and ryegrass) or by the introgression of specific loci from wild relative into the recipient crop. Allopolyploidy may generate intergenomic heterosis, which results in a competitive advantage over diploid progenitors (Comai, 2005), and it may mask deleterious recessive alleles and increase mutational robustness (Madlung, 2013). Newly formed hybrids often display broader adaptation to new environmental niches relative to their parents and may show greater ability to colonize disturbed and harsher habitats (Rieseberg et al., 2007; te Beest et al., 2012). This, in turn, may increase the invasiveness of newly formed hybrids (Pandit et al., 2011).

Despite the evolutionary success of allopolyploids, many newly developed hybrids display a phenomenon known as hybrid lethality. Dobzhansky (1936) proposed a model explaining the paradox of hybrid vigor (or evolutionary success) and incompatibility as the interactions of the parental genomes; this is now explained by the role of divergent small RNAs (Ha et al., 2009). New plant hybrids undergo multiple changes at the genome, chromosome, and gene levels. This includes genome downsizing, structural chromosome rearrangements, amplifications and/or reactivation of repetitive elements, modification of the gene expression patterns, and concerted evolution of multigene families (such as rDNAs immediately after the formation of the hybrid individuals). Furthermore, divergence of small RNAs in parental genomes may contribute to multiple heritable (epigenetic) changes, not associated with the changes in the DNA sequence (Bartel, 2004; Rao et al., 2009). The magnitude of all changes associated with hybridization is probably dependent on the degree of genome differences (Garsmeur et al., 2014; Bird et al., 2018).

Allopolyploidization has been intensively studied since its discovery, and several outstanding review papers concerning various aspects of allopolyploidization have been published in recent years (Mallet, 2005; Soltis and Soltis, 2009, 2012; Schubert and Lysak, 2011; Bottani et al., 2018; Li et al., 2019). In this review, we endeavor to provide a different view focusing on the fusion of two parental genomes and their competition in the single nucleus. With an increasing number of reports on the hybrid genome structure and evolution, it is evident that one of the parental genomes becomes dominant, and the other is rather submissive/suppressed in successive generations. This phenomenon called "genome dominance" (sometimes referred as "subgenome dominance") (Thomas et al., 2006) may include a plethora of the features including (a) an increase in genome size of the dominant genome and/or reduction of the submissive one, (b) elimination of chromosomes of the submissive genome, (c) replacement of chromosomes from the submissive genome by those of the dominant genome, (d) preferential loss (deletion) or silencing (by epigenetic processes) of alleles from the submissive genome resulting in homoeologous expression bias and the expression level dominance (ELD), and (e) preferential activation of transposable elements (TEs) and (f) global methylation changes (all these events are listed in **Figure 1**). Because the divergence in the mating system and parental conflict (acting as a barrier to hybridization) are discussed elsewhere (Brandvain and Haig, 2005), we will discuss the features of genome dominance at the level of genomic changes, chromosomal pairing, and gene regulation in relation to the stability of hybrid genomes.

GENOMIC CHANGES AND TE DYNAMICS IN ALLOPOLYPLOIDS

Genomic stress represented by interspecific hybridization (and polyploidization) affects genome reorganization, genetic changes, and epigenetic repatterning (histone modifications and DNA methylation). Global genomic changes go hand in hand with the establishment of genome dominance (Edger et al., 2017) and "new" gene expression, TEs reactivation, and TEs new insertions (Parisod et al., 2010b; Yaakov and Kashkush, 2012). Activation of TEs might occur immediately after a WGD event and seems to play a major role in all genomic changes in allopolyploids (Parisod et al., 2010a). In fact, the activation of TEs seems to be dependent on the qualitative/quantitative imbalance between the parental TE loads (**Figure 1A**). Such imbalance then may result in weak suppression of TEs and conflict between (sub)genomic elements within one nucleus (Parisod et al., 2012). Lim et al. (2004) suggested that greater imbalance is leading to stronger genome shock intensity. Recently, Mhiri et al. (2019) found that the proportion of new loci correlates with the extent of each TE load imbalance in different *Nicotiana* accessions and thus supported the influence of the genome shock intensity on TE activation. Genome dominance at the genomic level may be achieved by different regulation and composition of dominant TEs in the parental lineages (Freeling et al., 2012). The (re)activation of TEs includes re-patterning of DNA methylation in CG, CHG, and CHH motives, as well as changes in methylation of lysine residues in histone H3 which are typical for heterochromatin, namely, H3K9me2 and H3K27me1 or H3K27me2 (Lindroth et al., 2004; Fuchs and Schubert, 2012; Rando, 2012). Additionally, it is hypothesized that TE activity can contribute to genome size increase in a hybrid formation event. Conversely, unequal homologous recombination and illegitimate recombination may reduce the TE genome content (Bennetzen and Wang, 2014), as described for *Veju* elements in allohexaploid wheat (Kraitshtein et al., 2010). An extreme case of TE reactivation after interspecific hybridization may result in an increase of postzygotic lethality and seed abortion accompanied by arrested embryo development (Josefsson et al., 2006).

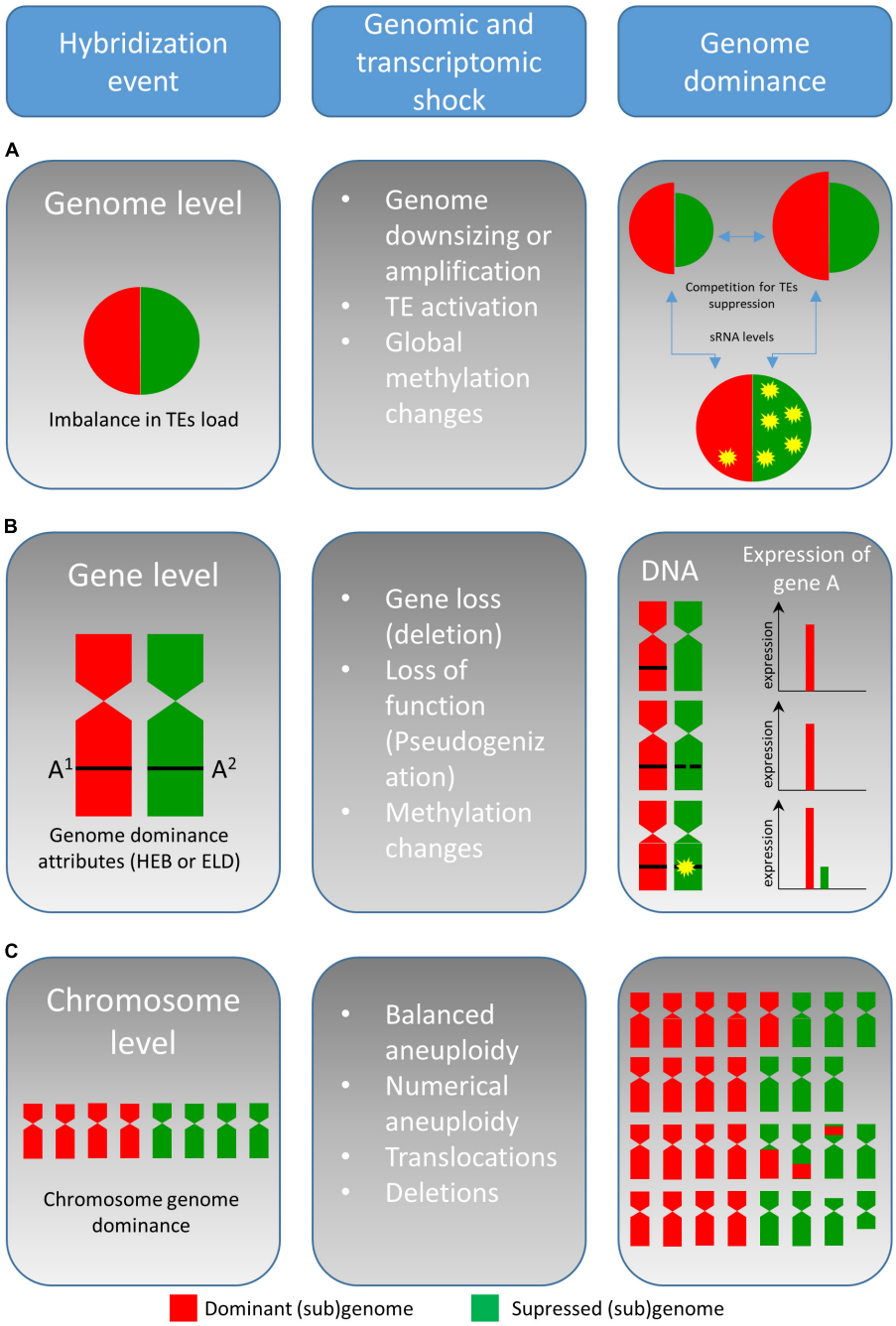

FIGURE 1 | Simplified model on the genetic and epigenetic changes associated with the interspecific hybridization in plants. The genome dominance described in plant interspecific hybrids can act on genome, chromosome, and gene level. Hybridization event, genome size, and TEs activation is affected greatly by imbalance(s) in TEs load and overall level of sRNA. The imbalance results in (sub)genome competition for TEs suppression **(A)**. Soon, after hybridization, functional conflicts between interacting genes impair the expression and due to gene imbalance, one genome becomes to be dominant at the expression level. Attributes facilitating the dominance at the transcriptome level are expression level dominance (ELD) and homoeolog expression bias (HEB) **(B)**. At the chromosomal level, genome dominance is affected by several factors (listed in section "Parental Chromosome Dominance") and by chromosomal aberrations, leading to imbalance in chromosome number **(C)**.

Newly resynthesized wheat, *Triticum aestivum* (with AABBDD genomes) represents a very good example of altered gene expression and altered DNA methylation affected by TEs activation after polyploidization. In this allopolyploid, the reactivated LTR *Wis 2-1A* retrotransposon deregulates expression of neighboring genes, driving the transcription of flanking regions (Kashkush et al., 2003). Similarly, the DNA methylation is altered in the case of *Veju* LTRs. About

∼43% of the tested insertion sites of *Veju* LTRs displayed hypo or hyper-methylation in the successive generations after allopolyploidization (Kraitshtein et al., 2010). In a later study, a similar heritable methylation repatterning was observed also for the *BARE-1* retroelement (Zhao et al., 2011). Given that both retroelements belong to Class I LTR (Wicker et al., 2007), it would be interesting to decipher how other retroelements and DNA transposons may affect the hybrid vigor in natural occurring hybrid populations. Interestingly, in naturally occurring wheat, the genome contains 2–10% less DNA than the sum of its putative diploid parents. The similar DNA elimination has been observed also in synthetic allopolyploids, showing such events can be studied in artificial system (Eilam et al., 2010). An intriguing question is how the parent-specific dominance of one genome can affect the properties of the newly formed hybrids and subsequent generations. In the allopolyploid *Nicotiana tabacum*, for example, the elimination of the paternally derived DNA was observed (Renny-Byfield et al., 2011).

Another unanswered question is how the alleles of paternal or maternal origin modulate the hybrids phenotype (vigor) and to which extent are paternal/maternal epigenetic mechanisms transmitted to the progeny. To answer such a complicated question, it will be necessary to study more hybrid systems and species (Song and Chen, 2015) and to formulate better hypotheses. We stress that high throughput sequencing methods now allow deeper understanding of the allopolyploidization process. Better understanding of the mechanisms underlying TE dynamics may allow development of new desired hybrids in near future.

GENOME AND NUCLEOLAR DOMINANCE

The gene number (orthologs or homoeologs) is duplicated after allopolyploidization (two diploid genomes are merged into a tetraploid individual). As a result of the duplicated genes after allopolyploidization, the hybrid genomes undergo extensive changes in gene expression, called "transcriptomic shock." This shock modifies the gene expression patterns, followed by unequal parental contribution and transgressive up- or down-regulation (Parisod et al., 2012; Yoo et al., 2013). From the long-term evolutionary perspective, there are three possible scenarios for ortholog genes (**Figure 1B**; Ma and Gustafson, 2005): (i) one copy becomes non-functional by genetic and/or epigenetic changes (non-functionalization), (ii) one copy acquires a novel, usually beneficial function, and is preserved by natural selection while the other copy retains the original function (neo-functionalization), or (iii) both copies become partially compromised by accumulations of mutations to the point where their total capacity is reduced to the level of the single-copy ancestral gene (sub-functionalization).

Parental genome which becomes dominant usually is the one that has lost fewer genes and therefore tends to express its genes to higher levels (Woodhouse et al., 2014). This phenomenon is called biased fractionation and is a result of functional conflicts between interacting genes and has been verified in many polyploid species (Emery et al., 2018). The dominance of a parental genome over the other genome at the gene expression level seems to be established in the first generations after hybridization and is transmitted over the generations (Schnable et al., 2011; Cheng et al., 2012; Zhu et al., 2017) or multiple rounds of polyploidy (Woodhouse et al., 2014). The genome dominance thus impairs the expression of various gene as demonstrated for rRNA genes (see below) and other genes, e.g., the genes for centromeric proteins. In fact, Talbert et al. (2002) used an antibody against *Arabidopsis thaliana* CenH3 and demonstrated that this antibody does not recognize the centromeres of *Arabidopsis arenosa* but recognizes the epitope in synthetic and natural allopolyploids originated from both these species. This clearly shows that the CenH3 gene from *A. thaliana* is dominant and its product is incorporated into the centromeres from both parental species. So, it seems that in a wide variety of hybrids one parent only recruits the kinetochore components (or its majority) which are expressed from the genes of the dominant parent. We hypothesize that the congruence between these components (kinetochore) and centromeric repeats probably determine the kinetochore function in newly formed hybrids and chromosome stability.

There are two different attributes which facilitate the genome dominance at the transcriptome level (**Figure 1B**), ELD, and homoeolog expression bias (HEB). While the ELD accounts for the overall expression of a single gene which resembles the expression level of one of its parents, the HEB represents preferential expression from one-parental allele (from one homoeolog) (Grover et al., 2012). Over the past decade, HEB and ELD were extensively studied in a number of plant allopolyploids (**Table 1**) (Bardil et al., 2011; Yoo et al., 2013; Edger et al., 2017; Wu et al., 2018). It should be stressed that majority of the genes can be expressed additively from both parental alleles even in cases with observed ELD and HEB (Chelaifa et al., 2013; Bertrand et al., 2015).

Generally, genome dominance is mediated by the upregulation of the dominant allele or downregulation of the submissive allele (Shi et al., 2012; Yoo et al., 2013; Combes et al., 2015; Zhu et al., 2017). If this is the case, then *trans* factors would likely be responsible for dominance, and the parent with more efficient *trans* factors would presumably be the one to establish the dominance (Hu and Wendel, 2019). Moreover, gene expression can be modified by epigenetic regulation mediated by TEs (see section "Genomic Changes and TE Dynamics in Allopolyploids") and small RNAs (McClintock, 1984; Ha et al., 2009; Lu et al., 2012). Specifically, siRNAs serve as guides for methyltransferases to perform *de novo* DNA methylation at CG, CHG, and CHH motifs (Haag and Pikaard, 2011). The methylation level of these motifs is often modified (either increased or reduced) after hybridization (Greaves et al., 2015; Zhu et al., 2017). Methylation of TEs within or close to a gene can lead to its silencing (Kim and Zilberman, 2014; Rodrigues and Zilberman, 2015). As an example, higher siRNA density at genes associated with TEs showed a negative effect on gene expression of the D genome in nascent allohexaploid wheat (Li et al., 2014). In monkeyflower allopolyploids, the dominantly expressed genome displayed a lower abundance of TEs and a

TABLE 1 | Examples of genome dominance in different hybrids and polyploids.

Species	Genome dominance	Genomes	References
Coffea arabica 2*n* = 4*x* = 44	ELD	*C. canephora* × *C. eugenioides*	Bardil et al., 2011
Brassica napus (natural) 2*n* = 4*x* = 38, AACC	HEB	*B. rapa* × *B. oleracea*	Ksiazczyk et al., 2011; Chalhoub et al., 2014
Brassica napus (resynthesized) 2*n* = 4*x* = 38, AACC	HEB, ELD	*B. rapa* × *B. oleracea*	Wu et al., 2018
Triticum aestivum 2*n* = 6*x* = 42, BBAADD	HEB	*T. turgidum* × *A. tauschii*	Pfeifer et al., 2014; Harper et al., 2016; Ramirez-Gonzalez et al., 2018
Triticum aestivum 2*n* = 6*x* = 42, BBAADD	ELD	AABB (*T. turgidum*) × DD (*A. tauschii*)	Li et al., 2014
Mimulus peregrinus (natural and resynthesized), ***Mimulus robertsii*** (resynthesized triploid hybrid) 2*n* = 6*x* = 92, GGLLLL	ELD and HEB	*M. guttatus* × *M. luteus*	Edger et al., 2017
Capsella bursa pastoris (*C. grandiflora* × *C. orientalis*) 2*n* = 4*x* = 32	ELD	*C. grandiflora* × *C. orientalis*	Kryvokhyzha et al., 2019
Gossypium hirsutum 2*n* = 4*x* = 52	HEB, ELD	*G. arboreum* × *G. raimondii* (?)	Yoo et al., 2013

ELD—expression level dominance, HEB—homoeolog expression bias.

reduced level of the CHH site methylations near genes (Edger et al., 2017). However, the role of methylation at the CG, CHG, and CHH motifs may be meager. In general, 24nt siRNAs are downregulated in the hybrids at loci in which parents differ in sRNA levels. Additionally, it has been shown that closely related parental lines had more additive expression of 21nt miRNAs and hybrids formed from more divergent lines display several non-additively expressed miRNAs, altering gene expression and phenotype of F1 population. Thus, it is believed that competition of parental hybrids for TEs regulation and overall levels of siRNAs (**Figure 1A**) are important for the hybrid vigor, and hybrid expression changes (reviewed in Groszmann et al., 2013). A comparative study of 3DL chromosome arms from wheat and its progenitor *Aegilops tauschii* revealed that methylation is responsible for only 11% of genes with altered gene expression. It seems that a reduced gene expression correlates more tightly with higher compaction and reduced accessibility of chromatin of this particular wheat chromosome arm (Lu et al., 2019). This was not surprising because gene expression changes have been linked with the spatial organization of chromatin and gene repositioning (Lanctot et al., 2007; Gonzalez-Sandoval and Gasser, 2016). Recently, Wang et al. (2018) described the dynamics of 3D genome architecture after polyploidization, showing the reorganization of topologically associated domains in allopolyploid cotton. In addition, they identified inter-subgenomic chromatin interactions between homoeologous gene pairs. Linking these interactions with expression of homoeologous gene pairs showed that some genes with extreme expression bias are associated with low number of chromatin interactions. Increased compactness, however, did not correlate with gene expression changes in allotetraploid *Arabidopsis suecica* (a hybrid of *A. thaliana* × *A. arenosa*) (Zhu et al., 2017).

An important manifestation of the genome dominance in plant hybrids is nucleolar dominance (ND) first described by Navashin (1934). Nucleoli are sites of transcription of rRNA, which is participating in the ribosome assembly. A typical feature of ND is that ribosomal genes inherited from one (dominant) parental species are expressed and those inherited from the other parent are silenced (Neves et al., 1997; Lawrence et al., 2004; Idziak and Hasterok, 2008; Ksiazczyk et al., 2011; Borowska-Zuchowska et al., 2016). Expression analysis of the rRNA transcription in two yeast strains revealed that the proportion of active rDNA is regulated by a dosage control mechanism (French et al., 2003). The level of rDNA expression is about the same in both strains, even though they significantly differ in the number of copies. Such dosage control is a result of higher occupancy of Pol I per gene, and the occupancy itself is linked to epigenetic marks regulating it (French et al., 2003). Increased methylation of CHG and CHH motifs and histone marks (H3K27me3 and H3K9me2) was observed for A genome NOR (nucleolar organizing region) loci in synthetic allotetraploid wheat, leading to their silencing and further elimination in later generations (Guo and Han, 2014). Similarly, epigenetic modifications such as DNA and histone methylations at lysine residues (H3K9me2 and H3K4me3) resulted in the silencing of NORs of *A. thaliana* genome in allotetraploid *A. suecica* (Lawrence et al., 2004). Conversely, reduction of CG and CHG DNA methylation probably influenced the reactivation of NOR of the submissive genome in *Tragopogon mirus* (Dobesova et al., 2015). Similarly, the deletion of the NOR region to about 4% of the normal length does not significantly decrease the level of rDNA expression in *T. mirus* (Dobesova et al., 2015). The rDNA loci frequently differ in numbers between genomes in allopolyploids, e.g., in *Triticale*, one locus is present in the rye genome, while the wheat genome possesses two major and several minor loci. Interestingly, all rRNAs are transcribed from wheat loci (Neves et al., 1997) but not necessarily from rye. Similarly, allotetraploid *A. suecica* expresses rRNA from *A. arenosa* while the *A. thaliana* rDNA genes are silenced. However, a backcross of *A. suecica* to *A. thaliana* reverts such patterns, and the *A. arenosa* rDNA cluster becomes suppressed (Chen et al., 1998). One may expect that the ND is correlated with the number of rDNA loci of the parental genomes. Nevertheless, translocation of the short arm of rye chromosome containing nucleolus organizing regions (NORs) to its homoeologous wheat chromosome made the rye NOR co-dominant with wheat NORs (Vieira et al., 1990). Thus, the number of these loci itself is presumably not the exclusive driving force for the ND, and another mechanism operates. In fact, other studies support the hypothesis that selective silencing of rRNA genes depends on the position on the chromosome and sequences that surround them (Chandrasekhara et al., 2016; Mohannath et al., 2016). Yet, another factor in ND seems to be also parent-allele specific

origin. In allopolyploids such as *Tragopogon* L., *Cardamine* L., and *Senecio* L., NOR is preferentially expressed from the maternal genome. On the other hand, some hybrids display a bias in the expression of rRNA genes toward the same genome in reciprocal crosses. In hybrid *Rosa agrestis* and *Rosa rubiginosa*, the expression dominance of the *Canina* type units was observed even if they were underrepresented in copy numbers (Khaitová et al., 2010; Herklotz et al., 2018). Overall, in some systems, one-genome-type rDNA is more vulnerable to repression, and in others, it is prone to be dominant. Such vulnerability is clearly linked with epigenetic marks (Costa-Nunes and Pontes, 2013; He and Deng, 2013). Nevertheless, it remains to be determined how such ND is established, and whether one parent always becomes dominant or factors such as neighbor regulatory sequences, chromosomes positioning, and chromatin organization during interphase play a role.

CHROMOSOME PAIRING IN HYBRIDS

Homoploid hybrids are rare in plants (Yakimowski and Rieseberg, 2014). This may well be due to problems in meiosis when homoeologous chromosomes fail to pair as bivalents, and random segregation of univalents produces non-functional gametes. Thus, WGD of sterile diploid F_1 hybrids is necessary for fertility restoration.

The homoploid hybrids occur in nature only sporadically. For this reason, we focus here on allopolyploids and provide only several known examples of chromosome rearrangements in artificial homoploid hybrids. Allopolyploid hybrids possess three or more chromosome sets from two or more species, e.g., *A. suecica* (Novikova et al., 2017). Assuming that the basic chromosome number is the same in both parents, each chromosome in an allopolyploid can pair either with its homolog or with one of the two homoeologs. Theoretically, in an allotetraploid, the ratio of homologous vs. homoeologous pairing should be 1:2, but very few hybrids exhibit such a ratio. The pairing bias depends on the level of the DNA sequence divergence of two parental genomes. Immediately or soon after initial hybridization, newly formed allopolyploid lineages often establish a system that may hamper the pairing of homoeologs. One of the such best-studied systems is *Ph1* (Pairing homoeologous 1) presented in polyploid wheats (Sears and Okamoto, 1958). Nevertheless, after 60 years of extensive research, the mode of action of *Ph1* is still not completely understood. There are two competing theories to its actual location and nature (Rey et al., 2017, 2018; Rawale et al., 2019). It is assumed that the *Ph1* locus contains a cluster of defective cyclin-dependent kinases (CDKs) and S-adenosyl methionine-dependent methyltransferase (SAM-MTases) genes and inserted paralog of the ZIP4 (Knight et al., 2010; Greer et al., 2012; Martin et al., 2014, 2017). Interestingly, *Ph1*, once introgressed from wheat to the relative species, has an ability to modify chromosome pairing of the host genome (Lukaszewski and Kopecky, 2010).Coupled with *Ph1*, a similar chromosome pairing control systems have been found in *Brassica* allopolyploids (*PrBn*), oats, fescues, and many other allopolyploids (Jenczewski and Alix, 2004). Nevertheless, the presence of a regulating system does not always preclude the elimination of the submissive genome as documented in triticales (allopolyploids of bread wheat and rye) (Tsunewaki, 1964; Orellana et al., 1984; Lukaszewski et al., 1987).

The mechanism of reduced chromosome pairing in allopolyploids is also not fully understood. In wheat-rye hybrids, a relationship between the behavior of telomeres and the success of chromosome pairing has been observed by Naranjo (2014), who reported that reduced pairing of rye chromosomes in wheat appeared to be a consequence of disturbed migration of rye telomeres into the leptotene bouquet. We have observed that the problem of aberrant rye telomeres is not limited to the initial stages of meiosis but may be systemic in nature. The frequency of out-of-bouquet rye telomere positions at leptotene was virtually identical to that in the nuclei of somatic cells, and that in turn correspond to the rate of chromosome pairing (Pernickova et al., 2019a,b).

Besides allopolyploids with evolved pairing regulators, several hybrids with extensive homoeologous chromosome pairing have been reported in well-established allopolyploids and synthetic F_1 hybrids (either homoploids or allopolyploids). Some allopolyploids with a molecular mechanism of diploid-like pairing behavior, when resynthesized from putative progenitors, display disrupted meiosis with homoeologous chromosomes pairing as described in resynthesized *Brassica napus* (Gaeta and Pires, 2010). Xiong et al. (2011) studied the karyotypes of resynthesized *B. napus* and found that the aneuploidy rate was increasing for ten successive generations. In addition, the authors found frequent homoeologous chromosome pairing and replacement of chromosomes of one parental species by the other (prevalence of C-genome). Intriguingly, two lines retaining the expected original chromosome constitution had the highest seed yield, and thus, the selection against aberrant chromosome constitutions with reduced fertility may be expected. Those lines which lack the control over the regular meiotic division might be the key factor in the establishment of natural *B. napus* with the stabilized genome.

PARENTAL CHROMOSOME DOMINANCE

The mechanism(s) responsible for the chromosomal genome dominance remains to be fully determined. There are several features that may hypothetically facilitate the replacement of chromosomes from the submissive genome by those from the dominant genome (**Figure 1C**). Possible scenarios involve differences in male meiosis, female meiotic drive, variation in the proliferation of pollen tube, germination, and fertility of pollen grains and seed yield. A combination of these processes is likely involved; however, technical issues hamper the ability to discriminate among individual factors clearly. The submissive genome may show reduced chromosome pairing; this leads to the formation of univalents, and univalents are frequently lost during meiosis. However, the substitution of chromosomes from one genome by homoeologs chromosomes from the

other, and not just uniparental chromosome elimination, is far more complicated. Such chromosomal substitution would require non-disjunction and unidirectional movement of both homologs from the dominant genome (associated in the bivalent) to one pole of the meiotic cell (coupled with concomitant elimination of the homoeolog). This does not appear very likely. Random migration of univalents to daughter nuclei in anaphase I during meiosis would probably be more likely to produce such single chromosome substitutions, coupled with selection for euploid gametes.

In our earlier studies, we have observed chromosomal dominance in all cultivars selected from *Lolium multiflorum* × *Festuca pratensis* hybrids. These hybrids exhibit the prevalence of the *Lolium* chromosomes (Kopecky et al., 2006). Zwierzykowski et al. (2006, 2011) conducted a study over eight successive generations of such hybrids and observed a slow but consistent gradual replacement of the *Festuca* by *Lolium* chromosomes. Such chromosome-level genome dominance appears to be present also in other hybrid systems. In homoploid onion hybrids of *Allium cepa* × *Allium roylei*, the *A. roylei* genome appears to replace the *A. cepa* genome in successive generations. van Heusden et al. (2000) reported that, on average, the *A. cepa*-specific markers were not amplified in 28% of the F_2 plants, while only 16% of the F_2 plants did not display amplification of *A. roylei*-specific markers. Thus, the contribution of *A. cepa* and *A. roylei* alleles in the F_2 population was 44 and 56%. Similar genome dominance on the chromosome level has been observed in hybrids of *Alstroemeria aurea* × *Alstroemeria inodora*, *Gasteria lutzii* × *Aloe aristata* and in various lily hybrids (Takahashi et al., 1997; Kamstra et al., 1999; Karlov et al., 1999; Khan et al., 2009).

The role of meiotic pairing in chromosomal genome dominance is speculative. In an F_1 hybrid originating from crossing autotetraploid *L. multiflorum* and allotetraploid *Ferocactus glaucescens*, we observed 101 univalents of the former and 161 univalents of the latter (Kopecky et al., 2009). On contrary, in tetraploid hybrids *Lolium perenne* × *Formica pratensis*, Zwierzykowski et al. (2008) found that *Lolium* univalents were more frequent. Thus, the role of male gametic selection does not seem to play a significant role. In contrast, asymmetrical female meiosis offers an unprecedented opportunity to modify the genomic composition in favor of the dominant genome. Far from plant hybrids, Akera et al. (2017) observed uneven positioning of the parental chromosomes on the meiotic spindle in hybrid mice. The positioning of chromosomes from the dominant genome toward the egg cell and of chromosomes from the submissive genome toward the polar body was more frequent than the reciprocal configuration. In female meiosis, only one of the four products forms the embryo sack and is transmitted to the next generation. Given that only one product of meiosis proceeds, preferential positioning of specific chromosomes in the first meiotic division can easily explain different rates of chromosome transmission. Shifts in proportions of the parental genomes in favor of the dominant one. The candidate molecular driver in this case is CDC42, which is signaling unequal regulation of microtubule tyrosination. This unequal tyrosination is probably caused by the difference in the copy number of kinetochores between the two genomes. Hence, the dominant genome (having more copies of centromeric repeats and kinetochore proteins) is preferentially transmitted to egg cell while the submissive one into the polar body (Chmatal et al., 2014).

DISCUSSION AND OUTLOOK

Our knowledge about the genome dominance and its consequences for the structure and evolution of hybrid genomes is still limited, and further studies are required to shed more light on this biologically intriguing phenomenon. A systematic and complex approach should be applied to several systems to uncover potential linkages among different attributes and their specific roles in genome dominance. In some plant systems, one genome dominates on all the possible levels such as seen in *Lolium* × *Festuca* hybrids: *Lolium* chromosomes predominate in consecutive generations (Zwierzykowski et al., 2006), *Lolium* alleles are overexpressed relative to those of *Festuca* (Stoces et al., 2016), seed yield is higher in plants with higher proportions of *Lolium* chromatin (Kubota et al., 2019), and rRNA is exclusively transcribed from the *Lolium* variant (Mahelka and Kopecky, unpublished). On the other hand, analyses of other allopolyploids often produce opposite or conflicting results. Perhaps the patterns of interactions in individual hybrids are specific to combinations of involved parents and allelic variants individual hybrids.

The fate of newly established allopolyploids and the dominance of one genome over the other is, so far, not fully predictable. Studies on *Brassica* indicate that lines with additive karyotypes, without any rearrangements, are the most fertile. Hence, they can be considered evolutionarily the most successful. In the situation of competition between the genomes, one could reasonably expect that the dominant genome would trigger elimination (or replacement) of the other genome and revert the hybrid to the parental form. In this situation, it would be in the best interest of the submissive genome to establish as quickly as possible a mechanism favoring homologous pairing over that of homoeologous. Consequently, if fertility (seed set and seed yield) is highest in the additive karyotype (both parental genomes present), it would be beneficial to establish control pairing mechanism for both genomes. In *B. napus*, an amphiploid of *Brassica rapa* (A genome) and *Brassica oleracea* (C genome), meiotic regulator *PrBn* is located on the linkage group 15 in genome C, the genome that clearly dominates on the chromosome and transcriptome levels in the first generations of resynthesized lines (Liu et al., 2006). In wheat, the B genome carrying the *Ph1* locus on its 5BL chromosome arm expresses its alleles at the same or very similar level as the other two genomes (Ramirez-Gonzalez et al., 2018). Although *PrBn* and *Ph1* models are so far only two best-known examples, further analyses are needed to shed more light on these phenomena. Another interesting question is if and how is regulated meiotic pairing in gymnosperms and gymnosperm hybrids (Sedel'nikova et al., 2011).

We stress that with an increased understanding of the principles and mechanisms of genome dominance in interspecific hybrids, it may be possible to predict which crosses will result in

stable introgression of desired traits in plant breeding. It is still largely unknown if the genome dominance is fully deterministic or if it can be manipulated by the external conditions. Detailed molecular analyses of many model species and a complex view are needed to understand fully the complex aspects of hybrid formation and polyploidy evolution. We expect that future implication of CRISPR/Cas9 technology to edit methylation states, TE insertions, or species-specific centromere binding proteins can lead to the development of stable hybrids without presumable suppression of one of the genomes.

AUTHOR CONTRIBUTIONS

MG, VB, RH, and DK wrote the first draft of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

FUNDING

This research was funded by the Czech Science Foundation (grant awards 17-13853S, 20-10019S, 16-08698S, 19-05445S, and 18-06147S) and by the European Regional Development Fund OPVVV Project "Plants as a tool for sustainable development" number CZ.02.1.01/0.0/0.0/16_019/0000827 supporting Excellent Research at CRH.

ACKNOWLEDGMENTS

We would like to thank Prof. Adam J. Lukaszewski (University of California, Riverside) for his critical reading and valuable comments. We would like to thank Francesco Muto for English-language correction and to the reviewers for their thoughtful comments and efforts toward improving our manuscript.

REFERENCES

Akera, T., Chmatal, L., Trimm, E., Yang, K., Aonbangkhen, C., Chenoweth, D. M., et al. (2017). Spindle asymmetry drives non-Mendelian chromosome segregation. *Science* 358, 668–672. doi: 10.1126/science.aan0092

Bardil, A., de Almeida, J. D., Combes, M. C., Lashermes, P., and Bertrand, B. (2011). Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. *New Phytol.* 192, 760–774. doi: 10.1111/j.1469-8137.2011.03833.x

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5

Bennetzen, J. L., and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* 65, 505–530. doi: 10.1146/annurev-arplant-050213-035811

Bertrand, B., Bardil, A., Baraille, H., Dussert, S., Doulbeau, S., Dubois, E., et al. (2015). The greater phenotypic homeostasis of the allopolyploid *Coffea arabica* improved the transcriptional homeostasis over that of both diploid parents. *Plant Cell Physiol.* 56, 2035–2051. doi: 10.1093/pcp/pcv117

Bird, K. A., VanBuren, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi: 10.1111/nph.15256

Borowska-Zuchowska, N., Kwasniewski, M., and Hasterok, R. (2016). Cytomolecular analysis of ribosomal DNA evolution in a natural allotetraploid *Brachypodium hybridum* and its putative ancestors - dissecting complex repetitive structure of intergenic spacers. *Front. Plant Sci.* 7:1499. doi: 10.3389/fpls.2016.01499

Bottani, S., Zabet, N. R., Wendel, J. F., and Veitia, R. A. (2018). Gene expression dominance in allopolyploids: hypotheses and models. *Trends Plant Sci.* 23, 393–402. doi: 10.1016/j.tplants.2018.01.002

Brandvain, Y., and Haig, D. (2005). Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. *Am. Nat.* 166, 330–338. doi: 10.1086/432036

Chalhoub, B., Denoeud, F., Liu, S. Y., Parkin, I. A. P., Tang, H. B., Wang, X. Y., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Gene. Dev.* 30, 177–190. doi: 10.1101/gad.273755.115

Chelaifa, H., Chague, V., Chalabi, S., Mestiri, I., Arnaud, D., Deffains, D., et al. (2013). Prevalence of gene expression additivity in genetically stable wheat allohexaploids. *New Phytol.* 197, 730–736. doi: 10.1111/nph.12108

Chen, Z. J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15, 57–71. doi: 10.1016/j.tplants.2009.12.003

Chen, Z. J., Comai, L., and Pikaard, C. S. (1998). Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14891–14896. doi: 10.1073/pnas.95.25.14891

Cheng, F., Wu, J., Fang, L., Sun, S. L., Liu, B., Lin, K., et al. (2012). Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* 7:e36442. doi: 10.1371/journal.pone.0036442

Chmatal, L., Gabriel, S. I., Mitsainas, G. P., Martinez-Vargas, J., Ventura, J., Searle, J. B., et al. (2014). Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in Mice. *Curr. Biol.* 24, 2295–2300. doi: 10.1016/j.cub.2014.08.017

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Combes, M. C., Hueber, Y., Dereeper, A., Rialle, S., Herrera, J. C., and Lashermes, P. (2015). Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biol. Evol.* 7, 1110–1121. doi: 10.1093/gbe/evv057

Costa-Nunes, P., and Pontes, O. (2013). "Chromatin and small RNA regulation of nucleolar dominance," in *Polyploid and Hybrid Genomics*, eds Z. J. Chen, and J. A. Birchler (Hoboken, NJ: Wiley), doi: 10.1002/9781118552872.ch18

Dobesova, E., Malinska, H., Matyasek, R., Leitch, A. R., Soltis, D. E., Soltis, P. S., et al. (2015). Silenced rRNA genes are activated and substitute for partially eliminated active homeologs in the recently formed allotetraploid, *Tragopogon mirus* (Asteraceae). *Heredity* 114, 356–365. doi: 10.1038/hdy.2014.111

Dobzhansky, T. (1936). Studies on Hybrid Sterility. II. Localization of sterility factors in *Rosophila pseudoobscura* hybrids. *Genetics* 21, 113–135.

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y. W., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid Monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010

Eilam, T., Anikster, Y., Millet, E., Manisterski, J., and Feldman, M. (2010). Genome size in diploids, allopolyploids, and autopolyploids of mediterranean triticeae. *J. Bot.* 2010:341380. doi: 10.1155/2010/341380

Emery, M., Willis, M. M. S., Hao, Y., Barry, K., Oakgrove, K., Peng, Y., et al. (2018). Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS Genet.* 14:e1007267. doi: 10.1371/journal.pgen.1007267

Freeling, M., Woodhouse, M. R., Subramaniam, S., Turco, G., Lisch, D., and Schnable, J. C. (2012). Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr. Opin. Plant Biol.* 15, 131–139. doi: 10.1016/j.pbi.2012.01.015

French, S. L., Osheim, Y. N., Cioci, F., Nomura, M., and Beyer, A. L. (2003). In exponentially growing *Saccharomyces cerevisiae* cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. *Mol. Cell. Biol.* 23, 1558–1568. doi: 10.1128/MCB.23.5.1558-1568.2003

Fuchs, J., and Schubert, I. (2012). "Chromosomal distribution and functional interpretation of epigenetic histone marks in plants," in *Plant Cytogenetics:*

Genome structure and chromosome function, Vol. 4, eds H. Bass, and J. Birchler (New York, NY: Springer), 231–253. doi: 10.1007/978-0-387-70869-0_9

Gaeta, R. T., and Pires, J. C. (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol.* 186, 18–28. doi: 10.1111/j.1469-8137.2009.03089.x

Garsmeur, O., Schnable, J. C., Almeida, A., Jourda, C., D'Hont, A., and Freeling, M. (2014). Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol.* 31, 448–454. doi: 10.1093/molbev/mst230

Gonzalez-Sandoval, A., and Gasser, S. M. (2016). On TADs and LADs: spatial control over gene expression. *Trends Genet.* 32, 485–495. doi: 10.1016/j.tig.2016.05.004

Greaves, I. K., Gonzalez-Bayon, R., Wang, L., Zhu, A. Y., Liu, P. C., Groszmann, M., et al. (2015). Epigenetic changes in hybrids. *Plant Physiol.* 168, 1197–1205. doi: 10.1104/pp.15.00231

Greer, E., Martin, A. C., Pendle, A., Colas, I., Jones, A. M. E., Moore, G., et al. (2012). The Ph1 locus suppresses Cdk2-type activity during premeiosis and meiosis in wheat. *Plant Cell* 24, 152–162. doi: 10.1105/tpc.111.094771

Groszmann, M., Greaves, I. K., Fujimoto, R., Peacock, W. J., and Dennis, E. S. (2013). The role of epigenetics in hybrid vigour. *Trends Genet.* 29, 684–690. doi: 10.1016/j.tig.2013.07.004

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196, 966–971. doi: 10.1111/j.1469-8137.2012.04365.x

Guo, X., and Han, F. P. (2014). Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. *Plant Cell* 26, 4311–4327. doi: 10.1105/tpc.114.129841

Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K. D., Chapman, E. J., et al. (2009). Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17835–17840. doi: 10.1073/pnas.0907003106

Haag, J. R., and Pikaard, C. S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* 12, 483–492. doi: 10.1038/nrm3152

Harper, A. L., Trick, M., He, Z. S., Clissold, L., Fellgett, A., Griffiths, S., et al. (2016). Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat. *Plant Biotechnol. J.* 14, 1207–1214. doi: 10.1111/pbi.12486

He, G., and Deng, X.-W. (2013). "Chromatin and gene expression mechanisms in hybrids," in *Polyploid and Hybrid Genomics*, eds Z. J. Chen, and J. A. Birchler (Alexandria, VA: NSF), doi: 10.1002/9781118552872.ch20

Herklotz, V., Kovarik, A., Lunerova, J., Lippitsch, S., Groth, M., and Ritz, C. M. (2018). The fate of ribosomal RNA genes in spontaneous polyploid dogrose hybrids *Rosa* L. sect. *Caninae* (DC.) Ser. exhibiting non-symmetrical meiosis. *Plant J.* 94, 77–90. doi: 10.1111/tpj.13843

Hu, G. J., and Wendel, J. F. (2019). *Cis-trans* controls and regulatory novelty accompanying allopolyploidization. *New Phytol.* 221, 1691–1700. doi: 10.1111/nph.15515

Idziak, D., and Hasterok, R. (2008). Cytogenetic evidence of nucleolar dominance in allotetraploid species of *Brachypodium*. *Genome* 51, 387–391. doi: 10.1139/G08-017

Jenczewski, E., and Alix, K. (2004). From diploids to allopolyploids: the emergence of efficient pairing control genes in plants. *Crit. Rev. Plant Sci.* 23, 21–45. doi: 10.1080/07352680490273239

Jiao, Y. N., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16, 1322–1328. doi: 10.1016/j.cub.2006.05.045

Kamstra, S. A., Kuipers, A. G. J., De Jeu, M. J., Ramanna, M. S., and Jacobsen, E. (1999). The extent and position of homoeologous recombination in a distant hybrid of Alstroemeria: a molecular cytogenetic assessment of first generation backcross progenies. *Chromosoma* 108, 52–63. doi: 10.1007/s004120050351

Karlov, G. I., Khrustaleva, L. I., Lim, K. B., and van Tuyl, J. M. (1999). Homoeologous recombination in 2n-gametes producing interspecific hybrids of Lilium (Liliaceae) studied by genomic in situ hybridization (GISH). *Genome* 42, 681–686. doi: 10.1139/g98-167

Kashkush, K., Feldman, M., and Levy, A. A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* 33, 102–106. doi: 10.1038/ng1063

Khaitová, L., Werlemark, G., Nybom, H., and Kovarik, A. (2010). Frequent silencing of rDNA loci on the univalent-forming genomes contrasts with their stable expression on the bivalent-forming genomes in polyploid dogroses (*Rosa* sect. *Caninae*). *Heredity* 104, 113–120. doi: 10.1038/hdy.2009.94

Khan, N., Barba-Gonzalez, R., Ramanna, M. S., Visser, R. G. F., and Van Tuyl, J. M. (2009). Construction of chromosomal recombination maps of three genomes of lilies (Lilium) based on GISH analysis. *Genome* 52, 238–251. doi: 10.1139/G08-122

Kim, M. Y., and Zilberman, D. (2014). DNA methylation as a system of plant genomic immunity. *Trends Plant Sci.* 19, 320–326. doi: 10.1016/j.tplants.2014.01.014

Knight, E., Greer, E., Draeger, T., Thole, V., Reader, S., Shaw, P., et al. (2010). Inducing chromosome pairing through premature condensation: analysis of wheat interspecific hybrids. *Funct. Integr. Genomics* 10, 603–608. doi: 10.1007/s10142-010-0185-0

Kopecky, D., Bartos, J., Zwierzykowski, Z., and Dolezel, J. (2009). Chromosome pairing of individual genomes in tall fescue (Festuca arundinacea Schreb.), its progenitors, and hybrids with Italian ryegrass (Lolium multiflorum Lam.). *Cytogenet. Genome Res.* 124, 170–178. doi: 10.1159/000207525

Kopecky, D., Loureiro, J., Zwierzykowski, Z., Ghesquiere, M., and Dolezel, J. (2006). Genome constitution and evolution in *Lolium x Festuca* hybrid cultivars (Festulolium). *Theor. Appl. Genet.* 113, 731–742. doi: 10.1007/s00122-006-0341-z

Kraitshtein, Z., Yaakov, B., Khasdan, V., and Kashkush, K. (2010). Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genetics* 186, 801–812. doi: 10.1534/genetics.110.120790

Kryvokhyzha, D., Milesi, P., Duan, T. L., Orsucci, M., Wright, S. I., Glemin, S., et al. (2019). Towards the new normal: transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (*Capsella bursa-pastoris*). *PLoS Genet.* 15:e1008131. doi: 10.1371/journal.pgen.1008131

Ksiazczyk, T., Kovarik, A., Eber, F., Huteau, V., Khaitova, L., Tesarikova, Z., et al. (2011). Immediate unidirectional epigenetic reprogramming of NORs occurs independently of rDNA rearrangements in synthetic and natural forms of a polyploid species *Brassica napus*. *Chromosoma* 120, 557–571. doi: 10.1007/s00412-011-0331-z

Kubota, A., Akiyama, Y., and Fujimori, M. (2019). The relationship between f ratio and seed yield-related traits in Festulolium. *Crop Sci.* 59, 1992–1996. doi: 10.2135/cropsci2019.02.0092

Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8, 104–115. doi: 10.1038/nrg2041

Lawrence, R. J., Earley, K., Pontes, O., Silva, M., Chen, Z. J., Neves, N., et al. (2004). A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol. Cell* 13, 599–609. doi: 10.1016/S1097-2765(04)00064-4

Li, A. L., Liu, D. C., Wu, J., Zhao, X. B., Hao, M., Geng, S. F., et al. (2014). mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell* 26, 1878–1900. doi: 10.1105/tpc.114.124388

Li, N., Xu, C., Zhang, A., Lv, R., Meng, X., Lin, X., et al. (2019). DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *New Phytol.* 223, 979–992. doi: 10.1111/nph.15820

Lim, K. Y., Matyasek, R., Kovarik, A., and Leitch, A. (2004). Genome evolution in allotetraploid *Nicotiana*. *Biol. J. Linn. Soc.* 82, 599–606. doi: 10.1111/j.1095-8312.2004.00344.x

Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., et al. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J.* 23, 4146–4155. doi: 10.1038/sj.emboj.7600430

Liu, Z. Q., Adamczyk, K., Manzanares-Dauleux, M., Eber, F., Lucas, M. O., Delourme, R., et al. (2006). Mapping PrBn and other quantitative trait loci responsible for the control of homeologous chromosome pairing in oilseed rape (*Brassica napus* L.) haploids. *Genetics* 174, 1583–1596. doi: 10.1534/genetics.106.064071

Lu, F. H., McKenzie, N., Gardiner, L. J., Luo, M., Hall, A., and Bevan, M. W. (2019). Reduced chromatin accessibility underlies gene expression differences in homologous chromosome arms of hexaploid wheat and diploid *Aegilops tauschii*. *bioRxiv* [Preprint]. doi: 10.1101/571133

Lu, J., Zhang, C. Q., Baulcombe, D. C., and Chen, Z. J. (2012). Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5529–5534. doi: 10.1073/pnas.1203094109

Lukaszewski, A. J., Apolinarska, B., Gustafson, J. P., and Krolow, K. D. (1987). Chromosome pairing and aneuploidy in tetraploid triticale. I. Stabilized karyotypes. *Genome* 29, 554–561. doi: 10.1139/g87-093

Lukaszewski, A. J., and Kopecky, D. (2010). The Ph1 locus from wheat controls meiotic chromosome pairing in autotetraploid rye (*Secale cereale* L.). *Cytogenet. Genome Res.* 129, 117–123. doi: 10.1159/000314279

Ma, X. F., and Gustafson, J. P. (2005). Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenet. Genome Res.* 109, 236–249. doi: 10.1159/000082406

Madlung, A. (2013). Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* 110, 99–104. doi: 10.1038/hdy.2012.79

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi: 10.1016/j.tree.2005.02.010

Martin, A. C., Rey, M. D., Shaw, P., and Moore, G. (2017). Dual effect of the wheat Ph1 locus on chromosome synapsis and crossover. *Chromosoma* 126, 669–680. doi: 10.1007/s00412-017-0630-0

Martin, A. C., Shaw, P., Phillips, D., Reader, S., and Moore, G. (2014). Licensing MLH1 sites for crossover during meiosis. *Nat. Commun.* 5:4580. doi: 10.1038/ncomms5580

Masterson, J. (1994). Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264, 421–424. doi: 10.1126/science.264.5157.421

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

Mhiri, C., Parisod, C., Daniel, J., Petit, M., Lim, K. Y., Dorlhac de Borne, F., et al. (2019). Parental transposable element loads influence their dynamics in young *Nicotiana* hybrids and allotetraploids. *New Phytol.* 221, 1619–1633. doi: 10.1111/nph.15484

Mohannath, G., Pontvianne, F., and Pikaard, C. S. (2016). Selective nucleolus organizer inactivation in *Arabidopsis* is a chromosome position-effect phenomenon. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13426–13431. doi: 10.1073/pnas.1608140113

Naranjo, T. (2014). Dynamics of rye telomeres in a wheat background during early meiosis. *Cytogenet. Genome Res.* 143, 60–68. doi: 10.1159/000363524

Navashin, M. (1934). Chromosomal alterations caused by hybridization and their bearing upon certain general genetic problems. *Cytologia* 5, 169–203. doi: 10.1508/cytologia.5.169

Neves, N., Silva, M., HeslopHarrison, J. S., and Viegas, W. (1997). Nucleolar dominance in triticales: control by unlinked genes. *Chromosome Res.* 5, 125–131. doi: 10.1023/A:1018470208730

Novikova, P. Y., Tsuchimatsu, T., Simon, S., Nizhynska, V., Voronin, V., Burns, R., et al. (2017). Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Mol. Biol. Evol.* 34, 957–968. doi: 10.1093/molbev/msw299

Orellana, J., Cermeno, M. C., and Lacadena, J. R. (1984). Meiotic pairing in wheat-rye addition and substitution lines. *Can. J. Genet. Cytol.* 26, 25–33. doi: 10.1139/g84-005

Pandit, M. K., Pocock, M. J. O., and Kunin, W. E. (2011). Ploidy influences rarity and invasiveness in plants. *J. Ecol.* 99, 1108–1115. doi: 10.1111/j.1365-2745.2011.01838.x

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., et al. (2010a). Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol.* 186, 37–45. doi: 10.1111/j.1469-8137.2009.03096.x

Parisod, C., Holderegger, R., and Brochmann, C. (2010b). Evolutionary consequences of autopolyploidy. *New Phytol.* 186, 5–17. doi: 10.1111/j.1469-8137.2009.03142.x

Parisod, C., Mhiri, C., Lim, K. Y., Clarkson, J. J., Chase, M. W., Leitch, A. R., et al. (2012). Differential dynamics of transposable elements during long-term diploidization of *Nicotiana* section Repandae (Solanaceae) allopolyploid genomes. *PLoS One* 7:e50352. doi: 10.1371/journal.pone.0050352

Pernickova, K., Kolackova, V., Lukaszewski, A. J., Fan, C. L., Vrana, J., Duchoslav, M., et al. (2019a). Instability of alien chromosome introgressions in wheat associated with improper positioning in the nucleus. *Int. J. Mol. Sci* 20:E1448. doi: 10.3390/ijms20061448

Pernickova, K., Linc, G., Gaal, E., Kopecky, D., Samajova, O., and Lukaszewski, A. J. (2019b). Out-of-position telomeres in meiotic leptotene appear responsible for chiasmate pairing in an inversion heterozygote in wheat (*Triticum aestivum* L.). *Chromosoma* 128, 31–39. doi: 10.1007/s00412-018-0686-5

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B. J., Rudi, H., Hvidsten, T. R., et al. (2014). Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345:1250091. doi: 10.1126/science.1250091

Ramirez-Gonzalez, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361:eaar6089. doi: 10.1126/science.aar6089

Rando, O. J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr. Opin. Genet. Dev.* 22, 148–155. doi: 10.1016/j.gde.2012.02.013

Rao, D. D., Vorhies, J. S., Senzer, N., and Nemunaitis, J. (2009). siRNA vs. shRNA: similarities and differences. *Adv. Drug Deliv. Rev.* 61, 746–759. doi: 10.1016/j.addr.2009.04.004

Rawale, K. S., Khan, M. A., and Gill, K. S. (2019). The novel function of the Ph1 gene to differentiate homologs from homoeologs evolved in *Triticum turgidum* ssp. dicoccoides via a dramatic meiosis-specific increase in the expression of the 5B copy of the C-Ph1 gene. *Chromosoma* 128, 561–570. doi: 10.1007/s00412-019-00724-6

Renny-Byfield, S., Chester, M., Kovaøík, A., Le Comber, S. C., Grandbastien, M. A., Deloger, M., et al. (2011). Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. *Mol. Biol. Evol.* 28, 2843–2854. doi: 10.1093/molbev/msr112

Rey, M. D., Martin, A. C., Higgins, J., Swarbreck, D., Uauy, C., Shaw, P., et al. (2017). Exploiting the ZIP4 homologue within the wheat Ph1 locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. *Mol. Breed.* 37:95. doi: 10.1007/s11032-017-0700-2

Rey, M. D., Martin, A. C., Smedley, M., Hayta, S., Harwood, W., Shaw, P., et al. (2018). Magnesium increases homoeologous crossover frequency during meiosis in ZIP4 (Ph1 Gene) mutant wheat-wild relative hybrids. *Front. Plant Sci.* 9:509. doi: 10.3389/fpls.2018.00509

Rieseberg, L. H., Kim, S., Randell, R. A., Whitney, K. D., Gross, B. L., Lexer, C., et al. (2007). Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129, 149–165. doi: 10.1007/s10709-006-9011-y

Rodrigues, J. A., and Zilberman, D. (2015). Evolution and function of genomic imprinting in plants. *Genes Dev.* 29, 2517–2531. doi: 10.1101/gad.269902.115

Ruprecht, C., Proost, S., Hernandez-Coronado, M., Ortiz-Ramirez, C., Lang, D., Rensing, S. A., et al. (2017). Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. *Plant J.* 90, 447–465. doi: 10.1111/tpj.13502

Schnable, J. C., Springer, N. M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4069–4074. doi: 10.1073/pnas.1101368108

Schubert, I., and Lysak, M. A. (2011). Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* 27, 207–216. doi: 10.1016/j.tig.2011.03.004

Sears, E. R., and Okamoto, M. (1958). "Intergenomic chromosome relationships in hexaploid wheat," in *Proceedings of the Xth International Congress of Genetics*, ed. J. W. Boyes (Toronto: University of Toronto Press), 258–259. doi: 10.1093/aob/mcm331

Sedel'nikova, T. S., Muratova, E. N., and Pimenov, A. V. (2011). Variability of chromosome numbers in gymnosperms. *Biol. Bull. Rev.* 1, 100–109. doi: 10.1134/S2079086411020083

Shi, X. L., Ng, D. W. K., Zhang, C. Q., Comai, L., Ye, W. X., and Chen, Z. J. (2012). *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat. Commun.* 3:950. doi: 10.1038/ncomms1954

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

Soltis, P. S., and Soltis, D. E. (2012). *Polyploidy and Genome Evolution*. Berlin: Springer, doi: 10.1007/978-3-642-31442-1

Song, Q. X., and Chen, Z. J. (2015). Epigenetic and developmental regulation in plant polyploids. *Curr. Opin. Plant Biol.* 24, 101–109. doi: 10.1016/j.pbi.2015.02.007

Stoces, S., Ruttink, T., Bartos, J., Studer, B., Yates, S., Zwierzykowski, Z., et al. (2016). Orthology guided transcriptome assembly of Italian ryegrass and meadow fescue for single-ncleotide polymorphism discovery. *Plant Genome* 9, 1–14. doi: 10.3835/plantgenome2016.02.0017

Takahashi, C., Leitch, I. J., Ryan, A., Bennett, M. D., and Brandham, P. E. (1997). The use of genomic in situ hybridization (GISH) to show transmission of recombinant chromosomes by a partially fertile bigeneric hybrid, *Gasteria lutzii x Aloe aristata* (Aloaceae), to its progeny. *Chromosoma* 105, 342–348. doi: 10.1007/s004120050193

Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L., and Henikoff, S. (2002). Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 14, 1053–1066. doi: 10.1105/tpc.010425

te Beest, M., Le Roux, J. J., Richardson, D. M., Brysting, A. K., Suda, J., Kubesova, M., et al. (2012). The more the better? The role of polyploidy in facilitating plant invasions. *Ann. Bot.* 109, 19–45. doi: 10.1093/aob/mcr277

Thomas, B. C., Pedersen, B., and Freeling, M. (2006). Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* 16, 934–946. doi: 10.1101/gr.4708406

Tsunewaki, K. (1964). Genetic studies of 6X-derivative from an 8x *Triticale*. *Can. J. Genet. Cytol.* 6, 1–11. doi: 10.1139/g64-001

van Heusden, A. W., van Ooijen, J. W., Vrielink-van Ginkel, R., Verbeek, W. H. J., Wietsma, W. A., and Kik, C. (2000). A genetic map of an interspecific cross in Allium based on amplified fragment length polymorphism (AFLP (TM)) markers. *Theor. Appl. Genet.* 100, 118–126. doi: 10.1007/s001220050017

Vieira, R., Queiroz, A., Morais, L., Barao, A., Mello-Sampayo, T., and Viegas, W. (1990). Genetic control of 1R nucleolus organizer region expression in the presence of wheat genomes. *Genome* 33, 713–718. doi: 10.1139/g90-107

Wang, M. J., Wang, P. C., Lin, M., Ye, Z. X., Li, G. L., Tu, L. L., et al. (2018). Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat. Plants* 4, 90–97. doi: 10.1038/s41477-017-0096-3

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165

Woodhouse, M. R., Cheng, F., Pires, J. C., Lisch, D., Freeling, M., and Wang, X. (2014). Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5283–5288. doi: 10.1073/pnas.1402475111

Wu, J., Lin, L., Xu, M. L., Chen, P. P., Liu, D. X., Sun, Q. F., et al. (2018). Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. *BMC Genomics* 19:586. doi: 10.1186/s12864-018-4966-5

Xiong, Z. Y., Gaeta, R. T., and Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7908–7913. doi: 10.1073/pnas.1014138108

Yaakov, B., and Kashkush, K. (2012). Mobilization of Stowaway-like MITEs in newly formed allohexaploid wheat species. *Plant Mol. Biol.* 80, 419–427. doi: 10.1007/s11103-012-9957-3

Yakimowski, S. B., and Rieseberg, L. H. (2014). The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *Am. J. Bot.* 101, 1247–1258. doi: 10.3732/ajb.1400201

Yoo, M. J., Szadkowski, E., and Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110, 171–180. doi: 10.1038/hdy.2012.94

Zhao, N., Zhu, B., Li, M. J., Wang, L., Xu, L. Y., Zhang, H. K., et al. (2011). Extensive and heritable epigenetic remodeling and genetic stability accompany allohexaploidization of wheat. *Genetics* 188, 499–510. doi: 10.1534/genetics.111.127688

Zhu, W. S., Hu, B., Becker, C., Dogan, E. S., Berendzen, K. W., Weigel, D., et al. (2017). Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific *Arabidopsis* hybrid. *Genome Biol.* 18:157. doi: 10.1186/s13059-017-1281-4

Zwierzykowski, Z., Kosmala, A., Zwierzykowska, E., Jones, N., Joks, W., and Bocianowski, J. (2006). Genome balance in six successive generations of the allotetraploid *Festuca pratensis x Lolium perenne*. *Theor. Appl. Genet.* 113, 539–547. doi: 10.1007/s00122-006-0322-2

Zwierzykowski, Z., Zwierzykowska, E., Taciak, M., Jones, N., Kosmala, A., and Krajewski, P. (2008). Chromosome pairing in allotetraploid hybrids of *Festuca pratensis x Lolium perenne* revealed by genomic in situ hybridization (GISH). *Chromosome Res.* 16, 575–585. doi: 10.1007/s10577-008-1198-6

Zwierzykowski, Z., Zwierzykowska, E., Taciak, M., Kosmala, A., Jones, R. N., Zwierzykowski, W., et al. (2011). Genomic structure and fertility in advanced breeding populations derived from an allotetraploid *Festuca pratensis x Lolium perenne* cross. *Plant Breed.* 130, 476–480. doi: 10.1111/j.1439-0523.2010.01839.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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 February 2020
doi: 10.3389/fpls.2020.00205

The Formation of Sex Chromosomes in *Silene latifolia* and *S. dioica* Was Accompanied by Multiple Chromosomal Rearrangements

Václav Bačovský[1], Radim Čegan[1,2], Denisa Šimoníková[2], Eva Hřibová[2] and Roman Hobza[1,2]**

[1] Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czechia, [2] Institute of Experimental Botany, Czech Academy of Sciences, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czechia

OPEN ACCESS

Edited by:
Martin A. Lysak, Masaryk University, Czechia

Reviewed by:
Ales Kovarik, Academy of Sciences of the Czech Republic (ASCR), Czechia
Andreas Houben, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany

***Correspondence:**
Václav Bačovský
xbacovs@ibp.cz
Roman Hobza
hobza@ibp.cz

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

Received: *07 October 2019*
Accepted: *11 February 2020*
Published: *28 February 2020*

Citation:
Bačovský V, Čegan R, Šimoníková D, Hřibová E and Hobza R (2020) The Formation of Sex Chromosomes in Silene latifolia and S. dioica Was Accompanied by Multiple Chromosomal Rearrangements. Front. Plant Sci. 11:205. doi: 10.3389/fpls.2020.00205

The genus *Silene* includes a plethora of dioecious and gynodioecious species. Two species, *Silene latifolia* (white campion) and *Silene dioica* (red campion), are dioecious plants, having heteromorphic sex chromosomes with an XX/XY sex determination system. The X and Y chromosomes differ mainly in size, DNA content and posttranslational histone modifications. Although it is generally assumed that the sex chromosomes evolved from a single pair of autosomes, it is difficult to distinguish the ancestral pair of chromosomes in related gynodioecious and hermaphroditic plants. We designed an oligo painting probe enriched for X-linked scaffolds from currently available genomic data and used this probe on metaphase chromosomes of *S. latifolia* (2n = 24, XY), *S. dioica* (2n = 24, XY), and two gynodioecious species, *S. vulgaris* (2n = 24) and S. *maritima* (2n = 24). The X chromosome-specific oligo probe produces a signal specifically on the X and Y chromosomes in *S. latifolia* and *S. dioica,* mainly in the subtelomeric regions. Surprisingly, in *S. vulgaris* and *S. maritima*, the probe hybridized to three pairs of autosomes labeling their p-arms. This distribution suggests that sex chromosome evolution was accompanied by extensive chromosomal rearrangements in studied dioecious plants.

Keywords: chromosome painting, double-translocation, pseudo-autosomal region, *Silene*, Y chromosome

INTRODUCTION

The genus *Silene* is a model system for sex chromosome evolution, including about 700 species varying greatly in their mating system, ecology and sex determination (Bernasconi et al., 2009). Inside the genus two groups are considered as invaluable for the study of sex chromosome evolution and sex determination; section *Melandrium* and subsection *Otites* (reviewed in Vyskot and Hobza, 2015). Two dioecious plants *S. latifolia* (24, XY) and *S. dioica* (24, XY) from *Melandrium* have heteromorphic sex chromosomes and sex determination similar to mammals (Ming et al., 2007; Charlesworth, 2016). In contrast, related gynodioecious species *S. vulgaris* and *S. maritima* with the same number of autosomes (2n = 24), possess no sex chromosomes having a smaller

genome compared to *S. latifolia* or *S. dioica* (Runyeon and Prentice, 1997; Charlesworth and Laporte, 1998; Široký et al., 2001; Stone et al., 2017).

It is generally accepted that the sex chromosomes are derived from an ordinary pair of autosomes (reviewed in Bachtrog, 2006). As a result of suppressed recombination and accumulation of deleterious mutations, the sex chromosomes differ in their structure, function and gene density. The X chromosome becomes hemizygous and X hemizygosity in males leads to special regulatory mechanisms to equalize the transcription ratio between the X chromosomes and autosomes (Charlesworth and Charlesworth, 2005; Muyle et al., 2017; Darolti et al., 2019). As a result of accumulation of deleterious mutations, the Y chromosome is degenerated and the sex chromosomes may differ even within closely related species as demonstrated in human and chimpanzee (Hughes et al., 2010). Interestingly, newly formed sex chromosomes show the same signs of sex chromosome evolutionary pathways, as described in *Drosophila* (Bachtrog et al., 2009) or stickleback species (Yoshida et al., 2014) in which the ancestral Y chromosome fused with an autosome.

In *S. latifolia* and *S. dioica*, sequence data showed that the sex chromosomes evolved from one pair of autosomes with the divergence of X and Y homologous sequences <10 million years ago (Filatov, 2005), estimating the age of older and younger strata (non-recombining part of the sex chromosomes that differ from each other in level of divergence) around 11 and 6.32 mya (Krasovec et al., 2018). The sex chromosomes in *S. latifolia* and *S. dioica,* vary greatly in size having Y chromosome 1.4 larger than X (heteromorphism) (Vyskot and Hobza, 2015). The Y chromosome has a large non-recombining region and the size of the PAR (pseudo-autosomal region) is less than 10% of its chromosome length (Filatov et al., 2009). Both sex chromosomes accumulated various transposable elements (Bergero et al., 2008b; Kubat et al., 2014) and satellites (Cermak et al., 2008; Kejnovský et al., 2013), and differ in histone modifications and DNA methylation (Rodríguez Lorenzo et al., 2018; Bačovský et al., 2019).

Previous studies suggested that the sex chromosomes in *S. latifolia*, especially the Y chromosome, were derived through multiple rearrangements (Bergero et al., 2008a). Deletion mapping revealed that at least one larger inversion occurred after recombination suppression on the Y chromosome (Zluvova et al., 2005), supported also by findings of four genetically mapped genes between *S. latifolia* and *S. dioica* (Filatov, 2005). Later, Hobza et al. (2007) used physical mapping and confirmed two large inversions on the Y chromosome. These findings were further verified by Y deletion mapping showing that at least one inversion had to have occurred during the formation of the Y chromosome (Kazama et al., 2016), accelerating the recombination suppression (Bergero and Charlesworth, 2009). In contrast, comparative mapping between *S. latifolia* and *S. vulgaris* revealed the existence of one large (SvLG12) and two relatively small (SvLG9, Sv small LG) linkage groups that accompanied the sex chromosomes formation in *S. latifolia* (Bergero et al., 2013; Campos et al., 2017). Yet, it is still not clear what pair(s) of autosomes were included in such translocation and if such linkage groups also exist in other gynodioecious species. Thus, this raises two important questions: if *S. vulgaris* possesses three putative parts of three linkage groups corresponding to the X chromosome in *S. latifolia*, what is the origin of these linkage groups and how many chromosomes were involved in sex chromosome formation?

Recent advances in fluorescence *in situ* hybridization (FISH) experiments have provided a variety of techniques which can be used to study the structure, dynamics and origin of certain loci, chromosomal arms and/or specific chromosomes (reviewed in Cui et al., 2016; Bačovský et al., 2018; Huber et al., 2018). Previous cytogenetic studies in *Silene* species were based mainly on physical mapping of satellite rDNA (Široký et al., 2001), repeats and transposable elements (Cermak et al., 2008; Kejnovsky et al., 2009). Although Lengerova et al. (2004) managed to produce discrete signals using various DNA repeats (satellites, rDNA) and specific BAC clones, this approach proved to be cost ineffective due to the large screening of BACs containing only a low amount of repetitive DNA. As an another option, Hobza et al. (2004) designed a protocol using microdissected X and Y chromosomes from *S. latifolia* for whole chromosome painting. These probes produced relatively discrete signals on both sex chromosomes, but high amount of suppressive unlabeled DNA with very strict hybridization conditions made the use of such method very problematic in other *Silene* species (Hobza and Vyskot, 2007). The recent development of oligo painting probes in plants has proved to be useful in the detection of chromosomal aberrations and in comparative cytogenetics (reviewed in Jiang, 2019). In principal, oligo painting probe can be used to label particular regions containing enough short unique oligo sequences to be computationally isolated, synthesized, pooled and labeled (Han et al., 2015). Such probes, designed from conserved sequences of one species were used e.g. for developing karyotype among genetically related *Solanum* species (Braz et al., 2018), for differentiating of A, B, and D genomes in wheat (Tang et al., 2018), in comparative physical mapping of sex chromosomes in *Populus* (Xin et al., 2018) and in examination of meiotic pairing in polyploid *Solanum* species (He et al., 2018).

In this work, we designed an X chromosome-specific oligo probe enriched by X-linked scaffolds based on the *S. latifolia* female genome. We show that such technique is useful for the detection of discrete signals in sex chromosomes in *S. latifolia* and closely related *S. dioica*. In addition, the probe works well in the related gynodioecious species of *S. vulgaris* and *S. maritima*. Based on our results, we discuss the origin of sex chromosomes from one autosomal pair and the possible application of oligo painting probe in further studies. Our findings support the general hypothesis that multiple chromosomal changes took place during the formation of X and Y chromosomes.

MATERIALS AND METHODS

Plant Material

Seedlings of the *Silene* species listed in **Supplementary Table S1** (seeds owned by The Institute of Biophysics of the Czech Academy of Sciences) were used for chromosome preparation following (Bačovský et al., 2019). Young seedlings (average

size = 1 cm) were synchronized for 16 h in 1.125 mM hydroxyurea at RT, washed 2× for 5 min in distilled water and incubated 4 h in distilled water at RT. Cells in metaphase were accumulated by 0.05 mM colchicine at RT 4 h. After 4 h, root tips were stored for 16 h in ice cold water according to Pan et al. (1993). This reduced the number of ball metaphases and increased the mitotic index. As a final step, synchronized seedlings were fixed in freshly prepared Clarke's fixative (ethanol:glacial acetic acid, 3:1, v:v) for 24 h and stored at −20°C in 96% ethanol until use.

Oligo Painting Probe Selection and Preparation

The oligo painting probe of *S. latifolia*, prepared for X chromosome, was designed using Chorus software as previously described by Han et al. (2015). Briefly, oligo sequences (45 nt; >75% similarity) specific to X chromosome, based on the *S. latifolia* female genome (PRJNA289891; Papadopulos et al., 2015), were selected throughout the X chromosome scaffolds anchored using an X genetic map. Repetitive sequences were discriminated and removed during oligo painting probe design by Chorus pipeline (Han et al., 2015). A total of 12 988 oligo sequences were selected to cover X-linked scaffolds. The oligo sequences were synthesized *de novo* as myTags 20K Immortal library by Arbor Biosciences (Ann Arbor, MI, United States; TATAA Biocenter, Göteborg, Sweden). Labeling and detection of the oligo painting probe followed the published protocol of Han et al. (2015). For labeling of oligo-RNA products, we used universal primers (Eurofins Genomics, Ebersberg, Germany) conjugated with the Cy3 (5′–Cy3–CGTGGTCGCGTCTCA–3′) or primers conjugated with the digoxigenin (5′–DIG CGTGGTCGCGTCTCA–3′), similarly as (Šimoníková et al., 2019). Digoxigenin was detected by FITC conjugated anti-DIG antibody (Roche Life Sciences).

The number of oligo sequences per scaffold, scaffold length, position on genetic map and scaffold ID are included in **Supplementary Table S2**.

Chromosome and Probe Preparation

Chromosome spreads were obtained from multiple individuals from one population of studied species listed in **Supplementary Table S1**. Chromosomes were prepared as described in Bačovský et al. (2019) with minor modifications. Briefly, fixed root tips were washed 2× in distilled water 5 min, 2× in 0.001M citrate buffer 5 min and digested for 45–50 min in 1% enzyme mix (**Supplementary Table S3**) diluted in 0.001M citrate buffer. Chromosomes were squashed on to slides, freezed in liquid nitrogen and incubated for 5 min in freshly prepared Clarke's fixative. Prepared slides were used directly for fluorescence *in situ* hybridisation (FISH) or stored at −20°C in 96% ethanol until use.

FISH was performed as described by Schubert et al. (2016) using four different stringencies (**Supplementary Table S3**). The centromeric *Silene* tandem arrayed repeat (STAR-C) and subtelomeric tandem repeat (X43.1) were used as reference probes described in Bačovský et al. (2019). STAR-C is primarily located in centromeres on the X chromosome and autosomes, and on the Y in an additional two clusters based on stringency conditions (Hobza et al., 2007). Chromosome pictures were captured with Olympus AX70 microscope equipped with the cold cube camera. After image capture, all channels were processed with the software Adobe Photoshop free version CS2. A color histogram for each X and Y chromosome image was drawn using RGB profiler in ImageJ 1.52i FiJi[1]. These histograms display the distribution of DNA probes along each chromosome arm. RGB profiler was used on the same plot, for each type of green, red or blue selection as described in Mathur et al. (2012). The oligo painting probe was used in at least three individual experiments and each labeling pattern was scored in 30 metaphases/interphases per experiment. We did not observe any variation in signal patterns in studied populations.

RESULTS

Test for X Chromosome-Specific Oligo Probe Stringency and Oligo Painting Probe Signal Strength

We developed the X chromosome-specific oligo probe from female genomic sequences described in Papadopulos et al. (2015), using the approach described in Han et al. (2015). A total of 12 988 oligo sequences was selected from the entire currently available genomic data (**Supplementary Figure S1**; Papadopulos et al., 2015), covering on average 2.5–3 oligo sequence/kb (1.8–5.5 oligo sequence/kb) for the selected loci. The total density of selected oligo sequences from the X chromosome is below the recommended level of oligo sequence number per kilobase (0.03 oligo sequence/kb) (Han et al., 2015; Jiang, 2019). Nevertheless, for selected regions an average density of 1.8–5.5 oligo sequence/kb and the average number of oligo sequences is higher than the recommended standard of an oligo painting probe for metaphase chromosomes and single loci, 0.1–0.5 oligo sequence/kb (Han et al., 2015; Jiang, 2019).

To study potential rearrangements accompanying sex chromosome evolution, we used an X chromosome-specific probe in four species in the genus *Silene*, two dioecious (*S. latifolia* and *S. dioica*) and two gynodioecious (*S. vulgaris* and *S. maritima*). In *S. latifolia*, *S. dioica* and *S. vulgaris,* the oligo painting probe yields identical pattern in each species using direct (Cy3-tagged oligo sequences) and indirect labeling (digoxigenin tagged oligo sequences), and different hybridization stringencies (**Supplementary Table S3**). Only minor changes were observed in signal strength if the amount of oligo painting probe in *S. vulgaris* was increased (set to 1 μg per slide due to the weak overall chromosomal coverage). Nevertheless, 87 and 77% stringency produced very faint signal on the chromosomes of *S. maritima* (data not showed), using direct or indirect labeling and using the same amount of DNA (1 μg of oligo painting probe per slide). Therefore, we tested additional two hybridization stringencies, 68 and 62%, respectively, and we detected a similar pattern in *S. maritima* as in *S. vulgaris* (**Figure 1** and **Supplementary Figure S3**) (signal on three pairs of autosomes). Therefore, 68% hybridization stringency was

[1]https://imagej.nih.gov/ij/plugins/

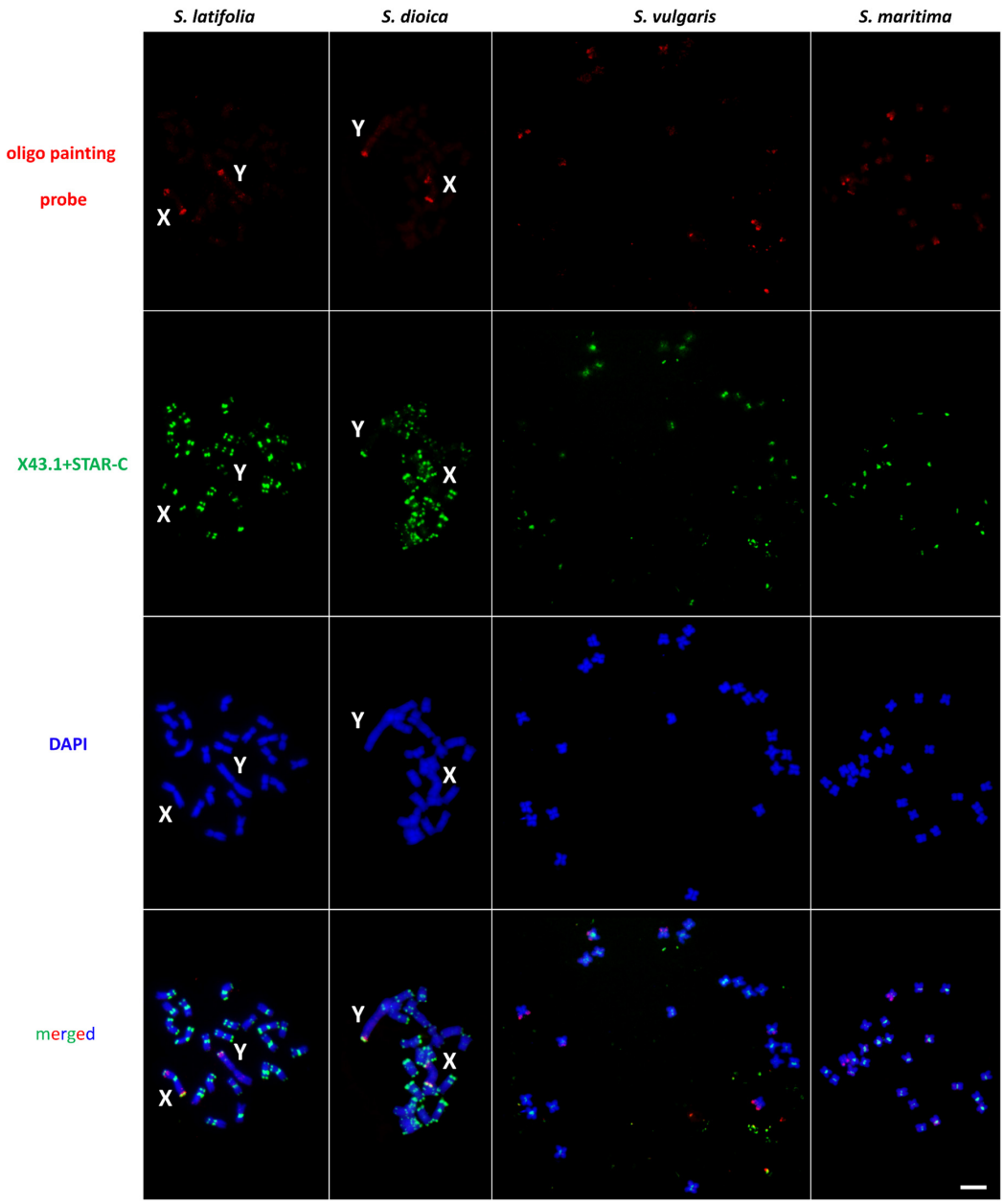

FIGURE 1 | Distribution of X chromosome-specific oligo probe on chromosomes in *Silene latifolia*, *S. dioica*, *S. vulgaris* and *S. maritima*. In *S. latifolia* and *S. dioica*, oligo painting probe has discrete signal only on X and Y chromosomes. In *S. vulgaris* and *S. maritima*, three pairs of autosomes can be differentiated from the whole genome. Scale bar = 5 μm.

applied in additional experiments for final analysis in all studied species using 1 μg of X chromosome-specific oligo painting probe per slide (**Figure 1** and **Supplementary Figure S3**).

X Chromosome-Specific Oligo Probe Pattern in *Silene* Species

The designed oligo painting probe hybridized to both ends of the X chromosome arms (including PAR on the p-arm) and to PAR located on Y q-arm in *S. latifolia* and *S. dioica*. In addition, an extra oligo painting probe signal is clearly visible on the X p-arm, suggesting potential gene-rich locus in this (sub)telomeric region (**Figure 2**). On the Y, the probe colocalizes with X43.1 (sub)telomeric probe band on the Y q-arm (PAR region). The additional oligo painting probe signal was visible on the Y, as an interstitial region, located on the p-arm in both species (**Figure 2**). We did observe extra (weak) signal on the autosomes, using both Cy3- and digoxigenin-conjugated primers and various hybridization stringencies (77, 68, and 62%). Nevertheless, the extra (weak) signal was affected by low hybridization stringency.

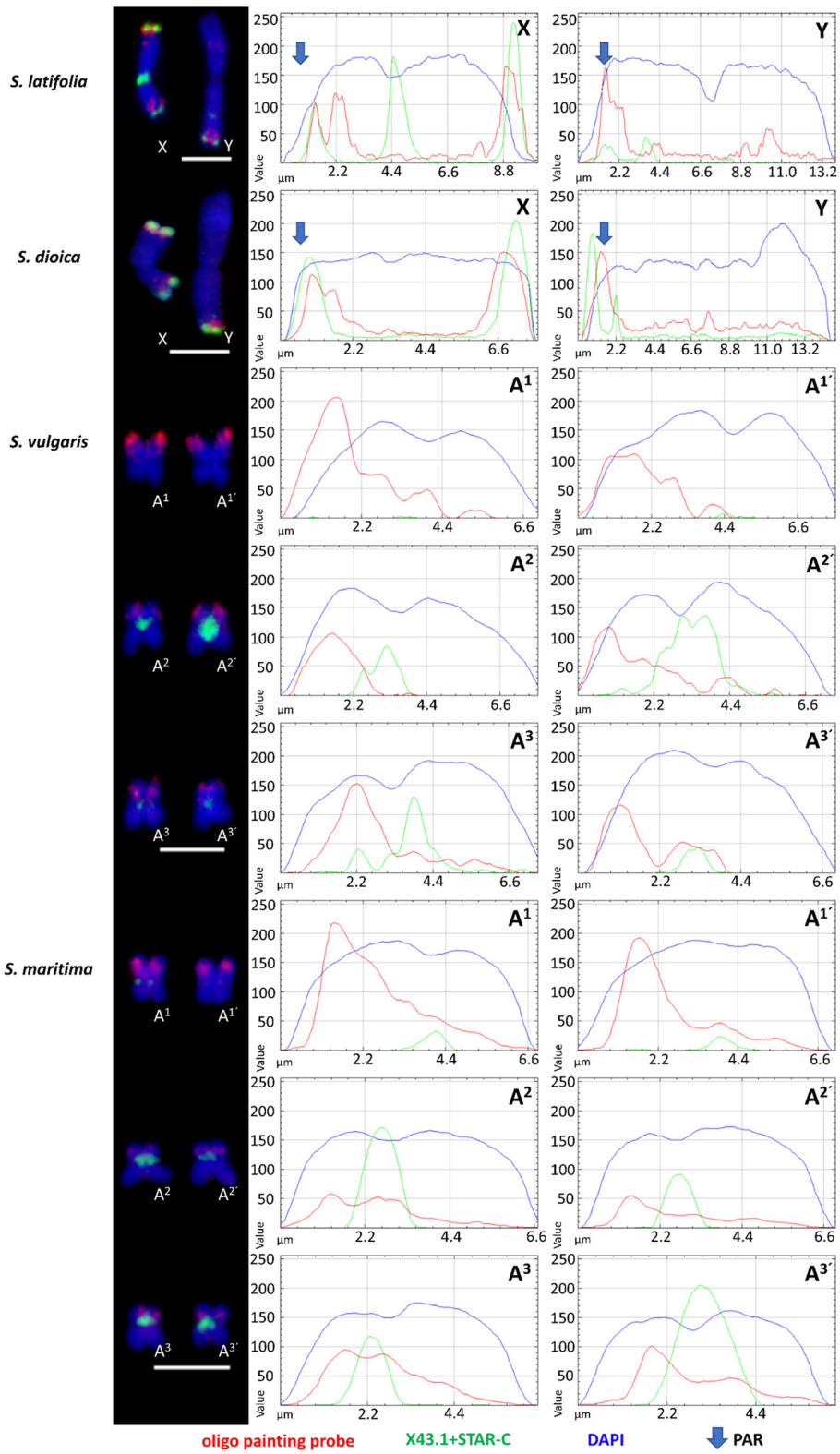

FIGURE 2 | Schematic distribution of X chromosome-specific oligo probe on sex chromosomes in *S. latifolia* and *S. dioica* and individual chromosomes in related gynodioecious species, *S. vulgaris* and *S. maritima*. Note the differences between **X** and **Y** chromosomes in *S. latifolia* and *S. dioica*. In S. *vulgaris* and *S. maritima,* the oligo painting probe signal is located on three pair of autosomes, numbered in this study as **A^1–A^3** and **A$^{1'}$–A$^{3'}$**. X43.1, a subtelomeric probe, is presented only on sex chromosomes and autosomes in *S. dioica* and *S. latifolia*. Scale bar = 5 µm.

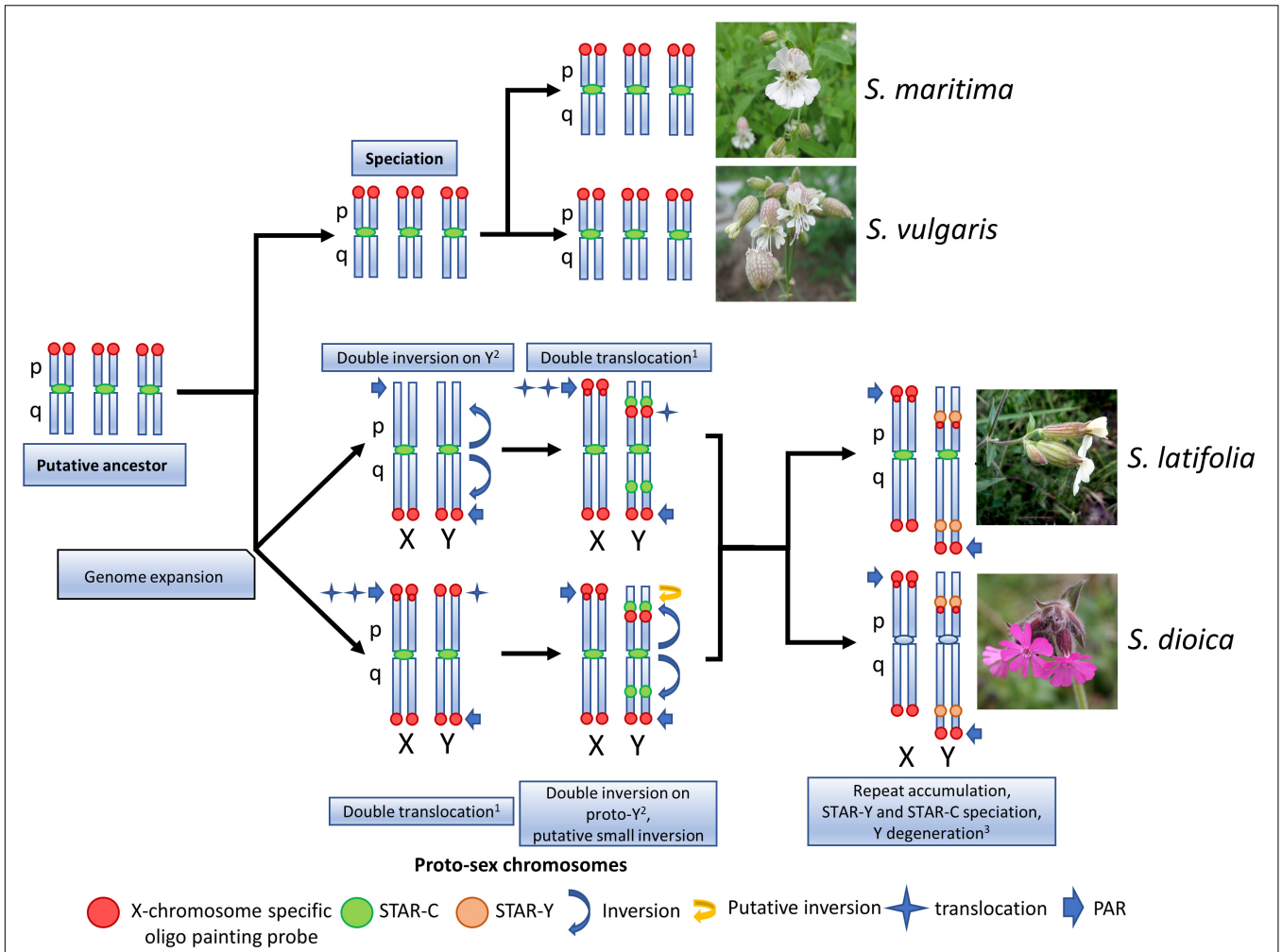

FIGURE 3 | Schematic view of the sex chromosomes evolution in *S. latifolia* and *S. dioica*. The sex chromosomes in *S. latifolia* and *S. dioica* underwent at least one double translocation[1] (Bergero et al., 2013; Campos et al., 2017) and double inversion[2] events (Hobza et al., 2007; Zluvova et al., 2007; Bergero et al., 2008a; Kazama et al., 2016). Compared to gynodioecious species (*S. vulgaris* and *S. maritima*), *S. latifolia* and *S. dioica* possess two sex chromosomes having signal from hybridized oligo painting probe. Based on the oligo painting probe signal in this study, two scenarios (supported by the literature) could led to the formation of sex chromosomes in *S. latifolia* and *S. dioica*. If the double translocation moving the segments of STAR-C satellite occurred first, then the autosomal parts have been translocated to proto-sex chromosome X PAR and interstitial Y p-arm region (upper scenario). In the second scenario, the autosomal parts were translocated to the sex chromosomes first, followed by double translocation and at least one putative translocation on the Y p-arm. Flower pictures downloaded on: https://commons.wikimedia.org/wiki/Main_Page.

In *S. latifolia* and *S. dioica* female karyotype, the oligo painting probe produced the same signal on both X chromosomes as on the X chromosome in males (**Supplementary Figure S4**). Thus, the X chromosome-specific oligo painting probe used in this work provides a highly reproducible signal in all studied species.

In *S. vulgaris* and *S. maritima*, application of oligo painting probe differentiated three pairs of autosomes, marked in this study as A^1–$A^{3'}$ (**Figure 2**). Compared to *S. maritima*, the decrease in hybridization stringency in *S. vulgaris* did not change the number of loci and signals on the chromosomes. In both gynodioecious species, the oligo painting probe labeled almost the entirety of the p-arms of A^1–$A^{1'}$, including (sub)centromere regions. Additionally, the oligo painting probe had a twofold stronger signal on the first pair of autosomes (A^1–$A^{1'}$), in *S. vulgaris* and *S. maritima*, than on the second and third (A^2–$A^{3'}$) autosomal pairs. The oligo painting probe hybridized to subtelomeric (A^2–$A^{2'}$) or more interstitial regions (A^3–$A^{3'}$) on these chromosomes (**Figures 1**, **2**).

In interphase, the X chromosome-specific probe differentiated the sex chromosome domains in *S. latifolia* and *S. dioica*, and the A^1–$A^{3'}$ autosomal regions in *S. vulgaris* and *S. maritima*. In the first two species, the oligo painting probe differentiated two subdomains located within one nucleus (**Supplementary Figures S2a,b**). In *S. vulgaris* and *S. maritima*, the oligo painting probe labeled three to six subdomains (**Supplementary Figures S2c,d**). Despite the average density being 1.8–5.5/kb in selected regions, the total coverage of the whole chromosome is only 0.03 oligo sequence/kb. The lower coverage is apparent (weaker signal)

in more relaxed chromatin state in early metaphase/prophase in all studied species (**Supplementary Figure S3**), showing that the oligo painting probe labeled (sub)telomere of the sex chromosomes and almost half of their length in the autosomes of *S. vulgaris* and *S. maritima*.

DISCUSSION

The oligo painting probe specifically hybridized to pseudo-autosomal region (PAR) located on the Y q-arm and interstitial Y p-arm region, and to both sub-telomeric regions on the X chromosome in *S. latifolia* and *S. dioica*, enriched on the X p-arm sub-telomere. This distribution of the oligo painting probe signal correlates with the pattern of specific histone modifications for active chromatin, namely H3K4me1, H3K4me2, H3K4me3, H3K9ac and gene repressive mark H3K27me3 (Bačovský et al., 2019). Since the Y chromosome still contains many active genes (Bergero and Charlesworth, 2011), the additional band on the X p-arm together with the interstitial Y p-arm region probably represent unique/important gene clusters (Hobza et al., 2018). In related *S. vulgaris* and *S. maritima*, the oligo painting probe clearly differentiates three pairs of autosomes. In these gynodioecious species, the oligo painting probe labels three autosomal p-arms as shown on early metaphase/prophase chromosomes (**Supplementary Figure S3**). These chromosomal patterns support previous evidence that parts of three linkage groups in *S. vulgaris* were translocated to the pseudo-autosomal region in *S. latifolia* through a double translocation event (Bergero et al., 2013). Our results support previous studies showing expansion of the PAR region in *S. latifolia* (Bergero et al., 2013; Campos et al., 2017) and newly in *S. dioica* (**Figure 3**). Alternatively, rearrangements could be accompanied by whole chromosome fusion(s) as documented in other species. In Japan Sea stickleback fish, an ancestral Y chromosome fused with the autosome, forming a neo-Y chromosome and the X_1X_2Y sex determination system (compared to closely related Pacific Ocean stickleback with XY system) (Yoshida et al., 2014). In *Rumex hastatulus*, North Carolina male race possesses XY_1Y_2, having the older Y (ancestral) chromosome heterochromatinised, and the younger Y with translocated autosomal part (Grabowska-Joachimiak et al., 2015). In *Silene* species "fusion" scenario is unlikely since it usually influences basic karyotype chromosome number (all studied species have $n = 12$). Moreover, existing genetic maps do not suggest large scale fusion events during karyotype evolution in studied species (Bergero et al., 2013). The existence of telomere-like sequences in *S. latifolia* Y chromosome (Uchida et al., 2002) can be explained as remnant of chromosomal inversion as was demonstrated in some *Silene* species (Filatov, 2005; Zluvova et al., 2005). It was reported that such chromosomal rearrangements included at least one paracentric and one pericentric inversion on the Y (Hobza et al., 2007; Bergero and Charlesworth, 2009). On the other hand, additional rearrangements that would suggest alternative scenarios cannot be excluded. Based on deletion mapping one inversion also occurred during the formation of the Y (Kazama et al., 2016). Bergero et al. (2013) and Campos et al. (2017) showed that PAR expanded through two step translocations. Our data support such translocation by the existence of three pair of autosomes (in gynodioecious species) corresponding to sex chromosomes in *S. latifolia* and *S. dioica*. Since sex chromosomes in studied dioecious species originated from one of these three chromosomal pairs, we assume that two additional loci were translocated on proto sex chromosomes in *S. latifolia* and *S. dioica* during sex chromosome evolution (**Figure 3**).

We have also tested the robustness of oligo painting method to study the dynamics of sex chromosomes and PAR in early metaphase/prophase (**Supplementary Figure S3**). In prophase/early metaphase in which the chromosomal (spatial) resolution limit is 10 times higher than in the metaphase and chromosomes are 10 times more extended (reviewed in Figueroa and Bass, 2010), the strength of the signal is relatively low. Therefore, it will be necessary to use more cytogenetic markers (together with oligo sequences) for discrimination of relaxed chromosomes such as e.g. specific antibodies against synaptonemal complexes in meiosis as described in Hurel et al. (2018).

Our X chromosome-specific oligo probe serves as useful tool to study the evolution of sex chromosomes in *S. latifolia*, *S. dioica* and their relatives. Since the genome of *S. latifolia* is still not fully sequenced we would like to leave the possibility that some sequences targeted by the probe might occur more than once in the genome open (e.g. low repetitive or duplicated). Nevertheless, compared to previous attempts and labeling of the sex chromosomes in *Silene* species using e.g. unique BAC clones (Lengerova et al., 2004) or dissected chromosomal probes (Hobza et al., 2004), the oligo painting probe provides an unique signal on both sex chromosomes and is also suitable to study other related species. In addition, the probe simplifies future analysis of chromosome pairing and facilitates the study of the dynamics of the PAR region in interphase or during cell division.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

VB, RČ, and RH conceived and designed the research. VB, RČ, EH, and DŠ conducted the experiments. DŠ and EH contributed the reagents. VB analyzed the data. VB wrote the manuscript. All the authors read and approved the manuscript.

FUNDING

This work was supported by the Czech Science Foundation grants No. 19-02476Y and 18-06147S.

ACKNOWLEDGMENTS

We would like to thank Chris Johnson for the English-language correction. We also like to thank the

reviewers for their thoughtful comments and efforts toward improving our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00205/full#supplementary-material

FIGURE S1 | The distribution of X chromosome-specific oligo probe in X chromosome genetic map in *S. latifolia*. The average coverage was chosen on 2–3 oligo sequences/kb (1.8–5.5). Total size of X chromosome is estimated around 400 Mb.

FIGURE S2 | Distribution of X chromosome-specific oligo in interphase in *S. latifolia*, *S. dioica*, *S. vulgaris* and *S. maritima*. Oligo painting probe differentiates two sub-domains in *S. latifolia* **(a)** and *S. dioica* **(b)**, and three to six sub-domains in *S. vulgaris* **(c)** and *S. maritima* **(d)**. Each sub-domain is enlarged and marked by separated colors (white/yellow/orange) in merged channel. X43.1, a sub-telomeric probe, is presented only on sex chromosomes and autosomes in *S. latifolia* and *S. dioica*. Note the distribution of STAR-C in *S. vulgaris* and *S. maritima*. Scale bar = 10 μm.

FIGURE S3 | Distribution of X chromosome-specific oligo probe on prophase and early metaphase chromosomes in *S. latifolia*, *S. dioica*, *S. vulgaris* and *S. maritima*. 12 988 has coverage 2.5–3 oligo sequences/kb, reaching 1.8–5.5 oligo sequences/kb on selected loci. Oligo painting probe (in red) hybridizes to very end of X and Y chromosomes in (sub)telomeres in *S. latifolia* and *S. dioica*. In *S. vulgaris* and *S. maritima*, the oligo painting probe labels six pairs of autosomes, hybridizing to their p-arm. Although the signal strength is weaker compared to condensed metaphase chromosomes, the oligo painting probe clearly marks sex chromosomes and autosomes in all studied species. Scale bar = 10 μm.

FIGURE S4 | Distribution of the oligo painting probe in *S. latifolia* and *S. dioica* female karyotype. Oligo painting probe was hybridized on metaphase chromosomes and on interphase nuclei in *S. latifolia* **(a,c)** and in *S. dioica* **(b,d)**. The remnant of a cytoplasm (signal not attached to any chromosome) is visible in the bottom of the *S. dioica* **(b)**. Scale bar = 10 μm.

TABLE S1 | Plant material.

TABLE S2 | Oligo probe sequence scaffolds ID.

TABLE S3 | The composition of the enzyme mix and hybridisation stringency.

REFERENCES

Bachtrog, D. (2006). A dynamic view of sex chromosome evolution. *Curr. Opin. Genet. Dev.* 16, 578–585. doi: 10.1016/j.gde.2006.10.007

Bachtrog, D., Jensen, J. D., and Zhang, Z. (2009). Accelerated adaptive evolution on a newly formed X chromosome. *PLoS Biol.* 7:e1000082. doi: 10.1371/journal.pbio.1000082

Bačovský, V., Hobza, R., and Vyskot, B. (2018). Technical review: cytogenetic tools for studying mitotic. *Methods Mol Biol.* 1675, 509–535. doi: 10.1007/978-1-4939-7318-7-30

Bačovský, V., Houben, A., Kumke, K., and Hobza, R. (2019). The distribution of epigenetic histone marks differs between the X and Y chromosomes in *Silene latifolia*. *Planta* 250, 487–494. doi: 10.1007/s00425-019-03182-7

Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* 24, 94–102. doi: 10.1016/j.tree.2008.09.010

Bergero, R., and Charlesworth, D. (2011). Preservation of the Y Transcriptome in a 10-Million-year-old plant sex chromosome system. *Curr. Biol.* 21, 1470–1474. doi: 10.1016/j.cub.2011.07.032

Bergero, R., Charlesworth, D., Filatov, D. A., and Moore, R. C. (2008a). Defining regions and rearrangements of the *Silene latifolia* Y chromosome. *Genetics* 178, 2045–2053. doi: 10.1534/genetics.107.084566

Bergero, R., Forrest, A., and Charlesworth, D. (2008b). Active miniature transposons from a plant genome and its nonrecombining Y chromosome. *Genetics* 178, 1085–1092. doi: 10.1534/genetics.107.081745

Bergero, R., Qiu, S., Forrest, A., Borthwick, H., and Charlesworth, D. (2013). Expansion of the pseudo-autosomal region and ongoing recombination suppression in the *Silene latifolia* sex chromosomes. *Genetics* 194, 673–686. doi: 10.1534/genetics.113.150755

Bernasconi, G., Antonovics, J., Biere, A., Charlesworth, D., Delph, L. F., Filatov, D., et al. (2009). Silene as a model system in ecology and evolution. *Heredity* 103, 5–14. doi: 10.1038/hdy.2009.34

Braz, G. T., He, L., Zhao, H., Zhang, T., Semrau, K., Rouillard, J.-M., et al. (2018). Comparative oligo-FISH mapping: an efficient and powerful methodology to reveal karyotypic and chromosomal evolution. *Genetics* 208, 513–523. doi: 10.1534/genetics.117.300344

Campos, J. L., Qiu, S., Guirao-Rico, S., Bergero, R., and Charlesworth, D. (2017). Recombination changes at the boundaries of fully and partially sex-linked regions between closely related Silene species pairs. *Heredity* 118, 395–403. doi: 10.1038/hdy.2016.113

Cermak, T., Kubat, Z., Hobza, R., Koblizkova, A., Widmer, A., Macas, J., et al. (2008). Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes. *Chromosom. Res.* 16, 961–976. doi: 10.1007/s10577-008-1254-2

Charlesworth, D. (2016). Plant sex chromosomes. *Annu. Rev. Plant Biol.* 67, 397–420. doi: 10.1146/annurev-arplant-043015-111911

Charlesworth, D., and Charlesworth, B. (2005). Sex chromosomes: evolution of the weird and wonderful. *Curr. Biol.* 15, R129–R131. doi: 10.1016/j.cub.2005.02.011

Charlesworth, D., and Laporte, V. (1998). The male-sterility polymorphism of Silene vulgaris: analysis of genetic dat from two populations and comparison with Thymus vulgaris. *Genetics* 150, 1267–1282.

Cui, C., Shu, W., and Li, P. (2016). Fluorescence in situ hybridization: cell-based genetic diagnostic and research applications. *Front. Cell Dev. Biol.* 4:89. doi: 10.3389/fcell.2016.00089

Darolti, I., Wright, A. E., Sandkam, B. A., Morris, J., Bloch, N. I., Farré, M., et al. (2019). Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proc. Natl. Acad. Sci. U.S.A.* 116, 19031–19036. doi: 10.1073/pnas.1905298116

Figueroa, D. M., and Bass, H. W. (2010). A historical andmodern perspective on plant cytogenetics. *Briefings Funct. Genom. Proteom.* 9, 95–102. doi: 10.1093/bfgp/elp058

Filatov, D. A. (2005). Evolutionary history of *Silene latifolia* sex chromosomes revealed by genetic mapping of four genes. *Genetics* 170, 975–979. doi: 10.1534/genetics.104.037069

Filatov, D. A., Howell, E. C., Groutides, C., and Armstrong, S. J. (2009). Recent spread of a retrotransposon in the *Silene latifolia* genome, apart from the Y chromosome. *Genetics* 181, 811–817. doi: 10.1534/genetics.108.099267

Grabowska-Joachimiak, A., Kula, A., Książczyk, T., Chojnicka, J., Sliwinska, E., and Joachimiak, A. J. (2015). Chromosome landmarks and autosome-sex chromosome translocations in *Rumex hastatulus*, a plant with XX/XY1Y2 sex chromosome system. *Chromosom. Res.* 23, 187–197. doi: 10.1007/s10577-014-9446-4

Han, Y., Zhang, T., Thammapichai, P., Weng, Y., and Jiang, J. (2015). Chromosome-specific painting in *Cucumis* species using bulked oligonucleotides. *Genetics* 200, 771–779. doi: 10.1534/genetics.115.177642

He, L., Braz, G. T., Torres, G. A., and Jiang, J. (2018). Chromosome painting in meiosis reveals pairing of specific chromosomes in polyploid *Solanum* species. *Chromosoma* 127, 505–513. doi: 10.1007/s00412-018-0682-9

Hobza, R., Hudzieczek, V., Kubat, Z., Cegan, R., Vyskot, B., Kejnovsky, E., et al. (2018). Sex and the flower – developmental aspects of sex chromosome evolution. *Ann. Bot.* 122, 1085–1101. doi: 10.1093/aob/mcy130

Hobza, R., Kejnovsky, E., Vyskot, B., and Widmer, A. (2007). The role of chromosomal rearrangements in the evolution of *Silene latifolia* sex

chromosomes. *Mol. Genet. Genomics* 278, 633–638. doi: 10.1007/s00438-007-0279-0

Hobza, R., Lengerova, M., Cernohorska, H., Rubes, J., and Vyskot, B. (2004). FAST-FISH with laser beam microdissected DOP-PCR probe distinguishes the sex chromosomes of *Silene latifolia*. *Chromosom. Res.* 12, 245–250. doi: 10.1023/B:CHRO.0000021929.97208.1c

Hobza, R., and Vyskot, B. (2007). Laser microdissection-based analysis of plant sex chromosomes. *Methods Cell Biol.* 82, 433–453. doi: 10.1016/S0091-679X(06)82015-7

Huber, D., Voith von Voithenberg, L., and Kaigala, G. V. (2018). Fluorescence in situ hybridization (FISH): history, limitations and what to expect from micro-scale FISH? *Micro Nano Eng.* 1, 15–24. doi: 10.1016/j.mne.2018.10.006

Hughes, J. F., Skaletsky, H., Pyntikova, T., Graves, T. A., van Daalen, S. K. M., Minx, P. J., et al. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* 463, 536–539. doi: 10.1038/nature08700

Hurel, A., Phillips, D., Vrielynck, N., Mézard, C., Grelon, M., and Christophorou, N. (2018). A cytological approach to studying meiotic recombination and chromosome dynamics in Arabidopsis thaliana male meiocytes in three dimensions. *Plant J.* 95, 385–396. doi: 10.1111/tpj.13942

Jiang, J. (2019). Fluorescence in situ hybridization in plants: recent developments and future applications. *Chromosom. Res.* 27, 153–165. doi: 10.1007/s10577-019-09607-z

Kazama, Y., Ishii, K., Aonuma, W., Ikeda, T., Kawamoto, H., Koizumi, A., et al. (2016). A new physical mapping approach refines the sex-determining gene positions on the *Silene latifolia* Y-chromosome. *Sci. Rep.* 6:18917. doi: 10.1038/srep18917

Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z., and Vyskot, B. (2009). The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* 102, 533–541. doi: 10.1038/hdy.2009.17

Kejnovský, E., Michalovova, M., Steflova, P., Kejnovska, I., Manzano, S., Hobza, R., et al. (2013). Expansion of microsatellites on evolutionary young Y chromosome. *PLoS One* 8:45519. doi: 10.1371/journal.pone.0045519

Krasovec, M., Chester, M., Ridout, K., and Filatov, D. A. (2018). The mutation rate and the age of the sex chromosomes in *Silene latifolia*. *Curr. Biol.* 28, 1832.e–1838.e. doi: 10.1016/j.cub.2018.04.069

Kubat, Z., Zluvova, J., Vogel, I., Kovacova, V., Cermak, T., Cegan, R., et al. (2014). Possible mechanisms responsible for absence of a retrotransposon family on a plant Y chromosome. *New Phytol.* 202, 662–678. doi: 10.1111/nph.12669

Lengerova, M., Kejnovsky, E., Hobza, R., Macas, J., Grant, S. R., and Vyskot, B. (2004). Multicolor FISH mapping of the dioecious model plant, *Silene latifolia*. *Theor. Appl. Genet.* 108, 1193–1199. doi: 10.1007/s00122-003-1568-6

Mathur, J., Griffiths, S., Barton, K., and Schattat, M. H. (2012). Chapter eight - green-to-red photoconvertible meosfp-aided live imaging in plants. *Methods Enzymol.* 504, 163–181. doi: 10.1016/B978-0-12-391857-4.00008-2

Ming, R., Wang, J., Moore, P. H., and Paterson, A. H. (2007). Sex chromosomes in flowering plants. *Am. J. Bot.* 94, 141–150. doi: 10.3732/ajb.94.2.141

Muyle, A., Shearn, R., and Marais, G. A. B. (2017). The evolution of sex chromosomes and dosage compensation in plants. *Genome Biol. Evol.* 9, 627–645. doi: 10.1093/gbe/evw282

Pan, W. H., Houben, A., and Schlegel, R. (1993). Highly effective cell synchronization in plant roots by hydroxyurea and amiprophos-methyl or colchicine. *Genome* 36, 387–390. doi: 10.1139/g93-053

Papadopulos, A. S. T., Chester, M., Ridout, K., and Filatov, D. A. (2015). Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13021–13026. doi: 10.1073/pnas.1508454112

Rodríguez Lorenzo, J. L., Hobza, R., and Vyskot, B. (2018). DNA methylation and genetic degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *BMC Genomics* 19:540. doi: 10.1186/s12864-018-4936-y

Runyeon, H., and Prentice, H. C. (1997). Genetic differentiation in the Bladder campions, *Silene vulgaris* and *S. uniflora* (Caryophyllaceae), in Sweden. *Biol. J. Linn. Soc.* 61, 559–584. doi: 10.1111/j.1095-8312.1997.tb01807.x

Schubert, V., Ruban, A., and Houben, A. (2016). Chromatin ring formation at plant centromeres. *Front. Plant Sci.* 7:28. doi: 10.3389/fpls.2016.00028

Šimoníková, D., Němečková, A., Karafiátová, M., Uwimana, B., Swennen, R., Doležel, J., et al. (2019). Chromosome painting facilitates anchoring reference genome sequence to chromosomes in situ and integrated karyotyping in banana (Musa Spp.). *Front. Plant Sci.* 10:1503. doi: 10.3389/fpls.2019.01503

Široký, J., Lysák, M. A., Doležel, J., Kejnovský, E., and Vyskot, B. (2001). Heterogeneity of rDNA distribution and genome size in *Silene* spp. *Chromosom. Res.* 9, 387–393. doi: 10.1023/A:1016783501674

Stone, J. D., Koloušková, P., Sloan, D. B., and Štorchová, H. (2017). Non-coding RNA may be associated with cytoplasmic male sterility in *Silene vulgaris*. *J. Exp. Bot.* 68, 1599–1612. doi: 10.1093/jxb/erx057

Tang, S., Tang, Z., Qiu, L., Yang, Z., Li, G., Lang, T., et al. (2018). Developing new oligo probes to distinguish specific chromosomal segments and the A, B, D genomes of wheat (*Triticum aestivum* L.) Using ND-FISH. *Front. Plant Sci.* 9:1104. doi: 10.3389/fpls.2018.01104

Uchida, W., Matsunaga, S., Sugiyama, R., Shibata, F., Kazama, Y., Miyazawa, Y., et al. (2002). Distribution of interstitial telomere-like repeats and their adjacent sequences in a dioecious plant, *Silene latifolia*. *Chromosoma* 111, 313–320. doi: 10.1007/s00412-002-0213-5

Vyskot, B., and Hobza, R. (2015). The genomics of plant sex chromosomes. *Plant Sci.* 236, 126–135. doi: 10.1016/j.plantsci.2015.03.019

Xin, H., Zhang, T., Han, Y., Wu, Y., Shi, J., Xi, M., et al. (2018). Chromosome painting and comparative physical mapping of the sex chromosomes in *Populus tomentosa* and *Populus deltoides*. *Chromosoma* 127, 313–321. doi: 10.1007/s00412-018-0664-y

Yoshida, K., Makino, T., Yamaguchi, K., Shigenobu, S., Hasebe, M., Kawata, M., et al. (2014). Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLoS Genet.* 10:e1004223. doi: 10.1371/journal.pgen.1004223

Zluvova, J., Georgiev, S., Janousek, B., Charlesworth, D., Vyskot, B., and Negrutiu, I. (2007). Early events in the evolution of the *Silene latifolia* Y chromosome: male specialization and recombination arrest. *Genetics* 177, 375–386. doi: 10.1534/genetics.107.071175

Zluvova, J., Janousek, B., Negrutiu, I., and Vyskot, B. (2005). Comparison of the X and Y chromosome organization in *Silene latifolia*. *Genetics* 170, 1431–1434. doi: 10.1534/genetics.105.040444

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

PERSPECTIVE
published: 10 March 2020
doi: 10.3389/fpls.2020.00251

Has the Polyploid Wave Ebbed?

*Donald A. Levin**

Department of Integrative Biology, University of Texas at Austin, Austin, TX, United States

There was a wave of whole genome duplications (WGD) during and subsequent to the K-Pg interface, which was followed by an increase in the proportion of species that were polyploid. I consider why this wave of polyploid speciation has continued to rise through the divergent evolution of polyploid lineages, and through rounds of homoploid and heteroploid chromosomal change. I also consider why the polyploid speciation wave is likely to rise in the next millennium. I propose that the speed of polyploid genesis through ploidal increase and through diversification among polyploids likely will be greater than the speed of diploid speciation. The increase in polyploid diversity is expected to lag well behind episodes of WGD, owing to the very long period required for species diversification either by lineage splitting or additional rounds of polyploidy, in addition to the long period of genomic adjustment to higher ploidal levels in neopolyploids.

Keywords: divergent evolution, diversification lag, dysploidy, extinction, polyploidy, reproductive isolation

OPEN ACCESS

Edited by:
Hanna Weiss-Schneeweiss, University of Vienna, Austria

Reviewed by:
Itay Mayrose, Tel Aviv University, Israel
Yves Van de Peer, Ghent University, Belgium

***Correspondence:**
Donald A. Levin
dlevin@austin.utexas.edu

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

Received: *11 October 2019*
Accepted: *18 February 2020*
Published: *10 March 2020*

Citation:
Levin DA (2020) Has the Polyploid Wave Ebbed? Front. Plant Sci. 11:251. doi: 10.3389/fpls.2020.00251

INTRODUCTION

The K-Pg interface was a period of intense global change associated with the mass extinction event that ended the Age of Dinosaurs (Nichols and Johnson, 2002). Terrestrial plant ecosystems were altered dramatically at the K-Pg interface 65.5 million years ago (Ma), in part because of the massive extinction of species, genera, and higher lineages (Wolfe and Upchurch, 1986; Vajda et al., 2001; Nichols and Johnson, 2002; McElwain and Punyasena, 2007). For example, paleobotanical studies of fossil leaves, pollen and spores from North American sites have shown that 18–30% of plant genera and up to 80% of all plant species died out during the K-Pg interface (Nichols and Johnson, 2002; Wilf and Johnson, 2004). Jablonski (1994) estimated that 47% of all marine genera died out during this period, as did 76% of all species.

Evolutionary rebounds following mass extinctions are important components of micro- and macroevolution (Jablonski, 1986; Benton, 1987). The recovery of biodiversity subsequent to mass extinctions is rapid by geological time scales, and frequently is coincident with significantly accelerated evolutionary rates (Miller and Sepkoski, 1988). The period during and subsequent to the K-Pg boundary was a particularly active one for plant polyploidizations, as there was a wave of whole genome duplications (WGD), i.e., a clustering of plant polyploidizations, during this time (Fawcett et al., 2009; Vanneste et al., 2014; Cannon et al., 2015). Many WGDs in different plant families occurred roughly between 60 and 65 million years ago (Van de Peer et al., 2017; Clark and Donoghue, 2018). The numerous independent WGD events across multiple angiosperm lineages may be correlated with major climatic perturbations, and are clustered at the base of some of the most successful and largest extant plant families and larger clades (Soltis et al., 2015). For example, Cai et al. (2019) found a significant correlation between the timing of WGD events in the Malpighiales and periods of global climatic change during the Paleocene–Eocene, ca. 56–54 million years ago.

Multiple adaptive WGDs have occurred in the history of very successful clades. For example, there have been 13 independent polyploid events in the Brassicaceae (Mandáková et al., 2017); and 43% of the species therein are neopolyploids (Hohmann et al., 2015). There have been 26 ancient and more recent polyploidy events in Caryophyllales (Yang et al., 2017); and numerous polyploidization events have occurred throughout the Asteraceae (Huang et al., 2016). Multiple, adaptive shifts involving physiological, morphological and/or ecological attributes have occurred in the annals of successful phylads. These shifts may be the immediate product of ploidal change, and/or they may be the result of selective responses within polyploid species. The more the environment changes, the more varied the types and the magnitude of adaptive shifts are likely to be. Changes in the physiological, morphological and/or ecological attributes within single phylads need not be concurrent, nor will a given type of change necessarily occur synchronously across phylads inhabiting the same environments.

THE PREMISE

If many WGDs in different plant families occurred within 10 million years or so after the K-Pg interface, and if there was a polyploid wave that commenced at that time, we may ask whether the polyploid wave has ebbed, and what the wave is likely to do in the future. I argue here that the polyploid wave has been growing since its inception, and that is likely to continue to do so.

Although WGD changes near the K-Pg boundary have received the most attention, there have been numerous pulses of polyploidy during the past 40 million years within several families. For example, repeated evolution of polyploids in the grass tribe Andropogoneae occurred during the expansion of C4 grasslands in the Late to Mid Miocene (Estep et al., 2014). A surge of polyploidy occurred in several tribes of the Asteraceae in the Early Middle Miocene (Huang et al., 2016), and in the Brassicaceae in the mid to late Miocene (Kagale et al., 2014). Jacob Landis (personal communication) estimated that 7 WGD events associated with diversification pulses occurred between 33 and 37 mya, and 6 such events between 12 and19 mya, compared to 14 WGD events between 50 and 60 mya.

Polyploidy has been an ongoing process. Some polyploids are roughly a million years old, as are allotetraploid *Melampodium strigosum* and two allohexaploids (McCann et al., 2018). Other polyploids have emerged within the past 25,000 years. For example, the allotetraploid *Arabidopsis kamchatica* has evolved around 20,000 years ago (Novikova et al., 2018). The tetraploid *A. arenosa* arose from one population about 15,000–19,000 years ago; *A. suecica* is about 16,000 years ago. Polyploid species also have arisen within the past 300 years. They include *Spartina anglica, Senecio squalidus, Senecio eboracensis, Senecio cambrensis, Mimulus peregrinus* (Thomas, 2015). *Tragopogon mirus* and *T. miscellus* are only about 100 years old (Soltis and Soltis, 2009). Some recent polyploids may prove to be the antecedents of a speciation surge within a genus or family; it depends on their persistence and rate of diversification.

Polyploid waves take millions of years to develop. Within a family or higher phylad, polyploidy originates within one genus, and then perhaps many others; and only much later polyploidy may become prominent. This process would be reflected in phylogenetic trees, where entire branches or groups of twigs will be polyploid.

As briefly summarized above, there is considerable documentation of the rise of polyploids subsequent to K-Pg interval, and their continued production from diploid progenitors. However, there are some aspects of the polyploid scenario which have received relatively little attention; and their consideration would expand our understanding of the polyploid dynamic. This paper considers three aspects, namely (Nichols and Johnson, 2002) mechanisms of polyploid diversification, Wolfe and Upchurch (1986) reasons why WGD are not followed by immediate diversification thrusts, and (Vajda et al., 2001) polyploidy going forward.

MECHANISMS OF POLYPLOID DIVERSIFICATION

Whereas the rapid transition from diploidy to tetraploidy has been the focal point in discussions of polyploid increase, ostensibly most polyploid species do not arise directly from diploid progenitors. There are two mechanisms which contribute most to an increase in the percentage of polyploid species. They are additional rounds of polyploidy, and divergence within polyploid lineages. Polyploid species can produce only more polyploids. They cannot beget diploid species, whereas diploid species can produce both diploid and polyploid species (Meyers and Levin, 2006). The percentage of speciation events that contributed to the origins of other polyploids through additional rounds of polyploidy or via divergent processes remains to be determined.

Consider first additional rounds of polyploidy. Hybridization between a tetraploid species and a diploid can yield sterile triploid hybrids, whose unreduced triploid gametes may fuse to produce fertile hexaploids (Ramsey and Schemske, 1998). Or sterile tetraploids from hybridization between chromosomally differentiated tetraploid species can produce unreduced tetraploid gametes which may fuse to produce fertile octaploid individuals. Some very high ploidal levels have been reached in genera such as *Viola*, in which *V. arborescens* is at least 20-ploid (2n = ca. 160), and two lineages are 14-ploid and one is 18-ploid lineage (Marcussen et al., 2012). Some genera are composed primarily of polyploid species [e.g., *Triticum* and *Festuca* (70%), and *Draba* (78%)] (Soltis et al., 2015). In old species complexes, members with lower ploidal levels may be absent. For example, in *Bromus* section Ceratochloa there are no diploid or tetraploid species, but there are 8-ploid and 12-ploid species (Grant, 1982). A shift from a predominance of diploid species in a species complex to one where polyploids prevail also may be indicative of polyploid superiority.

Polyploid species also may generate other such species via divergent evolution. Allopatric and peripatric speciation may occur within tetraploid or hexaploid genera (for example)

just as it does within diploid genera. Intraspecific lineages gradually diverge and accumulate genetic differences which confer ecological isolation or which result in the emergence of post-pollination barriers such as cross-incompatibility, hybrid inviability, and hybrid sterility. What we know about divergent evolution in diploid species via genetic change in response to disparate environmental challenges or opportunities also pertains to polyploids. The tempo of barrier building in polyploid lineages through genetic change alone is not well understood. In diploid lineages, strong ecological barriers typically emerge only after tens if not hundreds of thousands of years; and cross-incompatibility and hybrid sterility typically arise only after many millions of years (Levin, 2012, 2013).

The strength of post-zygotic isolation may be lower between polyploids than between diploids and polyploids (Hersch-Green, 2012; Hülber et al., 2015; Sutherland and Galloway, 2017). For example, in *Senecio carniolicus*, plants with intermediate ploidal levels were absent in diploid/tetraploid and diploid/hexaploid contact zones, but were frequent in a tetraploid/hexaploid zone (Hülber et al., 2015). In the *Aster occidentalis* complex, viable hybrids were obtained from 26% of the crosses between diploid species versus 66% of the crosses between polyploids (Allen et al., 1983). Tetraploid/hexaploid hybrids within *Jacobaea carnifolica* had greater seed germination and seedling viability than diploid/polyploid hybrids within the species (Sonnleitner et al., 2013). High ploidal entities may act as evolutionary "sponges," expanding their base of variation through introgression across or within ploidal levels (Baduel et al., 2018).

Speciation among polyploids also may involve chromosomal change without a shift in ploidal level (Leitch and Bennett, 2004; Leitch and Leitch, 2008). The fixation of chromosomal novelty most often will be associated with migration and population bottlenecks. Fixation involves stochastic processes, wherein the greater the bottleneck or the fewer the number of founders of new chromosomal populations, the lower is the fixation probability (Lande, 1985; Rieseberg, 2001; Jackson et al., 2016). Correlatively, the greater the level of relatedness among survivors of bottlenecks, the higher the likelihood that a new chromosomal arrangement will be fixed (Wright, 1969; Hedrick and Levin, 1984). The species prone to chromosome breakage could experience a series of novel rearrangements in one population that is unlikely to arise elsewhere. Moreover, their fixation would be associated with loss of genetic diversity, which would be exacerbated by subsequent migration/colonization bottlenecks (Excoffier et al., 2009). Accordingly, chromosomally differentiated lineages/species are likely to be narrowly distributed, short-lived evolutionary entities; and the new karyotype likely would have a narrow geographical footprint.

Past polyploidizations often have been followed by additions or subtractions in chromosome number within a ploidal level, i.e., dysploidy (Mandáková et al., 2017; Mandáková and Lysak, 2018). Descending dysploidy relies on translocations among homoelogous chromosomes and/or non-homologous chromosomes. Ascending dysploid based on centric fusions also is possible (e.g., in palms, Barrett et al., 2019), but it is less frequent than the descending mode. Dysploidy is one of the most important avenues to diploidization. Progressive diploidization may reduce chromosome numbers, such that the chromosome number of a species with a WGD in its history may be the same as relatives without such (e.g., n = 10 in maize with a WGD and in sorghum without WGD, 50). The extent and rate of dysploidy is expected to be positively correlated with ploidal level and the time since a ploidal increase (Mandáková and Lysak, 2018). Accordingly, ancient hexaploids should have higher levels of dysploidy than tetraploids. Indeed, this relationship was found across several clades in the Brassicaceae (Mandá ková et al., 2017).

If dysploidy is associated with polyploidy, it follows that there would a secondary wave of dysploidy lagging behind the post K-Pg polyploid wave. The magnitude of the dysploid wave remains to be determined. Since dysploid species may serve as platforms for additional rounds of polyploidy (Wendel, 2015), their presence likely added impetus to the polyploid wave. The dysploid wave actually may be much higher than the polyploid wave. Whereas the number of ploidal jumps that genera are likely to accommodate are limited, there could be well over ten euploid number shifts within a genus (Mandá ková et al., 2017; Mandáková and Lysak, 2018).

WHY WGD ARE NOT FOLLOWED BY IMMEDIATE DIVERSIFICATION THRUSTS

Ren et al. (2018) found that lineages with early WGD such as Brassicaceae Malvaceae, Fabaceae, Asteraceae and Poaceae have significantly greater species diversity than their sister lineages. They also detected that species radiation was higher in orders with a greater percentage with WGDs than smaller orders for Asparagales, core Lamiales, and core monocots. Correlatively, in *Allium* higher species diversification rates occurred in lineages with high polyploid frequencies (Han et al., 2020). Whether polyploidy is a universal promoter of species diversification remains to be determined.

If polyploidy promotes species diversification, there are several reasons why that boost probably would not be immediate. It has been proposed that diversification via ploidal change requires genome stabilization through diploidization, the genesis of novel key traits, and the buildup of genetic diversity; and these processes take a millions of years (Landis et al., 2018). The pace of diploidization varies among phylads and components therein, as does the marshaling of genetic diversity. Once the genome has settled and variation has accrued, speciation may proceed through the evolution of ecologically and geographically distinctive intraspecific lineages. Subsequent genetic and/or chromosomal divergence would lead to their reproductive isolation (Clausen, 1951; Levin, 2000). Even additional rounds of polyploidy would take a long time to have a substantive effect on diversity, because neopolyploids are likely to be short-lived, and because the products of one episode of ploidal increase may be similar to those of another (Levin and Soltis, 2018). The components of some species may exploit similar types of habitats due to genetic constraints and poor niche availability or opportunities. As such, their subsequent rate of

radiation will be retarded relative to more ecologically flexible species. These species most likely will originate from multiple origins in which the parental entities were adapted to somewhat different habitats. Finally, the breeding system may contribute to the diversification lag in some clades. Families that are exclusively self-compatible have lower net diversification rates than those that are predominantly self-incompatible (Ferrer and Good, 2012). Ostensibly selfing species having higher extinction rates than outcrossers (Wright et al., 2013) in part due to a reduced ability to adapt, especially to new environments (Hartfield et al., 2017).

The absence of immediate speciation acceleration after polyploidy also may be the product of species extinction, because diversification is measured as speciation minus extinction. Consider first a tree (phylogenetic or otherwise) with a single stem. Branching into two axes may subsequently occur, and each branch (A and B) may bear additional branches, and so forth. However, for tree complexity to increase, each new branch must remain on the tree. Branch loss will delay or prevent subsequent tree radiation. The greater the incidence of branch loss the longer the time for a complex tree to develop. It follows that species extinction would thwart diversification. The establishment of a polyploid genus may be followed by long intervals in which only one or a few species remain viable. Then the subsequent generation of diversity would lag well behind the establishment of the first polyploid(s). If a genus radiated from an original polyploid, and then experienced a major and prolonged contraction, a subsequent expansion would lag behind the origin of the first polyploid.

The likelihood of polyploid species extinction is an inverse function of the number of ecogeographically differentiated lineages within a given species (Rosenblum et al., 2012). Divergent gene pools offer alternate points of departure for selective responses to different abiotic and biotic challenges, thus improving chances of local survival or migration in the face of environmental change (Oney et al., 2013; Moran et al., 2016). This is because ecogeographically differentiated lineages are unlikely to have similar geographical footprints. Indeed these lineages may live in different areas; and they are unlikely to respond to environmental change in the same manner or degree (Bennett and Provan, 2008; Stewart, 2009; Stewart et al., 2010). Polyploid species with many highly divergent lineages are likely to have broader ranges than species with few such lineages (Gaston and Fuller, 2009; Morin and Lechowicz, 2013; Slatyer et al., 2013). Range size tends to be positively correlated with species longevity (McKinney, 1997; Saupe et al., 2015).

POLYPLOIDY GOING FORWARD

Given that the percentage of polyploids has been increasing, what might be expected of the contribution of polyploids to future flowering plant diversity? I propose that the percentage of species that are polyploids almost certainly will increase during the next several millennium. If 25 to 30% of extant flowering plants are recent polyploids (Wood et al., 2009; Barker et al., 2016), then several millennia from now it would not be surprising to see that number approach 35–40%. A significant increase is likely because the factors that contributed to the Anthropocene polyploid formation such as habitat disturbance, transport, and domestication (Thomas, 2015; Bull and Maron, 2016; Vellend et al., 2017; Otto, 2018) are likely to have a larger impact in the future. Moreover climate change will alter community structure and create new habitats in which polyploids may thrive (Vanneste et al., 2014; Van de Peer et al., 2017). Hundreds of species in various parts of the world have already become locally extinct (Wiens, 2016). The background rate of extinction is 1000–10,000 times higher than it has ever been (Pimm and Joppa, 2015). The extent to which polyploids have outlived diploids over the past 5000 years remains to be determined.

Polyploidy can produce the antecedents of new species in one generation (autopolyploidy) or two generations (allopolyploidy) through the production of unreduced gametes or doubled somatic cells (Ramsey and Schemske, 1998; Mason and Pires, 2015). Autopolyploids can arise anywhere in a species' range from single individuals via the formation of unreduced gametes. Because autopolyploid genesis bypasses the requirement for species contact and hybridization, autopolyploids almost certainly will evolve much more frequently than allopolyploids. This is not to suggest that over the longer term autopolyploid species will prevail, which they do not (Barker et al., 2016). Allopolyploids tend to have broader niche tolerances than autopolyploids, and to be more ecologically divergent from their progenitors than are autopolyploids (Levin, 2002).

Speciation among polyploids also may involve chromosomal change without a shift in ploidal level. The fixation of chromosomal novelty and dysploid numbers most often will be associated with migration and population bottlenecks. Such change may occur in only a few hundred generations or less; and this form of speciation may be more important than a change in ploidal level. New polyploids also may originate from existing polyploids via the divergence of gene pools without chromosomal change. This is likely to be the least important speciation mode over the next few millennia, because ecological and incompatibility barriers take very long periods to evolve, as discussed above. It is possible, however, that some divergent lineages within species will become reproductively isolated within the next few thousand years.

A persistence/extinction differential between diploid and polyploid species would have an impact on the ploidal balance. If polyploid species were more tolerant of the upcoming climatic changes than diploids, as they seem to be and have been (Fawcett et al., 2009), then polyploid representation would increase. It is likely that the climatic shocks near and following the K-Pg boundary contributed to a diploid survival trough, and to a decline in the proportion of diploid species (Vanneste et al., 2014; Van de Peer et al., 2017; Levin and Soltis, 2018). If polyploids increase proportionally, then it will be most pronounced in herbaceous species, whose rates of chromosome doubling were six times higher than in woody species (Zenil-Ferguson et al., 2017). A polyploid elevation also is likely to be more pronounced in perennial herbs than in annuals (Otto and Whitton, 2000).

If polyploids had a higher survival rate than diploids over the past 5000 years, then we would surmise that the

diploid:polyploid balance would be tilted toward the latter in the future. Unfortunately past survival rates are unknowable. However, an insight into a possible differential may be obtained from a comparison of the ploidal levels of endangered species and their invasive relatives. Endangered plants are more likely to be diploid than their invasive congeners or invasive species as a whole (Pandit, 2006; Pandit et al., 2011). Pandit et al. (2011) found that being endangered is 14% more likely for diploids than for polyploids. The bases for this differential remains to be determined. Polyploids might benefit from the immediate effects of ploidal increase alone, or through a higher level of genetic diversity, or a combination of both. Perhaps polyploids will have a higher survival rate than they did during preceding periods.

CONCLUSION

The polyploid wave which began roughly 60 mya continues to rise. Pulses of polyploidy have occurred many times since then. Roughly 30% of all flowering plants are polyploid; and this number is likely to increase in the coming several millennia, perhaps approaching 40%. The factors promoting polyploidy in the past surely will be important in the future. These factors include climate and environmental change, which lead to community disassembly, and migration and contact between previously isolated congeners. Anthropogenic drivers will have an increasing impact going forward. In addition to elevated polyploid production, the percentage of polyploid species is likely to increase because diploid species may be more vulnerable to environmental change than polyploids.

The increase in polyploidy in the coming millennia will involve not only the doubling of diploid chromosome numbers, but also the genesis of higher level polyploids and dysploids from existing polyploids. A wave of dysploidy almost certainly accompanied the polyploid wave of the past, because polyploids are much more prone to chromosomal change than diploids, and because one dysploid species could parent another with a different chromosome number. Divergent evolution within a polyploid or dysploid species also may propel the genesis of more polyploids. However, that process is slow, and thus unlikely to yield large numbers of species within a few thousand years.

Some products of ploidal shifts several million years ago were the antecedents of major lineages/clades. However, diversification in the latter lagged well the WGD themselves, because new polyploids tend to have high extinction rates regardless of mode of origin, and because diversification in its various forms takes a very long time. Accordingly, new polyploid species that evolve within the next few millennia are unlikely to parent many species by the end of that period. Over millions of years, however, some may prove to be the antecedents of major lineages/clades.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The author is most grateful to Jacob Landis for permitting reference to his unpublished work on temporal aspects of WGD, and to Doug Soltis for his critical reading of the manuscript. The author also thanks the reviewers for their thoughtful comments.

REFERENCES

Allen, G. A., Dean, M. L., and Chambers, K. L. (1983). Hybridization studies in the *Aster occidentalis* (Asteraceae) polyploid complex of western North America. *Brittonia* 35, 353–361.

Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F., and Yant, L. (2018). The "Polyploid Hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* 6:117. doi: 10.3389/fevo.2018.00117

Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., and Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 210, 391–398. doi: 10.1111/nph.13698

Barrett, C. F., McKain, M. R., Sinn, B. T., Ge, X. J., Zhang, Y., Antonelli, A., et al. (2019). Ancient polyploidy and genome evolution in palms. *Genome Biol. Evol.* 11, 1501–1511. doi: 10.1093/gbe/evz092

Bennett, K. D., and Provan, J. (2008). What do we mean by 'refugia'? *Quat. Sci. Rev.* 27, 2449–2455. doi: 10.1016/j.quascirev.2008.08.019

Benton, M. J. (1987). Progress and competition in macroevolution. *Biol. Rev.* 62, 305–338. doi: 10.1111/j.1469-185x.1987.tb00666.x

Bull, J. W., and Maron, M. (2016). How humans drive speciation as well as extinction. *Proc. R. Soc. Lond. B Biol. Sci.* 283, 2016060. doi: 10.1098/rspb.2016.0600

Cai, L., Xi, Z., Amorim, A. M., Sugumaran, M., Rest, J. S., Liu, L., et al. (2019). Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytol.* 221, 565–576. doi: 10.1111/nph.15357

Cannon, S. B., McKain, M. R., Harkess, A., Nelson, M. N., Dash, S., Deyholos, M. K., et al. (2015). Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol. Biol. Evol.* 32, 193–210. doi: 10.1093/molbev/msu296

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

Clausen, J. (1951). *Stages in the Evolution of Plant Species*. Ithaca, NY: Cornell University Press.

Estep, M. C., McKain, M. R., Diaz, D. V., Zhong, J. S., Hodge, J. G., Hodkinson, T. R., et al. (2014). Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15149–15154. doi: 10.1073/pnas.1404177111

Excoffier, L., Foll, M., and Petit, R. J. (2009). Genetic consequences of range expansions. *Ann. Rev. Ecol. Evol. Syst.* 40, 481–501. doi: 10.1146/annurev.ecolsys.39.110707.173414

Fawcett, J. A., Maere, S., and Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5737–5742. doi: 10.1073/pnas.0900906106

Ferrer, M. M., and Good, S. V. (2012). Self-sterility in flowering plants: preventing self-fertilization increases family diversification rates. *Ann. Bot.* 110, 535–553. doi: 10.1093/aob/mcs124

Gaston, K. J., and Fuller, R. A. (2009). The sizes of species' geographic ranges. *J. Appl. Ecol.* 46, 1–9. doi: 10.1111/j.1365-2664.2008.01596.x

Grant, V. (1982). *Plant Speciation*, 2nd Edn. New York, NY: Columbia University Press.

Han, T.-S., Zheng, Q. J., Onstein, R. E., Rojas-Andrés, B. M., Hauenschild, F., Muellner-Riehl, A. N., et al. (2020). Polyploidy promotes species diversification of *Allium* through ecological shifts. *New Phytol.* 225, 571–583. doi: 10.1111/nph.16098

Hartfield, M., Bataillon, T., and Glémin, S. (2017). The evolutionary inter— play between adaptation and self—fertilization. *Trends Genet.* 33, 420–431. doi: 10.1111/evo.13426

Hedrick, P. W., and Levin, D. A. (1984). Kin-founding and the fixation of chromosomal variants. *Am. Nat.* 124, 789–797. doi: 10.1086/284317

Hersch-Green, E. (2012). Polyploidy in Indian paintbrush (*Castilleja*; Orobanchaceae) species shapes but does not prevent gene flow across species boundaries. *Am. J. Bot.* 99, 1680–1690. doi: 10.3732/ajb.1200253

Hohmann, N., Wolf, E. M., Lysak, M. A., and Koch, M. A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* 27, 2770–2784. doi: 10.1105/tpc.15.00482

Huang, C. H., Zhang, C., Liu, M., Hu, Y., Gao, T., Qi, J., et al. (2016). Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol. Biol. Evol.* 33, 2820–2835. doi: 10.1093/molbev/msw157

Hülber, K., Sonnleitner, M., Suda, J., Krej € cıkova, J., Sch€onswetter, P., Schneeweiss, G. M., et al. (2015). Ecological differentiation, lack of hybrids involving diploids, and asymmetric gene flow between polyploids in narrow contact zones of *Senecio carniolicus* syn. *Jacobaea carniolica*, Asteraceae. *Ecol. Evol.* 56, 1224–1234. doi: 10.1002/ece3.1430

Jablonski, D. (1986). "Evolutionary consequences of mass extinctions," in *Patterns and Processes in the History of Life*, eds D. M. Raup and D. Jablonski (Berlin: Springer-Verlag), 313–329. doi: 10.1007/978-3-642-70831-2_17

Jablonski, D. (1994). Extinctions in the fossil record. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 344, 11–17.

Jackson, B., Butlin, R., Navarro, A., and Fria, R. (2016). "Speciation, chromosomal rearrangements and," in *Encyclopedia of Evolutionary Biology*, ed. R. M. Kliman (Cambridge, MA: Elsevier), 149–158. doi: 10.1016/b978-0-12-800049-6.00074-3

Kagale, S., Robinson, S. J., Nixon, J., Xiao, R., Huebert, T., Condie, J., et al. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell* 26, 2777–2791. doi: 10.1105/tpc.114.126391

Lande, R. (1985). The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity* 54, 323–332. doi: 10.1038/hdy.1985.43

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

Leitch, A. R., and Leitch, I. J. (2008). Genomic plasticity and the diversity of polyploid plants. *Science* 320, 481–483. doi: 10.1126/science.1153585

Leitch, I. J., and Bennett, M. D. (2004). Genome downsizing in polyploid plants. *Biol. J. Linn. Soc.* 82, 651–663. doi: 10.1111/j.1095-8312.2004.00349.x

Levin, D. A. (2000). *The Origin, Expansion, and Demise of Plant Species*. New York, NY: Oxford University Press.

Levin, D. A. (2002). *The Role of Chromosomal Change in Plant Evolution*. New York, NY: Oxford University Press.

Levin, D. A. (2012). The long wait for hybrid sterility in flowering plants. *New Phytol.* 196, 666–670. doi: 10.1111/j.1469-8137.2012.04309.x

Levin, D. A. (2013). The timetable for allopolyploidy in flowering plants. *Ann. Bot.* 112, 1201–1208. doi: 10.1093/aob/mct194

Levin, D. A., and Soltis, D. E. (2018). Factors promoting polyploid persistence and diversification and limiting diploid speciation during the K-Pg interlude. *Curr. Opin. Plant Biol.* 42, 1–7. doi: 10.1016/j.pbi.2017.09.010

Mandá ková, T., Pouch, M., Harmanová, K., Zhan, S. H., Mayrose, I., and Lysak, M. A. (2017). Multi-speed genome diploidization and diversification after an ancient allopolyploidization. *Mol. Ecol.* 26, 6445–6462. doi: 10.1111/mec.14379

Mandáková, T., Li, Z., Barker, M. S., and Lysak, M. A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *Plant J.* 91, 3–21. doi: 10.1111/tpj.13553

Mandáková, T., and Lysak, M. A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Curr. Opin. Plant Biol* 42, 55–65. doi: 10.1016/j.pbi.2018.03.001

Marcussen, T., Jakobsen, K. S., Danihelka, J., Ballard, H. E., Blaxland, K., Brysting, A. K., et al. (2012). Inferring species networks from gene trees in high-polyploid North American and Hawaiian violets (*Viola*, Violaceae). *Syst. Biol.* 61, 107–126. doi: 10.1093/sysbio/syr096

Mason, A. S., and Pires, P. C. (2015). Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends Genet.* 31, 5–10. doi: 10.1016/j.tig.2014.09.011

McCann, J., Jang, T.-S., Macas, J., Schneeweiss, G. M., Matzke, N. J., Novák, P., et al. (2018). Dating the species network: allopolyploidy and repetitive DNA evolution in American daisies (*Melampodium* sect. *Melampodium,* Asteraceae). *Syst. Biol.* 67, 1010–1024. doi: 10.1093/sysbio/syy024

McElwain, J. C., and Punyasena, S. W. (2007). Mass extinction events and the plant fossil record. *Trends Ecol. Evol.* 22, 548–557. doi: 10.1016/j.tree.2007.09.003

McKinney, M. L. (1997). Extinction vulnerability and selectivity: combining ecological and paleontological views. *Ann. Rev. Ecol. Syst.* 28, 495–516. doi: 10.1146/annurev.ecolsys.28.1.495

Meyers, L. A., and Levin, D. A. (2006). On the abundance of polyploids in flowering plants. *Evolution* 60, 1198–1206. doi: 10.1111/j.0014-3820.2006.tb01198.x

Miller, A. I., and Sepkoski, J. J. Jr. (1988). Modeling bivalve diversification: the effect of interaction on a macroevolutionary system. *Paleobiology* 14, 364–369. doi: 10.1017/s0094837300012100

Moran, E. V., Hartig, F., and Bell, D. M. (2016). Intraspecific trait variation across scales: implications for understanding global change responses. *Glob. Change Biol.* 22, 137–150. doi: 10.1111/gcb.13000

Morin, X., and Lechowicz, M. J. (2013). Niche breadth and range area in North American trees. *Ecography* 36, 300–312. doi: 10.1111/j.1600-0587.2012.07340.x

Nichols, D. J., and Johnson, K. R. (2002). Palynology and microstratigraphy of Cretaceous-Tertiary boundary sections in southwestern North Dakota. *Geol. Soc. Am. Spec. Pap.* 361, 95–143.

Novikova, P. Y., Hohmann, N., and Van de Peer, Y. (2018). Polyploid *Arabidopsis* species originating around recent glaciation maxima. *Curr. Opin. Plant Biol* 42, 8–15. doi: 10.1016/j.pbi.2018.01.005

Oney, B., Reineking, B., O'Neill, G., and Kreyling, J. (2013). Intraspecific variation buffers projected climate change impacts on *Pinus contorta*. *Ecol. Evol.* 3, 437–449. doi: 10.1002/ece3.426

Otto, S. P. (2018). Adaptation, speciation and extinction in the Anthropocene. *Proc. R. Soc. Lond. B Biol. Sci.* 285:20182047. doi: 10.1098/rspb.2018.2047

Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Ann. Rev. Genet.* 34, 401–437. doi: 10.1146/annurev.genet.34.1.401

Pandit, M. K. (2006). Continuing the search for pattern among rare plants: are diploid species more likely to be rare? *Evol. Ecol. Res.* 8, 543–552.

Pandit, M. K., Pocock, M. J. O., and Kunin, W. E. (2011). Ploidy influences rarity and invasiveness in plants. *J. Ecol.* 99, 1108–1115. doi: 10.1111/j.1365-2745.2011.01838.x

Pimm, S. L., and Joppa, L. N. (2015). How many plant species are there, where are they, and at what rate are they going extinct? *Ann. Mo. Bot. Gard.* 100, 170–176. doi: 10.3417/2012018

Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Ann. Rev. Ecol. Syst.* 29, 467–502. doi: 10.1104/pp.16.01768

Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., et al. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* 11, 414–428. doi: 10.1016/j.molp.2018.01.002

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16, 351–358. doi: 10.1016/s0169-5347(01)02187-5

Rosenblum, E. B., Sarver, B. A., Brown, J. W., Des Roches, S., Hardwick, K. M., Hether, T. D., et al. (2012). Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. *Evol. Biol.* 39, 255–261. doi: 10.1007/s11692-012-9171-x

Saupe, E. E., Qiao, H., Hendricks, J. R., Portell, R. W., Hunter, S. J., Soberón, J., et al. (2015). Niche breadth and geographic range size as determinants of species survival on geological time scales. *Glob. Ecol. Biogeogr.* 24, 1159–1169. doi: 10.1111/geb.12333

Slatyer, R. A., Hirst, M., and Sexton, J. P. (2013). Niche breadth predicts geographic range size: a general ecological pattern. *Ecol. Lett.* 16, 1104–1114. doi: 10.1111/ele.12140

Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125.

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Ann. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Sonnleitner, M., Weis, B., Flatscher, R., García, P. E., Suda, J., Krejčíková, J., et al. (2013). Parental ploidy strongly affects offspring fitness in heteroploid crosses among three cytotypes of autopolyploid *Jacobaea carniolica* (Asteraceae). *PLoS One* 8:e78959. doi: 10.1371/journal.pone.0078959

Stewart, J. R. (2009). The evolutionary consequence of the individualistic response to climate change. *J. Evol. Biol.* 22, 2363–2375. doi: 10.1111/j.1420-9101.2009.01859.x

Stewart, J. R., Lister, A. M., Barnes, I., and Dalen, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proc. R. Soc. Lond. B Biol. Sci.* 277, 661–671. doi: 10.1098/rspb.2009.1272

Sutherland, B. L., and Galloway, L. F. (2017). Postzygotic isolation varies by ploidy level within a polyploid complex. *New Phytol.* 213, 404–412. doi: 10.1111/nph.14116

Thomas, C. D. (2015). Rapid acceleration of plant speciation during the Anthropocene. *Trends Ecol. Evol.* 30, 448–455. doi: 10.1016/j.tree.2015.05.009

Vajda, V., Raine, J. I., and Hollis, C. J. (2001). Indication of global deforestation at the Cretaceous-Tertiary boundary by New Zealand fern spike. *Science* 294, 1700–1702. doi: 10.1126/science.1064706

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

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res.* 24, 1334–1347. doi: 10.1101/gr.168997.113

Vellend, M., Baeten, L., Becker-Scarpitta, A., Boucher-Lalonde, V., McCune, J. L., Messier, J., et al. (2017). Plant biodiversity change across scales during the Anthropocene. *Ann. Rev. Plant Biol.* 68, 563–586. doi: 10.1146/annurev-arplant-042916-040949

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320

Wiens, J. J. (2016). Climate-related local extinctions are already wide-spread among plant and animal species. *PLoS Biol.* 14:e2001104. doi: 10.1371/journal.pbio.2001104

Wilf, P., and Johnson, K. R. (2004). Land plant extinction at the end of the Cretaceous: a quantitative analysis of the North Dakota megafloral record. *Paleobiologyogy* 30, 347–368. doi: 10.1666/0094-8373(2004)030<0347:lpeate>2.0.co;2

Wolfe, J. A., and Upchurch, G. R. (1986). Vegetation, climatic and floral changes at the Cretaceous-Tertiary boundary. *Nature* 324, 148–152. doi: 10.1038/324148a0

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., Rieseberg, L. H., et al. (2009). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13875–13879. doi: 10.1073/pnas.0811575106

Wright, S. (1969). *The Theory of Gene Frequencies. Evolution and the Genetics of Populations: A Treatise in Three Volumes*. Chicago: The University of Chicago Press.

Wright, S. I., Kalisz, S., and Slotte, T. (2013). Evolutionary consequences of self-fertilization in plants. *Proc. R. Soc. Lond. B Biol. Sci.* 280:20130133. doi: 10.1098/rspb.2013.0133

Yang, Y., Moore, M., Brockington, S., Mikenas, J., Olivieri, J., Walker, J., et al. (2017). Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytol.* 217, 855–870. doi: 10.1111/nph.14812

Zenil-Ferguson, R., Ponciano, J. M., and Burleigh, J. G. (2017). Testing the association of phenotypes with polyploidy: an example using herbaceous and woody eudicots. *Evolution* 51, 1138–1148. doi: 10.1111/evo.13226

Conflict of Interest: The author declares 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

REVIEW
published: 18 March 2020
doi: 10.3389/fpls.2020.00296

Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants

Kaichi Huang and Loren H. Rieseberg**

Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada

Chromosomal inversions have the potential to play an important role in evolution by reducing recombination between favorable combinations of alleles. Until recently, however, most evidence for their likely importance derived from dipteran flies, whose giant larval salivary chromosomes aided early cytogenetic studies. The widespread application of new genomic technologies has revealed that inversions are ubiquitous across much of the plant and animal kingdoms. Here we review the rapidly accumulating literature on inversions in the plant kingdom and discuss what we have learned about their establishment and likely evolutionary role. We show that inversions are prevalent across a wide range of plant groups. We find that inversions are often associated with locally favored traits, as well as with traits that contribute to assortative mating, suggesting that they may be key to adaptation and speciation in the face of gene flow. We also discuss the role of inversions in sex chromosome formation, and explore possible parallels with inversion establishment on autosomes. The identification of inversion origins, as well as the causal variants within them, will advance our understanding of chromosomal evolution in plants.

Keywords: inversions, comparative genomics, reduced recombination model, secondary contact, comparative genetic mapping

OPEN ACCESS

Edited by:
Martin A. Lysak,
Masaryk University, Czech

Reviewed by:
Ingo Schubert,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany
Hans De Jong,
Wageningen University & Research, Netherlands
David B. Lowry,
Michigan State University, United States

****Correspondence:***
Kaichi Huang
kaichi.huang@botany.ubc.ca
Loren H. Rieseberg
lriesebe@mail.ubc.ca

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

Received: *10 December 2019*
Accepted: *27 February 2020*
Published: *18 March 2020*

Citation:
Huang K and Rieseberg LH (2020) Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants. Front. Plant Sci. 11:296. doi: 10.3389/fpls.2020.00296

INTRODUCTION

Species and ecotypes are often differentiated by chromosomal rearrangements, such as translocations and inversions. The latter were first discovered by Sturtevant (1921) when comparing genetic linkage maps of closely related *Drosophila* species. Sturtevant further deduced that inversions reduce the rate of recombination in heterozygotes (which is key to their main evolutionary role), and validated this claim through observations of the giant larval salivary chromosomes found in *Drosophila*. Inverted regions were subsequently identified from banding patterns of chromosomes in many other species and became the first genetic markers used to reconstruct phylogenies (Krimbas and Powell, 1992). The abundance of inversion polymorphisms detected in these studies also inspired population geneticists to investigate patterns of variation within and between species of *Drosophila* (Dobzhansky, 1970).

Until recently, most evidence regarding the frequency and evolutionary role of inversions came from studies of Dipteran systems, such as *Drosophila* (Noor et al., 2001; Ortiz-Barrientos et al., 2002), *Anopheles* (Ayala and Coluzzi, 2005; Ayala et al., 2010) and *Rhagoletis* (Feder et al., 2003a,b). This was partly due to the ease of identifying inversions in Dipteran salivary gland chromosomes, but also because of the widespread recognition of the importance of inversions in this group.

Over the past two decades, however, comparative genetic mapping and genomic approaches have revealed that inversions are ubiquitous across the plant and animal kingdoms, either fixed between or polymorphic within species (Wellenreuther and Bernatchez, 2018). Detailed studies of the genetic contents and establishment of inversions have not only confirmed a longstanding hypothesis that inversions play an important role in adaptation by reducing recombination between favorable combinations of alleles (Kirkpatrick and Barton, 2006; Lowry and Willis, 2010), but also that they contribute to speciation in a similar way - by suppressing recombination between local adapted alleles and those causing assortative mating (Trickett and Butlin, 1994).

In this paper, we first review various approaches that have been employed to detect inversions within and between plant species, as well as studies that report on inversion abundance across a wide range of plant groups. We then discuss possible scenarios for the origin and spread of inversions inspired by theoretical and empirical studies. We further illustrate their important role in speciation with case studies that have associated inversions with traits known to underlie ecological adaptation and reproductive isolation. We also discuss the role of inversions in sex chromosome formation and whether the stepwise establishment of inversions seen on sex chromosomes might also occur on autosomes. Finally, we suggest avenues for future studies to bridge gaps in our understanding of the evolutionary role of inversions in plants.

DETECTION OF INVERSIONS AND THEIR PERVASIVENESS IN PLANTS

Cytogenetic Studies

Most early evidence of inversions in plants came from two sources. First, as alluded to above, inversions could sometimes be inferred from the chromosome banding patterns seen in karyotypes (Greilhuber and Speta, 1976; Konishi and Linde-Laursen, 1988; Rodriguez et al., 2000). This approach worked well in plants with small numbers of large and distinctive chromosomes, but was impractical in most species of plants. Also, even when feasible, only very large inversions were typically detectable. Nonetheless, these karyotypic analyses indicated that inversions were not uncommon in plants.

A second source of information about inversions came from studies of chromosome pairing in meiosis (Lewis and Roberts, 1956; Sybenga, 1975; Ahmad et al., 1979; Gopinathan and Babu, 1986; Anderson et al., 2010). This approach relied on that fact that crossing over in inversion heterozygotes creates distinctive meiotic configurations (Sybenga, 1975; **Figure 1**). While such an approach is feasible in taxa with small and/or morphologically similar chromosomes, it under-estimates inversion abundance because recombination within inversions is required for their detection. Thus, small inversions, or those in low-recombining regions of the genome, will be missed.

In recent decades, *in situ* hybridization approaches have been widely employed to study karyotype evolution within and between species. Such approaches are most powerful in groups such as the Brassicaceae and Solanaceae, in which virtually repeat-free BAC contigs covering much of the genome are available for use as probes, permitting "comparative chromosome mapping" (Lysak and Lexer, 2006). Successful application of this method has led to the discovery of numerous inversions across various clades of the Brassicaceae, but especially *Arabidopsis* and *Brassica* (Lysak et al., 2006, 2007; Mandakova and Lysak, 2008; Mandakova et al., 2015; Lee et al., 2017), as well as among and within species in *Solanum* (Szinay et al., 2012). However, suitable sets of chromosome-specific painting probes are needed for the broader application of this approach in other plant groups.

Comparative Genetic Mapping

With the development of DNA markers in the latter part of the 20th century, it became feasible to develop genetic linkage maps that were sufficiently dense to permit detection of chromosomal rearrangements between plant genomes. More recently, advances in high throughput sequencing and computational methods permit high-resolution genetic mapping and inference of large structural variants from low coverage sequence data (e.g., Flagel et al., 2019). Comparative genetic mapping has been broadly applied, especially in crop rich families such as Solanaceae, Poaceae, and Brassicaceae. While successful, the number of rearrangements detected depends in part on marker density and recombination rates. If either is low, then rearrangements will be missed. We summarize a few well known examples below, both to illustrate that inversions are common in essentially all plant taxa analyzed, but also for comparison to genomic studies (below), which show that comparative mapping, like chromosome banding and meiotic analyses, greatly under-estimated inversion numbers.

In Solanaceae, for example, the genomes of potato and tomato were found to differ by only five paracentric inversions (Bonierbale et al., 1988; Tanksley et al., 1992), while at least 19 inversions and 6 chromosome translocations differentiate potato and pepper (Wu et al., 2009; summarized in Szinay et al., 2012). In Poaceae, genetic maps based on restriction fragment length polymorphism markers identified an inversion on the short arm of chromosome 9 between sorghum and maize, which also differentiates maize from its close relative *Zea mexicana* (Berhan et al., 1993). Likewise, Ahn and Tanksley (1993) showed that while the rice and maize genomes are largely conserved in gene order, multiple inversions and translocations occurred after the polyploidization of maize. Analyses of different intraspecific maps of *Brassica oleracea* revealed that small inversions among morphotypes were the most frequent form of rearrangements followed by translocations (Kianian and Quiros, 1992). A comparison between *Arabidopsis thaliana* and *Brassica nigra* also identified numerous translocations and even more inversions between the two genera (Lagercrantz, 1998).

The comparative mapping approach has also been applied to ecological and evolutionary model systems, including *Mimulus*, *Populus*, and *Helianthus*. This work has identified numerous chromosomal rearrangements from the ecotypic to interspecific level. For example, Lowry and Willis (2010) discovered an inversion between annual and perennial ecotypes of *Mimulus guttatus* by comparing maps from multiple F2 mapping

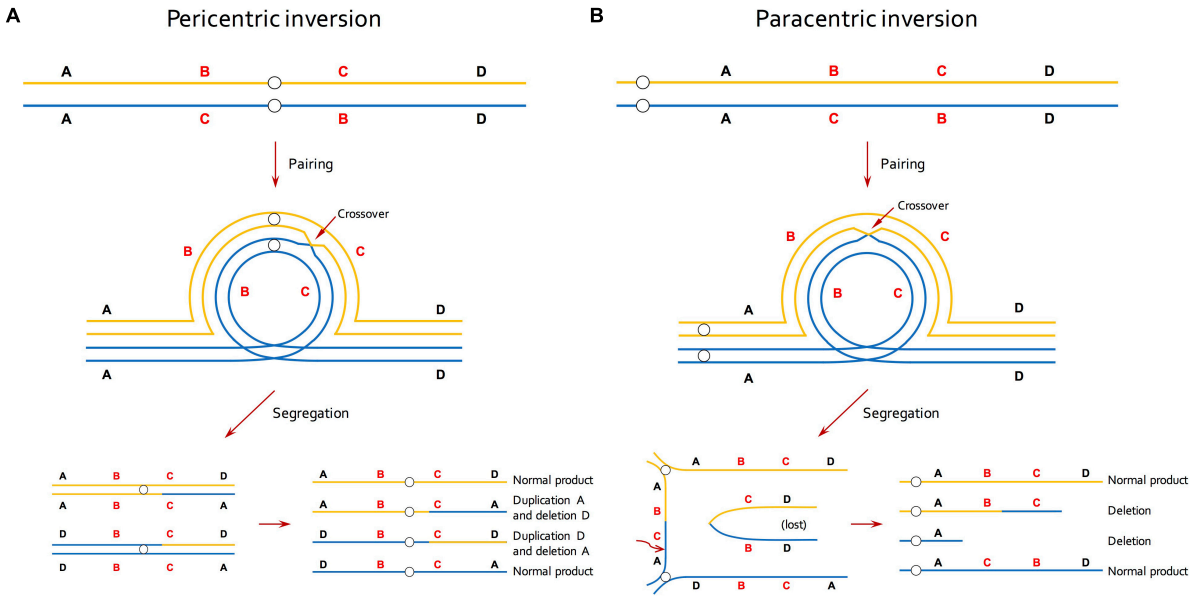

FIGURE 1 | Effective reduction in recombination of inversion by selection against recombinant gametes in meiosis. **(A)** In individuals that are heterozygous for a pericentric inversion, a single crossover within the inversion generates unbalanced gametes that contain a duplication and a deletion. **(B)** In individuals that are heterozygous for a paracentric inversion, a single crossover within the inversion produces a dicentric bridge and an acentric fragment. The acentric fragment is lost because it cannot be drawn to either end and the chromosomal bridge breaks at random point during segregation, resulting in two deletion products. Lines of blue and orange colors represent homologous chromosomes and small circles indicate centromeres.

populations. Using a similar approach, Fishman et al. (2013) identified two reciprocal translocations and three inversions between *Mimulus cardinalis* and *M. lewisii*. Using evidence from both linkage and physical maps, possible inversions were also inferred between *Populus* species (Drost et al., 2009; Tong et al., 2016). In *Helianthus*, low density linkage maps of *Helianthus annuus* and *H. petiolaris* suggested that three inversions and as many as eight translocations differentiated the species (Rieseberg et al., 1995; Burke et al., 2004). However, a recent follow-up study (Ostevik et al., 2019), which employed higher density genetic maps and a novel algorithm for synteny block detection, found 50-60 inversions between the species and 6-8 translocations. Thus, low-density maps appear better able to detect translocations than inversions, presumably because detection of the former requires fewer markers and is less sensitive to marker ordering errors. Ostevik et al. (2019) also applied their algorithm to comparisons of new maps for the two subspecies of *H. petiolaris*, as well as high-density genetic maps previously published for three other species (Barb et al., 2014). Up to 74 inversions and 15 translocations were found across the five taxa. Lastly, Huang et al. (2019) developed genetic maps for dune and non-dune ecotypes of *H. petiolaris* and successfully identified multiple inversions, but no translocations.

Comparative Genomic Studies

In recent years, the ever-increasing number of high quality genome assemblies and other genomic datasets have greatly facilitated the detection of chromosomal rearrangements and uncovered very large numbers of inversions between closely related species (**Table 1**). For example, a *de novo* assembly of *Arabidopsis thaliana* Ler-0 strain revealed 47 inversions between its genome and that of the widely used Col-0 accession (Zapata et al., 2016), although some unknown fraction of these inversions might have been introduced by mutagenesis. A reference genome of *Arabidopsis lyrata* was compared to that of *A. thaliana,* and 154 inversions were identified, as well as two reciprocal translocations and three chromosomal fusions previously revealed by genetic mapping (Yogeeswaran et al., 2005; Hu et al., 2011). Several inversions were also found between cucumber and melon (Huang et al., 2009; Garcia-Mas et al., 2012), and five paracentric and one pericentric inversions were revealed between cultivated and wild cucumber with the aid of comparative fluorescence *in situ* hybridization (Yang et al., 2012). Whole-genome sequencing of pepper confirmed previously reported large inter-chromosomal translocations and identified 367 inversions between pepper and potato (Qin et al., 2014), about 20x more than were identified via comparative mapping. In addition, a total of 214 inversions were identified between rice (*Oryza sativa*) and its close relative *O. brachyantha* (Chen et al., 2013). And a comprehensive study using homologous gene sequences showed that short paracentric inversions and short intra-chromosomal translocations were the most common rearrangements in the grass family Poaceae (Dvorak et al., 2018). Other well characterized examples come from comparisons of different cultivars of cotton (Yang et al., 2019) and grapevine (Zhou et al., 2019); details in **Table 1**.

Using the number of inversions reported in these studies and species' divergence times obtained from the literature, we estimated the rate of inversion evolution to be about 15–30 inversions per million years (**Table 1**). However, this estimate should be treated with caution since the quality of genome

TABLE 1 | Summary of inversions from comparative genomics studies.

Species	Common name	Number of inversions	Size	Number of translocations or fusions	Divergence time (MYA)	Evolutionary rate (inversion/MYA)	References
Arabidopsis thaliana strains	Thale cress	47	115 bp-1.17 Mbp	0	–	–	Zapata et al., 2016
Arabidopsis lyrata/A. thaliana	Thale cress/lyrate rockcress	154	–	5	10	15.4	Hu et al., 2011
Cucumis sativus	Wild and cultivated cucumbers	5 paracentric, 1 pericentric	–	0	–	–	Yang et al., 2012
C. sativus/Cucumis melo	Cucumber/melon	Several	–	–	10	–	Huang et al., 2009; Garcia-Mas et al., 2012
Capsicum annuum/Solanum lycopersicum	Pepper/tomato	468	–	612	20	23.4	Qin et al., 2014
Capsicum annuum/ Solanum tuberosum	Pepper/potato	367	–	430	20	18.35	
Oryza sativa/O. brachyantha	Rice	214	–	0	15	14.27	Chen et al., 2013
Gossypium hirsutum	Upland cotton	60	–	1314	–	–	Yang et al., 2019
Gossypium hirsutum/G. arboreum	Cottons	39	–	35	1-2	18.5-39	
Gossypium hirsutum/G. raimondii	Cottons	15	–	29	1-2	7.5-15	
Vitis vinifera cultivars	Grapevine	1513	–	3786	–	–	Zhou et al., 2019
Aegilops tauschii	Tausch's goatgrass	44	1.6–8.0 Mbp	–	3	14.67	Dvorak et al., 2018
Triticum turgidum subgenome A	Wild emmer wheat	91			3	30.33	
Triticum turgidum subgenome B		65			3	21.67	
Brachypodium distachyon	Purple false brome	82			35	2.34	
Oryza sativa	Rice	20			47	0.43	
Sorghum bicolor	Sorghum	33			53	0.62	

assemblies and the methods employed to identify chromosome rearrangements varied among studies. Also inversion sizes typically were not reported in these studies, except for Zapata et al. (2016) and Dvorak et al. (2018), who showed that most inversions are small in size. Some of the variation in rates of inversion evolution reported for different groups likely derives from different size cut-offs used to report inversions. For example, the range of inversion sizes reported by Zapata et al. (2016) and Dvorak et al. (2018) are almost completely non-overlapping (**Table 1**). In the future, it would be useful if studies would report the size distributions of inversions, as well as the extent of sequence divergence between inversion haplotypes.

Population Genomic Approaches

While these comparative approaches offer a means for determining the number of inversions between species, they typically tell us little about the distribution of inversion polymorphisms within species or the traits that are associated with the inversions. However, this information is needed to understand how inversions are established, as well as their role in adaptive evolution and speciation. Fortunately, two population genomic approaches have recently been developed that permit inference of inversions from resequencing data when paired with high quality reference sequences.

One approach detects potential inversions by scanning the genome for regions of high linkage disequilibrium (LD) among linked markers (Kemppainen et al., 2015). The rationale for this approach is that recombination suppression should produce very high LD among SNPs within inversions. Although gene conversion and double recombination can break down LD in the middle of old and large inversions, such as that observed for some inversions in *Drosophila* (Korunes and Noor, 2019), whether this is common in plants is unclear. Other mechanisms that reduce recombination, such as pericentromeric heterochromatin, will also lead to high LD regions. Thus, the LD scan should be complemented by an analysis of genotypic relationships within the predicted inversion using principle component analysis (PCA) or a similar method. Because inversions only suppress recombination in heterozygotes, three distinct genotypic clusters should be detected within an inversion representing each inversion orientation (0/0, 1/1), plus heterozygotes between inversion haplotypes (0/1). While LD scans have been employed to search for inversions in animals (Faria et al., 2019), we are unaware of their application to plants.

A second approach takes advantage of the effects of inversions on population structure. This approach assumes that the lack of gene flow between inversion haplotypes will result in differences in patterns of genetic relatedness between inverted and collinear regions. These outlier regions can be detected by conducting windowed analyses of population structure across the genome, such as that implemented by the Local PCA/population structure (lostruct) program developed by Li and Ralph (2019). As with the LD method, analyses of genotypic relationships within the predicted inversion can offer further support for the putative inversion. The local population structure approach has been used to detect polymorphic inversions from RAD sequencing data or whole-genome shotgun data within the wild sunflower species, *Helianthus annuus*, *H. argophyllus,* and *H. petiolaris* (Huang et al., 2019; Todesco et al., 2019). While many of the inversions predicted by this method in sunflower have been subsequently confirmed via comparisons of reference sequences, comparative mapping, or Hi-C sequencing, two were not, indicating that these population genomic approaches can offer suggestive evidence of inversions, but are less diagnostic. Other mechanisms, such as recent introgression, could generate patterns similar to those of an inversion. Conversely, small inversions, or inversions lacking elevated population structure or high LD outside of inversion breakpoints, might not be detected by these methods.

Using these population genomic approaches, the genotypes of multiple individuals can be simultaneously determined for all detected inversions, which provide useful information on their frequency and geographic distribution. Mapping the breakpoints of inversions will also be helpful for developing PCR markers to determine patterns of inversion polymorphism across large numbers of individuals. This sets the stage for associating traits and environmental factors with inversion haplotypes, thereby revealing the evolutionary forces that shape the pattern of inversion polymorphisms. Although additional independent lines of evidence are encouraged to confirm putative inversions suggested by such methods, population genomic approaches, coupled with ever-expanding population sequencing data, have great potential to further our understanding of the prevalence and evolutionary role of inversions in plants, especially in non-model species.

Different Likelihood of the Establishment of Inversions and Translocations

As discussed above, comparative mapping of plant species typically identified more inversions than inter-chromosomal translocations or "fusions" (terminal reciprocal translocations). In general, as mapping or sequencing resolution increases, so does the number of inversions detected (Wu and Tanksley, 2010; Hu et al., 2011; Chen et al., 2013; Qin et al., 2014; Ostevik et al., 2019). In contrast, little or no increase in the numbers of inter-chromosomal translocations is reported with increasing resolution. Some studies (not discussed here) have focused exclusively on chromosomal-scale translocations and fusions/fissions or did not clearly differentiate rearrangement types and thus are not relevant to this question.

We suspect that variation in the abundance of major inter-chromosomal translocations versus inversions relates more to differences in the likelihood of their establishment than to variation in mutation rates. Translocation heterozygotes involving different chromosomes will show mis-segregation during meiosis and produce unbalanced and inviable gametes (King, 1993). This strong heterozygous disadvantage (underdominance) of inter-chromosomal translocations makes them difficult to establish. On the other hand, plants seem to be more tolerant of intra-chromosomal rearrangements such as inversions. While recombination between inversion orientations is predicted to result in inviable gametes, the evidence for this is surprisingly sparse and comes mainly from interspecific crosses. Meiotic abnormalities diagnostic of inversions, along with reduced pollen viability, have been reported, for example,

in hybrids of *Gibasis venustula* and *G. speciose* (Kenton, 1981), *Vigna umbellate* and *V. minima* (Gopinathan and Babu, 1986), as well as between races of *Paspalum notatum* (Stein et al., 2004), but the fertility loss is typically much smaller than for most translocations. Surprisingly, inversions segregating within species often have no visible effect on fertility, such as reported for *Brassica oleracea* (Kianian and Quiros, 1992) and maize (Fang et al., 2012). In *Mimulus* and *Helianthus*, crosses between ecotypes that are separated only by inversions do not show reduced pollen viability (Lowry and Willis, 2010; Ostevik et al., 2016; Huang et al., 2019), although meiotic abnormalities diagnostic for inversions have been reported for interspecific crosses (Chandler et al., 1986). This suggests that the reduction in recombination associated with inversions within plant species is typically achieved by disrupting pairing and crossing over between inverted regions (Searle, 1993) rather than selection against inviable recombinant gametes. Regardless of the cause, the minimal underdominance of many inversions should ease their establishment.

In a number of comparative genomic studies, more translocations were reported than inversions (**Table 1**). Variation in the abundance of inversions and translocations seen in **Table 1** stems partly from differences in methods, criteria (e.g., size cut-offs), and power for detecting structural variants, as opposed to real differences in their frequency. Some studies (Yang et al., 2019; Zhou et al., 2019) applied whole-genome alignment, long-read alignment and short-read alignment to detect both inter- and intra-chromosomal translocations of various sizes (transposed genomic segments), while others (Hu et al., 2011) have focused exclusively on large inter-chromosomal reciprocal translocations. More robust conclusions about the prevalence of inversions and translocations will not only require that studies be more parallel in terms of data and methodology employed, but also that they take rearrangement size into account.

ORIGIN AND ESTABLISHMENT OF INVERSIONS

There are a number of different molecular mechanisms by which inversions can arise, including ectopic recombination between copies of repeated sequences such as transposable elements, tRNA genes or segmental duplications, or by chromosomal breakage and repair by non-homologous end-joining (Gray, 2000; Feschotte and Pritham, 2007; Delprat et al., 2009). Both mechanisms have been shown to occur in plants, especially in maize (Lister et al., 1993; Ziolkowski et al., 2003; Zhang and Peterson, 2004; Yu et al., 2011; Knoll et al., 2014). Epigenetic modification, given its role in transposable element de-activation and heterochromatin formation, may also play an important role in chromosome evolution in plants (Li et al., 2017). Given the high fraction of plant genomes occupied by transposable elements and other duplicated sequences, inversion mutation rates are likely to be high. However, the relative importance of these different mechanisms and the overall incidence of inversions in natural populations remain to be explored.

Like other genetic mutations, inversions can change in frequency as a consequence of genetic drift or selection. Early models of chromosomal evolution assumed that most underdominant rearrangements became established through drift (White, 1973; Lande, 1979). However, the fixation of a strongly underdominant mutation through drift is unlikely, except under extreme conditions, such as can be found in small founder populations and/or through inbreeding. The conditions for fixation are relaxed for neutral or weakly underdominant mutations such as inversions. Nonetheless, the fact that inversions are frequent in outcrossing species with large effective population sizes, co-vary with ecological variation, and underlie important adaptive traits (Hoffmann and Rieseberg, 2008; Lowry and Willis, 2010; Todesco et al., 2019), suggests that the establishment and spread of large inversions is most likely driven by natural selection. The jury is still out for the many small inversions that differentiate plant genomes.

Meiotic drive has been proposed as another possible mechanism for the fixation of chromosomal rearrangements. While meiotic drive may very well explain the establishment of large inter-chromosomal translocations, it appears to be too infrequent to account for the abundance of inversions seen both within and between species (Coyne, 1989). It also has been hypothesized that inversions could be favored if breakpoints disrupt an open reading frame or alter gene expression (Hoffmann and Rieseberg, 2008). While we are unaware of a case in plants where such changes have been shown to be adaptive, it is important to keep in mind that few breakpoints have been characterized for inversions with clear phenotypic effects. In the best-studied examples in plants (or animals), selection favoring the establishment of inversions appears to arise indirectly from their impact on reducing recombination within the inverted region. Thus, most recent evolutionary models for the establishment of inversions have focused on this property (Kirkpatrick and Barton, 2006; Burger and Akerman, 2011; Feder et al., 2011; Charlesworth and Barton, 2018).

The importance of recombination to the establishment and spread of inversions was initially put forward by Dobzhansky (1970) based on studies of *Drosophila*, in which inversions typically have little impact on fertility. Dobzhansky argued that genes within inversions were co-adapted, meaning that the fitness of the alleles held together by the inversion would be greater than the sum of their independent effects. A newly arisen inversion carrying a co-adapted set of alleles would spread to fixation in a population unless it was under balancing selection or there was ongoing migration from other populations. Unfortunately, we do not know whether such co-adaptation (i.e., favorable epistatic interactions) is common within inversions.

Newer models suggest that such epistatic interactions are not required if the inversions bring together two or more alleles that are adapted to the same local environment and there is ongoing migration between environments (Kirkpatrick and Barton, 2006; **Figures 2A,B**). In this situation, the newly derived inversion will have a selective advantage over the ancestral collinear arrangement that carries mixtures of adapted and maladapted alleles (Kirkpatrick and Barton, 2006; Burger and Akerman, 2011). However, a recent re-examination of the model showed

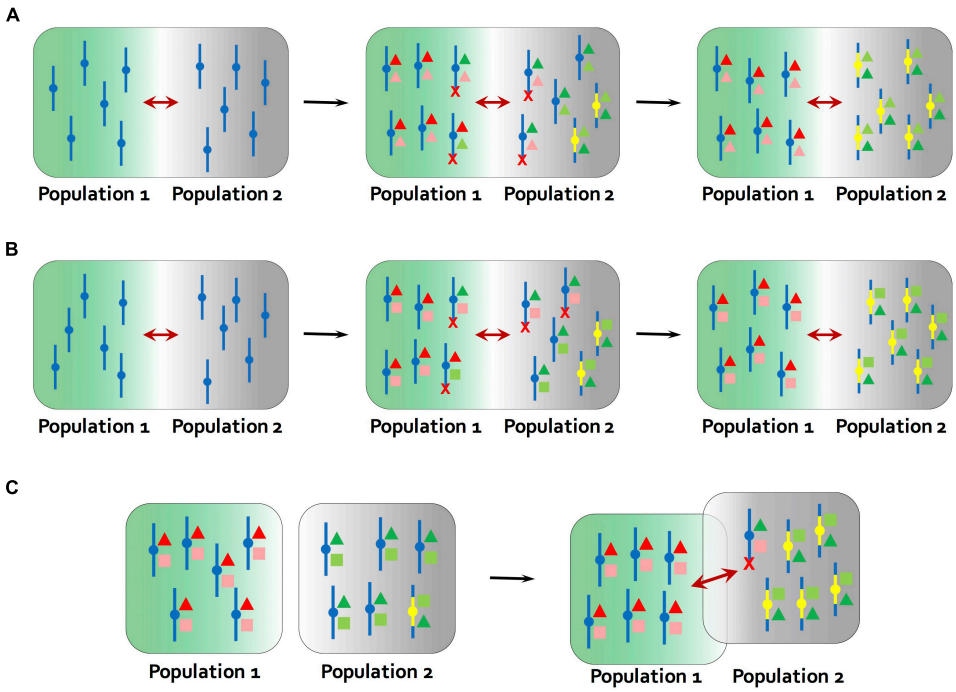

FIGURE 2 | Models for the establishment of inversions. **(A)** Kirkpatrick and Barton (2006) model. At the starting point, population 1 and 2 occur in different environments, but are connected by gene flow (maroon arrows). Different alleles (red and green colors) at multiple genes underlying the same locally adapted trait (deep color and light color triangles) are favored in local environments (green and gray backgrounds). The ancestral chromosome carries mixtures of adapted and maladapted alleles in the face of gene flow, while a new inversion carries only the locally adapted alleles (yellow bars). The inversion is therefore favored and rises to high frequency in population 2. **(B)** Inversions become established through a process similar to **(A)** but by carrying a combination of alleles at two loci that are adapted to different aspects of the local environment (triangles and squares in different colors). For example, in a dune ecotype of the prairie sunflower (*Helianthus petiolaris*), larger seed size and tolerance to low nutrient soils were found to map to the same inversions (Huang et al., 2019; Todesco et al., 2019). **(C)** Mixed geographic model proposed by Feder et al. (2011). At the starting point, population 1 and 2 are allopatric. Multiple locally adapted alleles (triangles and squares in different colors) are fixed due to lack of gene flow, and an inversion carrying a full complement of these alleles is present at low frequency in population 2 through mutation-purifying selection balance or genetic drift. At secondary contact, the reduction in recombination caused by the inversion results in a selective advantage over collinear regions, leading to rise of inversion frequency. Red crosses indicate that chromosomes carrying maladaptive combinations of alleles are eliminated in each environment.

that the selective advantage of an inversion will be small if the loci contained within the inversion are already tightly linked (Charlesworth and Barton, 2018). Thus, the conditions under which an inversion is favored in this model are less permissive that previously thought.

Empirical studies that associate multiple locally adapted traits or genes with inversions offer indirect support for this model. For example, Lowry and Willis (2010) showed that the chromosomal inversion differentiating annual and perennial ecotypes in *Mimulus guttatus* was associated with flowering time and morphological traits, as well as fitness in inland and coastal environments. Follow-up studies indicated that the inversion was associated with life history divergence and environmental variation, as well as adaptive trade-offs among growth, reproduction, and herbivore resistance (Oneal et al., 2014; Lowry et al., 2019). Similarly, in wild *Zea mays* an inversion on chromosome 1 showed a strong altitudinal cline in population frequency and statistical association with phenotypic traits such as culm diameter (Fang et al., 2012).

However, it is often unclear whether the inversions have captured pre-existing combinations of locally adapted alleles or whether such allelic combinations accumulated after inversion establishment. A number of studies in plants have successfully addressed this question, thereby offering more direct support for the Kirkpatrick and Barton (2006) model. Lee et al. (2017) made use of available collinear local adapted genotypes in *Boechera stricta* for genetic mapping and showed that pre-existing locally adaptive alleles may be captured by young inversions and contribute to local adaptation and incipient speciation. Likewise, Coughlan and Willis (2019) showed that key life history QTLs mapping to an inversion differentiating annual and perennial *Mimulus guttatus* mapped to the same region in a population involving annual *M. guttatus* and a collinear perennial species, *M. tilingii*, thereby showing that loci contributing to local adaptation predate the inversion in this system as well. Inversions on chromosome NC6 of *Noccaea caerulescens* are found to group pre-existing metal homeostasis genes, which may explain their fixation and role in speciation (Mandakova et al., 2015).

Inversion establishment in the Kirkpatrick and Barton (2006) model is also constrained by migration rates. Gene flow between different environments must be sufficiently high to generate a selective advantage for the new inversion. But high gene flow will

lead to recombination between adapted and non-adapted alleles, reducing the likelihood that a new inversion would bring together a complete set of locally adapted alleles. A possible solution to this issue was suggested by Feder et al. (2011), who developed a mixed geographic model, in which adaptation to different environments occurs in allopatry, so that it is straightforward for a new inversion to capture a full cassette of adaptive alleles. Subsequent range expansion and secondary contact would give the new inversion a selective advantage over collinear regions (as in the Kirkpatrick and Barton model), leading to its establishment (**Figure 2C**). Given that range fluctuations are common in plants, and that this model permits inversion establishment from standing variation, we suspect that it might be a common mechanism.

Evidence that secondary contact promotes the spread of inversions has been found in birds (Hooper and Price, 2017), but to date there has been little relevant data in plants. However, new data from *Helianthus* sunflowers implies that secondary contact and hybridization may contribute importantly to the establishment of large inversions. We used a combination of population genomic, comparative mapping, and HiC sequencing to detect numerous polymorphic inversions within *Helianthus annuus*, *H. argophyllus* and *H. petiolaris* (Huang et al., 2019; Todesco et al., 2019), which are sympatric and known to hybridize with multiple other species. However, when we applied a similar population genomic approach to the analyses of two *Helianthus s*pecies that are largely (*H. bolanderi*; data from Owens et al., 2016) or completely (*H. niveus*; data from Zhang et al., 2019) allopatric, we failed to find clear signals of inversions (**Figure 3**). A mixed geographic model might explain why inversions are only found in *Helianthus* species that have extensive range overlap with others taxa.

Secondary contact can also shape current pattern of inversions within species. Phylogenomic analyses of inversions segregating within *Helianthus* species revealed that these inversion haplotypes typically are highly divergent, pre-dating the split between species (Todesco et al., 2019). While such a pattern could be due to balancing selection, the lack of trans-specific inversion polymorphisms (i.e., none of the inversions are polymorphic in more than one species), suggests that they might have been acquired from other, possibly extinct, species instead. This could have occurred via introgression or species fusion. Note that the latter would also account for the "extinction" of donor species. Evidence for the origin of inversions through introgression is known from animals (Feder et al., 2003b; Tuttle et al., 2016), but evidence in plants is slim. Clearly, phylogenomic analyses of inversion origins and ages in other plant groups should be a priority for future studies.

ROLE IN SPECIATION

Early models of chromosomal speciation were based on the assumption that inversions and other chromosomal rearrangements reduced gene flow between taxa through their effects on hybrid fitness. However, due to the paucity of evidence of reduced fitness in hybrids heterozygous for inversions, as well as the theoretical difficulties associated with fixing strongly underdominant mutations, another class of models was developed based on the effects of inversions on recombination rates within the inverted region (Trickett and Butlin, 1994; Noor et al., 2001; Rieseberg, 2001). These recombination suppression models offer a means for resolving the widely recognized antagonism between divergent natural selection and recombination, permitting adaptive divergence and speciation in the presence of gene flow (Ortiz-Barrientos et al., 2016).

In this simplest model (Trickett and Butlin, 1994), an inversion reduces recombination between loci contributing to local adaptation and those causing assortative mating, permitting adaptive divergence and potentially speciation in the presence of gene flow (**Figure 4**). Such a genetic architecture appears to be common in plants. For example, the chromosomal inversion that contributes to local adaptation in inland and coastal environments in *Mimulus guttatus* is also associated with flowering time and other life history differences (Lowry and Willis, 2010; Oneal et al., 2014), thereby contributing to assortative mating between the annual and perennial ecotypes. In *M. lewisii* and *M. cardinalis*, inversions were also found to co-localize with a series of floral trait QTLs, such as corolla length and flower color, which are important in both prezygotic and postzygotic isolation in this species pair (Fishman et al., 2013). Similarly, in wild *Arabidopsis thaliana*, an inversion on chromosome 4 was reported to be strongly associated with fecundity under drought and an early flowering allele (Fransz et al., 2016). In *Boechera stricta*, Lee et al. (2017) found multiple linked phenology QTLs, including flowering differences, within an inversion that differentiates subspecies. In sunflower, inversions were associated with multiple ecological relevant traits, such as seed size and various soil and climate characteristics, as well as flowering time, revealing their role in ecotype formation and ecological speciation (Huang et al., 2019; Todesco et al., 2019).

While all of these examples demonstrate recombination suppression between locally adapted alleles and an assortative mating trait (most frequently flowering time), it is important to keep in mind that plant ecotypes and species often exhibit eco-geographic isolation. For example, in dune versus non-dune ecotypes of sunflower, the strongest reproductive barriers are immigrant inviability and extrinsic selection against hybrids (Ostevik et al., 2016). Strong selection against small seeds on the dunes, combined with a negative trade-off between seed size and seed number, underlies both barriers. Thus, isolation in this system is mainly due to a classic locally adapted trait (seed size), which maps to three inversions (Todesco et al., 2019), rather than assortative mating traits such as flowering time variation or conspecific pollen precedence. This situation is not unique to sunflowers and implies that the Trickett and Butlin (1994) model should be applicable to any trait that causes reproductive isolation, not just those contributing to assortative mating.

In addition to the Trickett and Butlin model, inversions have also been proposed as a means for maintaining hybrid incompatibilities in the face of ongoing gene flow (Noor et al., 2001), facilitating the accumulation of additional hybrid

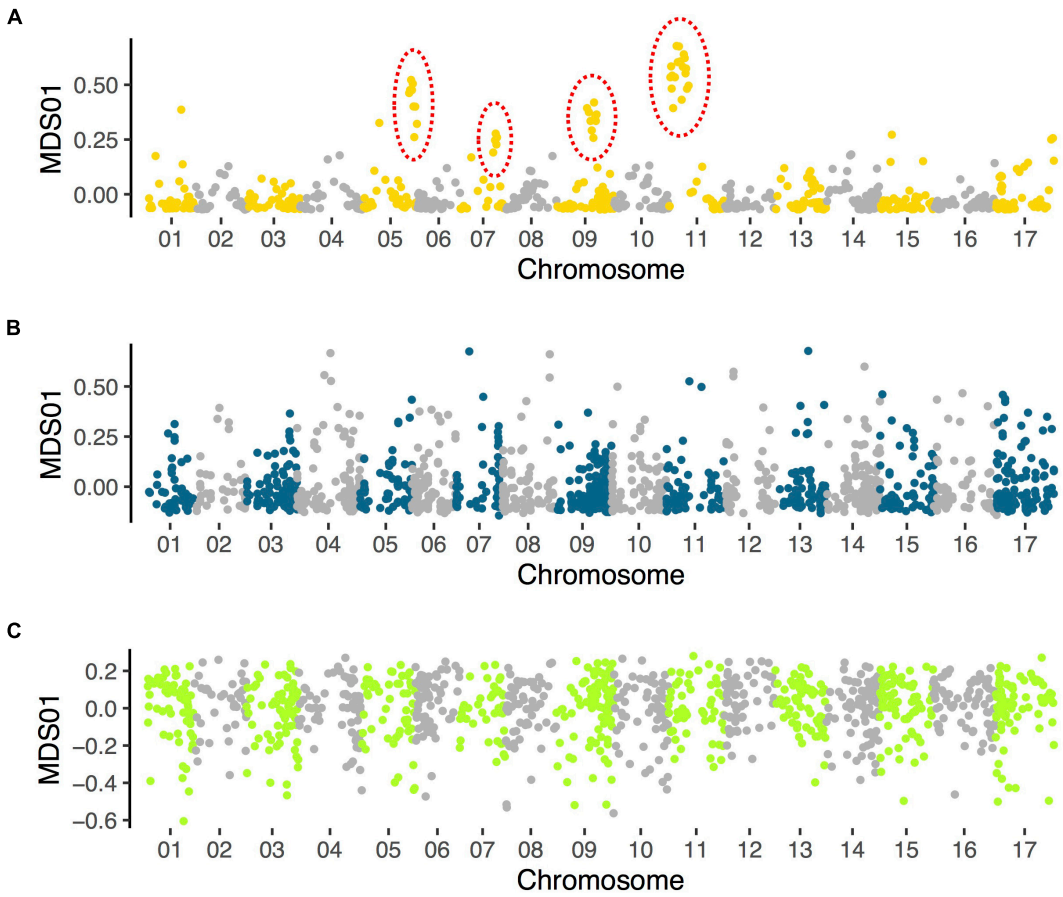

FIGURE 3 | Results of local population structure analyses in **(A)** *Helianthus petiolaris* (data from Huang et al., 2019), **(B)** *H. bolanderi* (data from Owens et al., 2016), and **(C)** *H. niveus* (data from Zhang et al., 2019). Variant calling and multidimensional scaling (MDS) follow the same methods described in Huang et al. (2019). Only the first MDS coordinate is plotted. Clusters of MDS outliers, which indicate putative inversions (and have been confirmed with other methods), are identified in *H. petiolaris* (indicated with dotted circles) but not in the other two species.

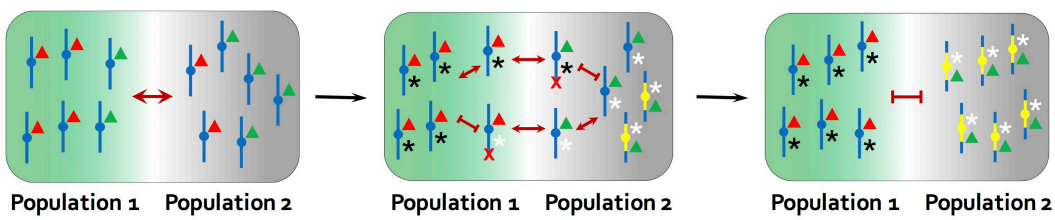

FIGURE 4 | The model for the role of inversions in speciation proposed by Trickett and Butlin (1994). An inversion facilitates speciation by suppressing recombination between genes involved in local adaptation (red and green triangles) and those underlying assortative mating traits, as such flowering time (black and white asterisks). The ancestral chromosome carries mixtures of adapted and maladapted alleles due to recombination. Individuals that are locally adapted to the environment of population 2, but carries the white assortative mating allele, will tend to mate with individuals adapted to the other environment and produce maladaptive offspring in population 2. Individuals with a new inversion that captures only the locally adapted alleles and black assortative mating allele do not suffer the reproductive cost from recombination. The inversion is therefore favored and contributes to further divergence between populations. Red crosses indicate that chromosomes carrying maladaptive combinations of alleles are eliminated in each environment.

incompatibilities (Navarro and Barton, 2003), and extending the time window for reinforcement to evolve (Servedio, 2000, 2009). Also, by suppressing recombination, inversions can extend the effects of genes that contribute in some way to isolation over larger genomic regions (Rieseberg, 2001), thereby generating "genomic islands of divergence" (Oneal et al., 2014; Twyford and Friedman, 2015; Huang et al., 2019). Lastly, by combining the effects of multiple locally adapted alleles, the selective advantage of an inversion is expected to be greater than that of individual alleles, permitting divergence under higher migration

rates (Rieseberg, 2001; Kirkpatrick and Barton, 2006). However, of these potential roles, only the association of inversions with genomic islands of divergence has been documented in plants.

SEX CHROMOSOME AND SEQUENTIAL INVERSIONS

Although most plants are hermaphrodite (co-sexual), some plant species have evolved separate male and female sex morphs (i.e., dioecy). The transition from a co-sexual breeding system to dioecy typically involves the formation of sex chromosomes, in which recombination suppression evolves between male and female sterility loci (Charlesworth, 2012). Inversions offer a straightforward means for suppressing recombination between newly formed X and Y (or Z and W) chromosomes and have been reported in numerous animal systems, as well as in a handful of plant species. For example, two large inversions were found to define the non-recombining region between Y and X chromosomes in papaya, suggestive of a role in sex chromosome formation (Wang et al., 2012).

The evolution of sexually antagonistic genes, which are favored in one sex but not in the other, provides additional selection pressure to reduce recombination between sex chromosome pairs; otherwise antagonistic alleles will be transmitted to the opposite sex. As a consequence, over time recombination suppression typically expands to cover most of the sex chromosome pair. Interestingly, in many animals and some plant species, such expansions appear to be episodic, producing "evolutionary strata" across sex chromosomes, i.e. spatial clusters of X-Y or Z-W orthologs with similar divergence estimates (Wright et al., 2016). Such strata are often closely associated with inversions, leading to suggestions that the stepwise establishment of inversions might be responsible for this pattern of divergence. Currently, the best evidence for this hypothesis in plants comes from papaya, in which two distinct evolutionary strata were discovered that correspond perfectly with the boundaries of the two inversions (Wang et al., 2012). In *Silene*, both strata (Nicolas et al., 2004; Bergero et al., 2007) and inversions (Zluvova et al., 2005; Hobza et al., 2007) have been reported, but they are not explicitly linked. Other authors have noted that recombination suppression associated with sex chromosome divergence sometimes occurs through other mechanisms such as transposable element insertion (e.g., Xu et al., 2019). Therefore, it is possible that suppressed recombination comes first, followed by the accumulation of inversions. An example in which recombination suppression precedes chromosomal rearrangements has been reported in fungi (Sun et al., 2017), but as far as we are aware, evidence for such a scenario is lacking in plants.

The stepwise accumulation of inversions need not be restricted to sex chromosomes. An inversion on an autosome could initially become established by capturing multiple locally adapted alleles, as proposed in the Kirkpatrick and Barton (2006) model. Subsequent inversions that added new locally adapted alleles into the non-recombining block would be favored by selection. This stepwise extension of recombination suppression presumably would create evolutionary strata similar to that seen on sex chromosomes. The apparent clustering of inversions seen in comparisons between cucumber and melon (Garcia-Mas et al., 2012), as well as between domesticated rice and *Oryza brachyantha* (Chen et al., 2013), are consistent with this hypothesis. Future dissection of the structure and divergence patterns of complex inversions should be a priority.

CONCLUDING REMARKS

Since the discovery of inversions in *Drosophila* close to a century ago, numerous verbal and quantitative models have explored their potential role in evolution and possible mechanisms for their establishment. Comparative genetic mapping and genomic studies have revealed that chromosomal inversions are far more prevalent than previously imagined. However, these studies often focus on large inversions only and/or fail to report on inversion sizes or the extent of sequence divergence between inversion haplotypes. These information gaps should be addressed in future comparisons of reference assemblies.

Likewise, an ever-increasing number of studies in plants suggest that inversions play a key role in adaptive divergence and speciation in the presence of gene flow. However, the genes and mutations underlying key traits associated with the inversions are difficult to identify because of strong linkage disequilibrium within the inverted region. Analyses of collinear genomes that are expected to differ for many of the same genes will aid in this process, as shown by two examples here. Population genetic and molecular tools would also help pinpoint the genetic changes within inversions that are responsible for adaptive differences or speciation. Of particular interest are regions near inversion breakpoints, since the inversions have the potential not only to disrupt open-reading frames or associations with regulatory elements, but also to change the local chromosome landscape of genes (e.g. potentially moving genes closer or further away from heterochromatic regions). Lastly, little is known in plants about potential downsides of inversions such as increased transposable element activity and the accumulation of deleterious mutations, both of which are a predicted consequence of suppressed recombination. This information gap that could also be addressed with population genomic analyses.

Despite the rapid accumulation of examples of the importance of inversions in a variety of ecological and evolutionary processes, information on their origin is scarce. Inversions can become established in several ways, but models based on the advantages of reducing recombination between locally adapted alleles when there is migration between different environments seem most plausible. This process is likely aided in some instances by a period of allopatry between hybridizing populations so that the full set of locally adapted alleles can be captured by the new inversion. Phylogenomic analyses of closely related species are needed to determine the origins and extent of divergence between inversion haplotypes, since evidence suggests that hybridization

and introgression may contribute to inversion establishment and subsequent evolutionary dynamics. Lastly, we urge students of chromosomal evolution to assess whether inversions are clustered in the genome and if evolutionary strata can be discovered on autosomes, similar to what has been reported for sex chromosomes.

AUTHOR CONTRIBUTIONS

LR conceived the idea. KH performed literature review, analyzed the data, and made the graphics. KH and LR wrote the manuscript.

FUNDING

This work was supported by China Scholarship Council (no. 201506380099) to KH and NSERC Discovery grant (327475) to LR

ACKNOWLEDGMENTS

The authors would like to thank Armando Geraldes for valuable comments on the manuscript, as well as Judith Mank, Greg Owens, Kate Ostevik, and Marco Todesco for general discussions about the role of inversions in evolution and how to study them.

REFERENCES

Ahmad, Q. N., Britten, E. J., and Byth, D. E. (1979). Inversion heterozygosity in the hybrid soybean × *Glycine soja*: evidence from a pachytene loop configuration and other meiotic irregularities. *J. Heredity* 70, 358–364. doi: 10.1093/oxfordjournals.jhered.a109277

Ahn, S., and Tanksley, S. D. (1993). Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci. U.S.A.* 90, 7980–7984. doi: 10.1073/pnas.90.17.7980

Anderson, L. K., Covey, P. A., Larsen, L. R., Bedinger, P., and Stack, S. M. (2010). Structural differences in chromosomes distinguish species in the tomato clade. *Cytogenet. Genome Res.* 129, 24–34. doi: 10.1159/000313850

Ayala, D., Fontaine, M. C., Cohuet, A., Fontenille, D., Vitalis, R., and Simard, F. (2010). Chromosomal inversions, natural selection and adaptation in the malaria vector *Anopheles funestus*. *Mol. Biol. Evol.* 28, 745–758. doi: 10.1093/molbev/msq248

Ayala, F. J., and Coluzzi, M. (2005). Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proc. Natl. Acad. Sci. U.S.A.* 102(Suppl. 1), 6535–6542. doi: 10.1073/pnas.0501847102

Barb, J. G., Bowers, J. E., Renaut, S., Rey, J. I., Knapp, S. J., Rieseberg, L. H., et al. (2014). Chromosomal evolution and patterns of introgression in *Helianthus*. *Genetics* 197, 969–979. doi: 10.1534/genetics.114.165548

Bergero, R., Forrest, A., Kamau, E., and Charlesworth, D. (2007). Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* 175, 1945–1954. doi: 10.1534/genetics.106.070110

Berhan, A. M., Hulbert, S. H., Butler, L. G., and Bennetzen, J. L. (1993). Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. *Theor. Appl. Genet.* 86, 598–604. doi: 10.1007/BF00838715

Bonierbale, M. W., Plaisted, R. L., and Tanksley, S. D. (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120, 1095–1103.

Burger, R., and Akerman, A. (2011). The effects of linkage and gene flow on local adaptation: a two-locus continent–island model. *Theor. Popul. Biol.* 80, 272–288. doi: 10.1016/j.tpb.2011.07.002

Burke, J. M., Lai, Z., Salmaso, M., Nakazato, T., Tang, S., Heesacker, A., et al. (2004). Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* 167, 449–457. doi: 10.1534/genetics.167.1.449

Chandler, J. M., Jan, C. C., and Beard, B. H. (1986). Chromosomal differentiation among the annual *Helianthus* species. *Syst. Bot.* 11, 354–371.

Charlesworth, B., and Barton, N. H. (2018). The spread of an inversion with migration and selection. *Genetics* 208, 377–382. doi: 10.1534/genetics.117.300426

Charlesworth, D. (2012). Plant sex chromosome evolution. *J. Exp. Bot.* 64, 405–420. doi: 10.1093/jxb/ers322

Chen, J., Huang, Q., Gao, D., Wang, J., Lang, Y., Liu, T., et al. (2013). Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat. Commun.* 4:1595. doi: 10.1038/ncomms2596

Coughlan, J. M., and Willis, J. H. (2019). Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. *Mol. Ecol.* 28, 1343–1357. doi: 10.1111/mec.14804

Coyne, J. A. (1989). A test of the role of meiotic drive in chromosome evolution. *Genetics* 123, 241–243.

Delprat, A., Negre, B., Puig, M., and Ruiz, A. (2009). The transposon Galileo generates natural chromosomal inversions in *Drosophila* by ectopic recombination. *PLoS One* 4:e7883. doi: 10.1371/journal.pone.0007883

Dobzhansky, T. G. (1970). *Genetics of the Evolutionary Process*. New York, NY: Columbia University Press.

Drost, D. R., Novaes, E., Boaventura-Novaes, C., Benedict, C. I., Brown, R. S., Yin, T., et al. (2009). A microarray-based genotyping and genetic mapping approach for highly heterozygous outcrossing species enables localization of a large fraction of the unassembled *Populus trichocarpa* genome sequence. *Plant J.* 58, 1054–1067. doi: 10.1111/j.1365-313X.2009.03828.x

Dvorak, J., Wang, L., Zhu, T., Jorgensen, C. M., Deal, K. R., Dai, X., et al. (2018). Structural variation and rates of genome evolution in the grass family seen through comparison of sequences of genomes greatly differing in size. *Plant J.* 95, 487–503. doi: 10.1111/tpj.13964

Fang, Z., Pyhäjärvi, T., Weber, A. L., Dawe, R. K., Glaubitz, J. C., González, J. D. J. S., et al. (2012). Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* 191, 883–894. doi: 10.1534/genetics.112.138578

Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., et al. (2019). Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Mol. Ecol.* 28, 1375–1393. doi: 10.1111/mec.14972

Feder, J. L., Berlocher, S. H., Roethele, J. B., Dambroski, H., Smith, J. J., Perry, W. L., et al. (2003a). Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10314–10319. doi: 10.1073/pnas.1730757100

Feder, J. L., Roethele, J. B., Filchak, K., Niedbalski, J., and Romero-Severson, J. (2003b). Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* 163, 939–953.

Feder, J. L., Gejji, R., Powell, T. H., and Nosil, P. (2011). Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution* 65, 2157–2170. doi: 10.1111/j.1558-5646.2011.01321.x

Feschotte, C., and Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368. doi: 10.1146/annurev.genet.40.110405.090448

Fishman, L., Stathos, A., Beardsley, P. M., Williams, C. F., and Hill, J. P. (2013). Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). *Evolution* 67, 2547–2560. doi: 10.1111/evo.12154

Flagel, L. E., Blackman, B. K., Fishman, L., Monnahan, P. J., Sweigart, A., and Kelly, J. K. (2019). GOOGA: a platform to synthesize mapping experiments and identify genomic structural diversity. *PLoS Comput. Biol.* 15:e1006949. doi: 10.1371/journal.pcbi.1006949

Fransz, P., Linc, G., Lee, C. R., Aflitos, S. A., Lasky, J. R., Toomajian, C., et al. (2016). Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. *Plant J.* 88, 159–178. doi: 10.1111/tpj.13262

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V. M., et al. (2012). The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. U.S.A.* 109, 11872–11877. doi: 10.1073/pnas.1205415109

Gopinathan, M. C., and Babu, C. R. (1986). Meiotic studies of the F 1 hybrid between rice bean (*Vigna umbellata*) and its wild relative *V. minima*. *Genetica* 71, 115–117. doi: 10.1007/bf00058693

Gray, Y. H. (2000). It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet.* 16, 461–468. doi: 10.1016/s0168-9525(00)02104-1

Greilhuber, J., and Speta, F. (1976). C-banded karyotypes in the *Scilla hohenackeri* group, *S. persica*, and *Puschkinia* (*Liliaceae*). *Plant Syst. Evol.* 126, 149–188.

Hobza, R., Kejnovsky, E., Vyskot, B., and Widmer, A. (2007). The role of chromosomal rearrangements in the evolution of *Silene latifolia* sex chromosomes. *Mol. Genet. Genomics* 278, 633–638. doi: 10.1007/s00438-007-0279-0

Hoffmann, A. A., and Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol Syst.* 39, 21–42. doi: 10.1146/annurev.ecolsys.39.110707.173532

Hooper, D. M., and Price, T. D. (2017). Chromosomal inversion differences correlate with range overlap in passerine birds. *Nat. Ecol. Evol.* 1, 1526–1534. doi: 10.1038/s41559-017-0284-6

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

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., and Rieseberg, L. H. (2019). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *bioRxiv* [Preprint] Available online at: https://www.biorxiv.org/content/10.1101/829622v1 (accessed November 7, 2019).

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., et al. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* 41, 1275–1281. doi: 10.1038/ng.475

Kemppainen, P., Knight, C. G., Sarma, D. K., Hlaing, T., Prakash, A., Maung Maung, Y. N., et al. (2015). Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. *Mol. Ecol. Resour.* 15, 1031–1045. doi: 10.1111/1755-0998.12369

Kenton, A. (1981). Chromosome evolution in the *Gibasis linearis* alliance (Commelinaceae). *Chromosoma* 84, 291–304. doi: 10.1007/bf00399139

Kianian, S. F., and Quiros, C. F. (1992). Generation of a *Brassica oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. *Theor. Appl. Genet.* 84, 544–554. doi: 10.1007/BF00224150

King, M. (1993). *Species Evolution: the Role of Chromosome Change*. Cambridge: Cambridge University Press.

Kirkpatrick, M., and Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics* 173, 419–434. doi: 10.1534/genetics.105.047985

Knoll, A., Fauser, F., and Puchta, H. (2014). DNA recombination in somatic plant cells: mechanisms and evolutionary consequences. *Chromosome Res.* 22, 191–201. doi: 10.1007/s10577-014-9415-y

Konishi, T., and Linde-Laursen, I. (1988). Spontaneous chromosomal rearrangements in cultivated and wild barleys. *Theor. Appl. Genet.* 75, 237–243. doi: 10.1007/bf00303959

Korunes, K. L., and Noor, M. A. (2019). Pervasive gene conversion in chromosomal inversion heterozygotes. *Mol. Ecol.* 28, 1302–1315. doi: 10.1111/mec.14921

Krimbas, C. B., and Powell, J. R. (1992). *Drosophila Inversion Polymorphism*. Boca Raton, FL: CRC Press.

Lagercrantz, U. (1998). Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 150, 1217–1228.

Lande, R. (1979). Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. *Evolution* 33(1 Part 1), 234–251. doi: 10.1111/j.1558-5646.1979.tb04678.x

Lee, C. R., Wang, B., Mojica, J. P., Mandáková, T., Prasad, K. V., Goicoechea, J. L., et al. (2017). Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat. Ecol. Evol.* 1:119.

Lewis, H., and Roberts, M. R. (1956). The origin of *Clarkia lingulata*. *Evolution* 10, 126–138. doi: 10.1111/j.1558-5646.1956.tb02839.x

Li, H., and Ralph, P. (2019). Local PCA shows how the effect of population structure differs along the genome. *Genetics* 211, 289–304. doi: 10.1534/genetics.118.301747

Li, S. F., Su, T., Cheng, G. Q., Wang, B. X., Li, X., Deng, C. L., et al. (2017). Chromosome evolution in connection with repetitive sequences and epigenetics in plants. *Genes* 8:290. doi: 10.3390/genes8100290

Lister, C., Jackson, D., and Martin, C. (1993). Transposon-induced inversion in *Antirrhinum* modifies *nivea* gene expression to give a novel flower color pattern under the control of *cycloidearadialis*. *Plant Cell* 5, 1541–1553. doi: 10.1105/tpc.5.11.1541

Lowry, D. B., Popovic, D., Brennan, D. J., and Holeski, L. M. (2019). Mechanisms of a locally adaptive shift in allocation among growth, reproduction, and herbivore resistance in *Mimulus guttatus*. *Evolution* 73, 1168–1181. doi: 10.1111/evo.13699

Lowry, D. B., and Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* 8:e1000500. doi: 10.1371/journal.pbio.1000500

Lysak, M. A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K., and Schubert, I. (2006). Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5224–5229. doi: 10.1073/pnas.0510791103

Lysak, M. A., Cheung, K., Kitschke, M., and Bures, P. (2007). Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiol.* 145, 402–410. doi: 10.1104/pp.107.104380

Lysak, M. A., and Lexer, C. (2006). Towards the era of comparative evolutionary genomics in Brassicaceae. *Plant Syst. Evol.* 259, 175–198. doi: 10.1007/s00606-006-0418-9

Mandakova, T., and Lysak, M. A. (2008). Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). *Plant Cell* 20, 2559–2570. doi: 10.1105/tpc.108.062166

Mandakova, T., Singh, V., Kraemer, U., and Lysak, M. A. (2015). Genome structure of the heavy metal hyperaccumulator *Noccaea caerulescens* and its stability on metalliferous and nonmetalliferous soils. *Plant Physiol.* 169, 674–689. doi: 10.1104/pp.15.00619

Navarro, A., and Barton, N. H. (2003). Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* 57, 447–459. doi: 10.1111/j.0014-3820.2003.tb01537.x

Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., et al. (2004). A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol.* 3:e4. doi: 10.1371/journal.pbio.0030004

Noor, M. A., Grams, K. L., Bertucci, L. A., and Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12084–12088. doi: 10.1073/pnas.221274498

Oneal, E., Lowry, D. B., Wright, K. M., Zhu, Z., and Willis, J. H. (2014). Divergent population structure and climate associations of a chromosomal inversion polymorphism across the *Mimulus guttatus* species complex. *Mol. Ecol.* 23, 2844–2860. doi: 10.1111/mec.12778

Ortiz-Barrientos, D., Engelstädter, J., and Rieseberg, L. H. (2016). Recombination rate evolution and the origin of species. *Trends Ecol. Evol.* 31, 226–236. doi: 10.1016/j.tree.2015.12.016

Ortiz-Barrientos, D., Reiland, J., Hey, J., and Noor, M. A. (2002). Recombination and the divergence of hybridizing species. *Genetica* 116, 167–178. doi: 10.1007/978-94-010-0265-3_2

Ostevik, K. L., Andrew, R. L., Otto, S. P., and Rieseberg, L. H. (2016). Multiple reproductive barriers separate recently diverged sunflower ecotypes. *Evolution* 70, 2322–2335. doi: 10.1111/evo.13027

Ostevik, K. L., Samuk, K., and Rieseberg, L. H. (2019). Ancestral reconstruction of sunflower karyotypes reveals dramatic chromosomal evolution. *bioRxiv* [Preprint] Available online at: https://www.biorxiv.org/content/10.1101/737155v1 (accessed August 15, 2019).

Owens, G. L., Baute, G. J., and Rieseberg, L. H. (2016). Revisiting a classic case of introgression: hybridization and gene flow in Californian sunflowers. *Mol. Ecol.* 25, 2630–2643. doi: 10.1111/mec.13569

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum*

domestication and specialization. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5135–5140. doi: 10.1073/pnas.1400975111

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16, 351–358. doi: 10.1016/s0169-5347(01)02187-5

Rieseberg, L. H., Van Fossen, C., and Desrochers, A. M. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375, 313–316. doi: 10.1038/375313a0

Rodriguez, S., Perera, E., Maestra, B., Diez, M., and Naranjo, T. (2000). Chromosome structure of *Triticum timopheevii* relative to *T. turgidum*. *Genome* 43, 923–930. doi: 10.1139/gen-43-6-923

Searle, J. B. (1993). "Chromosomal hybrid zones in eutherian mammals," in *Hybrid Zones and the Evolutionary Process*, ed. R. G. Harrison (Oxford: Oxford University Press), 309–353.

Servedio, M. R. (2000). Reinforcement and the genetics of nonrandom mating. *Evolution* 54, 21–29. doi: 10.1111/j.0014-3820.2000.tb00003.x

Servedio, M. R. (2009). The role of linkage disequilibrium in the evolution of premating isolation. *Heredity* 102, 51–56. doi: 10.1038/hdy.2008.98

Stein, J., Quarin, C. L., Martínez, E. J., Pessino, S. C., and Ortiz, J. P. A. (2004). Tetraploid races of *Paspalum notatum* show polysomic inheritance and preferential chromosome pairing around the apospory-controlling locus. *Theor. Appl. Genet.* 109, 186–191. doi: 10.1007/s00122-004-1614-z

Sturtevant, A. H. (1921). A case of rearrangement of genes in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 7, 235–237. doi: 10.1073/pnas.7.8.235

Sun, Y., Svedberg, J., Hiltunen, M., Corcoran, P., and Johannesson, H. (2017). Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat. Commun.* 8:1140. doi: 10.1038/s41467-017-01317-6

Sybenga, J. (1975). *Meiotic Configurations*. Berlin: Springer-Verlag.

Szinay, D., Wijnker, E., van den Berg, R., Visser, R. G., de Jong, H., and Bai, Y. (2012). Chromosome evolution in *Solanum* traced by cross-species BAC-FISH. *New Phytol.* 195, 688–698. doi: 10.1111/j.1469-8137.2012.04195.x

Tanksley, S. D., Ganal, M. W., Prince, J. P., De Vicente, M. C., Bonierbale, M. W., Broun, P., et al. (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132, 1141–1160.

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J. S., Soudi, S., Burge, D. O., et al. (2019). Massive haplotypes underlie ecotypic differentiation in sunflowers. *bioRxiv* [Preprint] Available online at: https://www.biorxiv.org/content/10.1101/790279v1 (accessed October 2, 2019).

Tong, C., Li, H., Wang, Y., Li, X., Ou, J., Wang, D., et al. (2016). Construction of high-density linkage maps of *Populus deltoides* × *P. simonii* using restriction-site associated DNA sequencing. *PLoS One* 11:e0150692. doi: 10.1371/journal.pone.0150692

Trickett, A. J., and Butlin, R. K. (1994). Recombination suppressors and the evolution of new species. *Heredity* 73(Pt 4), 339–345. doi: 10.1038/hdy.1994.180

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., et al. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* 26, 344–350. doi: 10.1016/j.cub.2015.11.069

Twyford, A. D., and Friedman, J. (2015). Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution* 69, 1476–1486. doi: 10.1111/evo.12663

Wang, J., Na, J. K., Yu, Q., Gschwend, A. R., Han, J., Zeng, F., et al. (2012). Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13710–13715. doi: 10.1073/pnas.1207833109

Wellenreuther, M., and Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* 33, 427–440. doi: 10.1016/j.tree.2018.04.002

White, M. J. D. (1973). *Animal Cytology and Evolution*. Cambridge: Cambridge University Press.

Wright, A. E., Dean, R., Zimmer, F., and Mank, J. E. (2016). How to make a sex chromosome. *Nat. Commun.* 7:12087.

Wu, F., Eannetta, N. T., Xu, Y., Durrett, R., Mazourek, M., Jahn, M. M., et al. (2009). A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *Theor. Appl. Genet.* 118, 1279–1293. doi: 10.1007/s00122-009-0980-y

Wu, F., and Tanksley, S. D. (2010). Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* 11:182. doi: 10.1186/1471-2164-11-182

Xu, L., Auer, G., Peona, V., Suh, A., Deng, Y., Feng, S., et al. (2019). Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nat. Ecol. Evol.* 3, 834–844. doi: 10.1038/s41559-019-0850-1

Yang, L., Koo, D. H., Li, Y., Zhang, X., Luan, F., Havey, M. J., et al. (2012). Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* 71, 895–906. doi: 10.1111/j.1365-313X.2012.05017.x

Yang, Z., Ge, X., Yang, Z., Qin, W., Sun, G., Wang, Z., et al. (2019). Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* 10:2989. doi: 10.1038/s41467-019-10820-x

Yogeeswaran, K., Frary, A., York, T. L., Amenta, A., Lesser, A. H., Nasrallah, J. B., et al. (2005). Comparative genome analyses of *Arabidopsis* spp: inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Res.* 15, 505–515. doi: 10.1101/gr.3436305

Yu, C., Zhang, J., and Peterson, T. (2011). Genome rearrangements in maize induced by alternative transposition of reversed *Ac*/*Ds* termini. *Genetics* 188, 59–67. doi: 10.1534/genetics.111.126847

Zapata, L., Ding, J., Willing, E. M., Hartwig, B., Bezdan, D., Jiao, W. B., et al. (2016). Chromosome-level assembly of *Arabidopsis thaliana* Ler reveals the extent of translocation and inversion polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.* 113, E4052–E4060. doi: 10.1073/pnas.1607532113

Zhang, J., and Peterson, T. (2004). Transposition of reversed Ac element ends generates chromosome rearrangements in maize. *Genetics* 167, 1929–1937. doi: 10.1534/genetics.103.026229

Zhang, J. Q., Imerovski, I., Borkowski, K., Huang, K., Burge, D., and Rieseberg, L. H. (2019). Intraspecific genetic divergence within *Helianthus niveus* and the status of two new morphotypes from Mexico. *Am. J. Bot.* 106, 1229–1239. doi: 10.1002/ajb2.1349

Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., et al. (2019). The population genetics of structural variants in grapevine domestication. *Nat. Plants* 5, 965–979. doi: 10.1038/s41477-019-0507-8

Ziolkowski, P. A., Blanc, G., and Sadowski, J. (2003). Structural divergence of chromosomal segments that arose from successive duplication events in the *Arabidopsis* genome. *Nucleic Acids Res.* 31, 1339–1350. doi: 10.1093/nar/gkg201

Zluvova, J., Janousek, B., Negrutiu, I., and Vyskot, B. (2005). Comparison of the X and Y chromosome organization in *Silene latifolia*. *Genetics* 170, 1431–1434. doi: 10.1534/genetics.105.040444

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: 31 March 2020
doi: 10.3389/fpls.2020.00362

Differential Genome Size and Repetitive DNA Evolution in Diploid Species of *Melampodium* sect. *Melampodium* (Asteraceae)

*Jamie McCann[1], Jiří Macas[2], Petr Novák[2], Tod F. Stuessy[3], Jose L. Villaseñor[4] and Hanna Weiss-Schneeweiss[1]**

[1] Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria, [2] Biology Centre, Czech Academy of Sciences, Institute of Plant Molecular Biology, České Budějovice, Czechia, [3] Herbarium and Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH, United States, [4] Department of Botany, National Autonomous University of Mexico, Mexico City, Mexico

OPEN ACCESS

Edited by:
Steven Dodsworth,
University of Bedfordshire,
United Kingdom

Reviewed by:
Sònia Garcia,
Botanical Institute of Barcelona (IBB, CSIC-Ajuntament de Barcelona), Spain
Terezie Mandakova,
Masaryk University, Czechia

***Correspondence:**
Hanna Weiss-Schneeweiss
hanna.schneeweiss@univie.ac.at

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

Received: 31 January 2020
Accepted: 12 March 2020
Published: 31 March 2020

Citation:
McCann J, Macas J, Novák P, Stuessy TF, Villaseñor JL and Weiss-Schneeweiss H (2020) Differential Genome Size and Repetitive DNA Evolution in Diploid Species of Melampodium sect. Melampodium (Asteraceae). Front. Plant Sci. 11:362. doi: 10.3389/fpls.2020.00362

Plant genomes vary greatly in composition and size mainly due to the diversity of repetitive DNAs and the inherent propensity for their amplification and removal from the host genome. Most studies addressing repeatome dynamics focus on model organisms, whereas few provide comprehensive investigations across the genomes of related taxa. Herein, we analyze the evolution of repeats of the 13 species in *Melampodium* sect. *Melampodium*, representing all but two of its diploid taxa, in a phylogenetic context. The investigated genomes range in size from 0.49 to 2.27 pg/1C (ca. 4.5-fold variation), despite having the same base chromosome number (x = 10) and very strong phylogenetic affinities. Phylogenetic analysis performed in BEAST and ancestral genome size reconstruction revealed mixed patterns of genome size increases and decreases across the group. High-throughput genome skimming and the RepeatExplorer pipeline were utilized to determine the repeat families responsible for the differences in observed genome sizes. Patterns of repeat evolution were found to be highly correlated with phylogenetic position, namely taxonomic series circumscription. Major differences found were in the abundances of the SIRE (*Ty1-copia*), Athila (*Ty3-gypsy*), and *CACTA* (DNA transposon) lineages. Additionally, several satellite DNA families were found to be highly group-specific, although their overall contribution to genome size variation was relatively small. Evolutionary changes in repetitive DNA composition and genome size were complex, with independent patterns of genome up- and downsizing throughout the evolution of the analyzed diploids. A model-based analysis of genome size and repetitive DNA composition revealed evidence for strong phylogenetic signal and differential evolutionary rates of major lineages of repeats in the diploid genomes.

Keywords: ancestral state reconstruction, Bayesian analysis, genome size, *Melampodium*, phylogenetics, repetitive DNA, tandem repeats, transposable elements

INTRODUCTION

Nuclear genome size is a strikingly variable characteristic of the flowering plants. Current estimates show more than 2000-fold genome size variation from the smallest known genomes of the carnivorous *Genlisea* (Lentibulariaceae; ca. 0.06 pg/1C) to the largest of *Paris japonica* (Melanthiaceae, 152.2 pg/1C) (Pellicer et al., 2010; Fleischmann et al., 2014). Although polyploidization, frequently occurring in plants, results in instant multiplication of the whole nuclear genome, differential evolution of the repetitive component of the genome explains the majority of the observed genome size variation in angiosperms (Bennetzen, 2005). However, the complexity of the repetitive fraction of the genome, both in terms of content and evolution and how it relates to species diversity, is still relatively unknown.

Repetitive DNA in plant genomes consists of two broad categories of repeat types, the dispersed mobile elements and tandem repeats (Bennetzen and Wang, 2014). Dispersed elements encompass mostly the DNA transposons and retrotransposons, which are commonly referred to as cut-and-paste and copy-and-paste elements, respectively. The long-terminal-repeat (LTR) retrotransposons are the most frequently occurring elements in most plant genomes and are typically distributed throughout the chromosomes. These repeats encompass several superfamilies, of which *Ty1-copia* and *Ty3-gypsy* are the most common in plants (Wicker et al., 2007). Which of these two superfamilies predominates in the genome can differ between plant groups/families (Novák et al., 2014; Kelly et al., 2015; Macas et al., 2015).

Tandem repeats include the ribosomal DNAs (5S and 35S) and satellite DNAs. In contrast to transposable elements, tandem repeats are typically found in distinct loci in the chromosomes. The ribosomal loci are the only known coding tandem repeats in the genome and can be localized in multiple places in the chromosomes. While both rDNA loci are useful phylogenetic and cytogenetic markers, 35S rDNA in particular is commonly used as a model for investigating processes of homogenization and epigenetic gene regulation (Kovařík et al., 2008; Zozomová-Lihová et al., 2014). Satellite DNAs are long arrays of repeated sequence monomers found in centromeric, telomeric and interstitial chromosomal regions of the genome. These repeats can make up a substantial portion of the plant genome (more than 30% in some species; olive – Barghini et al., 2014; *Frittilaria* – Kelly et al., 2015) and evolve quickly as evidenced by their high sequence variability among species (Melters et al., 2013; Garrido-Ramos, 2015). The latter phenomenon is explained by the library hypothesis, where the species of a related group contain different amounts of a common set of satellite repeats that are differentially amplified and might evolve independently after speciation (Ugarković and Plohl, 2002; Ugarković, 2008).

The extent and rate of amplification and removal of various repeats in the genome is the key driver of genome size evolution. Seed plant genomes, in general, contain more repetitive DNA than their animal counterparts, which makes them excellent systems for studying the evolution of repetitive DNA. Few comparative studies of the repeat landscape from closely related groups of organisms are available as of yet, as this has only recently become feasible with the advent of high-throughput sequencing (HTS) and the RepeatExplorer pipeline (Novák et al., 2010, 2013). This pipeline uses genome skimming data and graph-based clustering algorithms to identify repeat families (clusters) in the genome and calculate their abundances. Sequence reads from multiple closely related species can also be analyzed together in a comparative analysis, whereby individual clusters contain sequence reads from multiple species, indicative of the presence of homologous repeat families in their genomes (Macas et al., 2015).

The utility of RepeatExplorer and the comparative approach has been demonstrated in several study systems thus far, focusing on different phylogenetic scales (Novák et al., 2014; Macas et al., 2015; Vu et al., 2015) and allopolyploid systems (Renny-Byfield et al., 2013; Dodsworth et al., 2016). These studies have shown, in addition to characterizing repeats in the focal genomes, that the repeat family abundances are phylogenetically informative. An analysis conducted using maximum parsimony and likelihood approaches to reconstructing phylogenies from continuous characters in several different plant groups showed that these phylogenies gave similar results to those from commonly used molecular markers, with similar node support (Dodsworth et al., 2015).

A good system for investigating repetitive DNA evolution is the plant genus *Melampodium* (Asteraceae). which comprises 40 species distributed in Mexico. The largest section in the group, section *Melampodium* (x = 10), comprises 22 species, with 13 diploid species and nine exclusively polyploid or having both diploid and polyploid cytotypes (Blöch et al., 2009). All species have had their karyotypes established including localization of 5S and 35S rDNA loci and genome sizes measured (Weiss-Schneeweiss et al., 2009, 2012). Although all diploid species have the same chromosome number, they exhibit approximately 4.5-fold variation in genome size (0.49 pg–2.27 pg/1C) and are classified into five phylogenetically distinct series (Stuessy et al., 2011).

Herein, we investigate repetitive DNA and genome size evolution in all but two of the diploid species (13 of 15) in *Melampodium* sect. *Melampodium*. Material for *M. sinuatum* was not available, and the isolated phylogenetic position of *Melampodium longipilum* (series *Longipila*) relative to the other species of section *Melampodium* inferred using nuclear markers (Blöch et al., 2009; Stuessy et al., 2011; McCann et al., 2018) precluded it from analyses. The main goal of this study is to elucidate patterns of genome size change along the phylogeny of the group through an investigation of repetitive DNA evolution. Therefore, we extend the aforementioned results to test hypotheses that (1) genome size evolution has included both up- and downsizing, (2) repetitive DNA composition of these diploid genomes is strongly correlated with the phylogeny of the section, and (3) repeat lineages have different rates of evolution. To this end, ancestral genome size of *Melampodium* sect. *Melampodium* is reconstructed using traditional model-based approaches, the composition of repeats in the analyzed diploid genomes is characterized using the RepeatExplorer pipeline, and a novel model-based approach in a Bayesian framework is applied to reconstruct a phylogeny using

comparative repeat abundances and to estimate rates of evolution for different repeat types.

MATERIALS AND METHODS

Sequence Acquisition and Phylogenetic Analysis

Sequences of the internal transcribed spacer (ITS1-5.8S-ITS2) of 35S rDNA, the non-transcribed spacer (NTS) of 5S rDNA, two paralogs of the *phosphoglucoisomerase* gene (*Pgi*CI and *Pgi*CII) and the chloroplast gene *maturase* K (*mat*K) from all available diploid species in *Melampodium* sect. *Melampodium* published in earlier works (Blöch et al., 2009; Stuessy et al., 2011; Weiss-Schneeweiss et al., 2012) were used (summarized in **Supplementary Table S1**). Alignments were performed in Muscle 3.8.31 (Edgar, 2004) and further refined manually in Geneious R6 (Kearse et al., 2012).

Phylogenetic inference of species trees using all markers was performed using the StarBEAST2 package implemented in BEAST version 2.4.4 (Bouckaert et al., 2014; Ogilvie and Drummond, 2016). The overall rate of molecular evolution for each marker was inferred separately using the species tree uncorrelated relaxed log-normal clock model (Ogilvie and Drummond, 2016). The prior on the mean of the log-normal relaxed clock was log-normal with mean 0.005 and standard deviation 0.35 (in log space). The standard deviation prior was exponential with mean equal to 1/3. A log-normal calibration with mean 5.5 million years and standard deviation 0.23 in real space (McCann et al., 2018) was used on the root of the tree.

Four independent MCMCs were run for 500 million generations, with burn-in of 10% and sampling every 100K generations. The log files were all checked for convergence (Effective Sample size [ESS] > 200 for all parameters) using Tracer 1.6 (Rambaut, 2007). A maximum clade credibility (MCC) tree was calculated from the combined set of trees (1125 from each run, 4500 total), with mean node heights and no limit on the posterior probability of each clade.

Ancestral Genome Size Reconstruction

Ancestral genome size was reconstructed using a set of twelve species, without *M. moctezumum* for which sequence data were not available. Reconstruction of ancestral genome size was performed on each tree in the post burn-in combined set from the BEAST runs (see above) in RevBayes, with the following three models: normal Brownian motion, Ornstein–Uhlenbeck, and a relaxed clock model. All three models shared a single rate parameter for genome size evolution. The latter two models, however, had an additional parameter allowing for either selection toward an optimum value or Brownian rate variation among the branches in the tree. A simple test of whether the estimate of these parameters was equal to zero was used to determine the validity of the model.

The data augmentation method was used for all models, and the script was adapted from the RevBayes tutorials (Horvilleur and Lartillot, 2014). This method allows reconstruction of the internal node states for each tree, with the need for prior specifications on the Brownian motion evolutionary rate and the root state of the tree. Wide uniform priors were placed on the logarithm of the Brownian motion rate (−5, 5) and the root genome size (−100, 100). These rate priors were used in all three analyses. For the Ornstein–Uhlenbeck model, the root genome size was used as the optimum value and the prior on the strength of selection was the same as that used for the Brownian motion rate. The mean of the log-normal for the relaxed clock analyses was set to 1 (as the Brownian motion rate was already estimated), and an exponential prior with mean 1/3 was used on the standard deviation.

For each tree and model a single MCMC run of 1 million generations with 100K burn-in was performed. Each run was checked for convergence (effective sample size > 200 of each model parameter) using custom python scripts. Posterior samples of ancestral genome sizes for all nodes in the MCC tree were extracted from the log files of the best model for each tree, provided the node was present (few nodes in the MCC had posterior support of 100%, see results), and combined into a single file for calculation of mean and 95% credible intervals using the coda R package (Plummer et al., 2006).

The program BayesTraits was used to estimate the δ, κ, and λ parameters, which are commonly used in ancestral state reconstructions to test for the presence of various evolutionary processes (Pagel and Meade, 2004). The δ parameter scales the tree in a way such that it can be detected if the rate of evolution of the trait in question changed as a function of distance from the root. The κ parameter also scales the branch lengths of the tree and lower values, for example, $\kappa = 0$, indicates that the branch lengths are not informative for the evolution of the trait. The λ parameter provides a measure of statistical independence of trait evolution and phylogeny, i.e., a value of 0 indicates that phylogenetic structure does not affect trait evolution. Four separate runs of BayesTraits were performed using the MCMC method of the program over the posterior set of trees. The different runs corresponded to independent estimates of each of the aforementioned parameters. The posterior distribution of these parameters was interpreted according to the BayesTraits manual (Pagel and Meade, 2004).

DNA Isolation and High-Throughput Sequencing

Genomic DNAs were isolated for all diploid species or cytotypes in *Melampodium* sect. *Melampodium* with the exception of *M. sinuatum* (excluded due to lack of material) and libraries were prepared as outlined in McCann et al. (2018). Briefly, genomic DNA samples from two to three individuals were isolated and checked for quality and concentration using a NanoDrop spectrophotometer and Quant-iT PicoGreen dsDNA assay kit (PeqLab, Erlangen, Germany). DNA samples were pooled (species-wise) in equal proportions and fragmented to 600–800 nt in length. All samples were sequenced on a single lane of an Illumina HiSeq2500 machine using the 150 nt paired-end technology. Fragmentation, library preparation and sequencing were all performed at the CSF-NGS sequencing facility (Vienna Biocenter, Austria).

Read pre-processing, including quality filtering and removal of reads with similarity to the *PhiX* spike-in DNA (Illumina) or chloroplast sequences, was performed as outlined in McCann et al. (2018). *M. longipilum* was expected to be very different in nuclear DNA composition, due to incongruence in its phylogenetic position between nuclear and chloroplast markers (Blöch et al., 2009). Initial analyses confirmed extensive differences in repeat composition compared to other species and therefore *M. longipilum* was not included in the comparative analysis (see Discussion). The reads were analyzed using the command line implementation of the RepeatExplorer pipeline with the default settings and using the maximum number of reads possible with 100 GB of RAM. Briefly, the RepeatExplorer performs an all-to-all blast comparison and clusters sequence reads based on similarity (Novák et al., 2010, 2013). The clusters containing a minimum of 0.01% of the total reads used were annotated using BLAST searches to manually curated transposable element domain databases, graph structure, dot-plot structure (Sonnhammer and Durbin, 1995) and paired-end read connections (Macas et al., 2015).

A comparative analysis was also performed where forward ends of paired-end reads used in the individual analyses were randomly sampled proportional to each species genome size and pooled into a single dataset. The use of single-end reads increased the randomness of each sample of genome, whereas using reads from the individual analyses allowed for automated annotation of the comparative analysis clusters by tracing reads back to clusters of their origin. The settings for the comparative analysis were essentially the same as those of the individual species. The analyses were repeated three times to check for the congruency of the results.

The number of reads from each species was quantified for each repeat cluster and used to compare the abundance of each cluster across all species. Hs/Ho ratios were calculated for each read in the comparative analysis as outlined in Macas et al. (2015). This statistic is the ratio of the frequency of BLAST hits to reads from the same species (*Hs*) to the frequency of hits to reads from all other species (*Ho*). The ratio was calculated for *Ho* where only species within the same series are included and for all other species regardless of serial classification.

Phylogeny Reconstruction Using Repeat Abundances

Genomic repeat abundances obtained from annotated repeats in the comparative analysis were cube-root transformed and used as continuous characters to reconstruct the phylogeny of section *Melampodium* using the program RevBayes (Höhna et al., 2016), developed specifically for phylogenetic inference. The method of phylogenetic independent contrasts using restricted maximum likelihood (Felsenstein, 1985) was applied to reconstruct phylogenies using several different model specifications with varying complexity. All analyses were adapted from scripts provided in the RevBayes tutorials.

Common to all models, priors on the diversification and turnover rates were chosen as follows: a log-normal distribution with mean 0 and standard deviation 1 (in real space) and a gamma distribution with shape and rate parameters equal to 4. Each model specification differed only in the number of rates estimated across the cluster abundance data matrix. Theoretically, this number could range from one, where all repeats evolve with the same rate, to the total number of repeat clusters analyzed, which assigns an independent rate parameter to each cluster.

In addition to single rate and all independent rate model specifications, a number of intermediate alternative models were also tested. These models imposed various constraints on the rates of evolution for different repeat types. For example, one model allowed different repeat types to evolve at different rates, but within a single type the rate of evolution was forced to be the same. Some within-repeat type variation was allowed, especially in satellite DNA repeats, where clusters within a particular repeat type are likely to have evolved differently among species. Each rate in all model specifications was assigned a wide uniform prior ranging from −5 to 5 in log space. Stepping stone sampling was performed to estimate the marginal likelihoods for each model. The ratio (or difference if log-transformed) of the marginal likelihoods of any two models gives the Bayes factor, which is a metric commonly used for scoring relative model fit. The marginal likelihoods were calculated for each model and Bayes Factors were calculated for the top models. The best model was used for final tree inference and interpretation of patterns of repeat evolution in *Melampodium*.

RESULTS

Phylogenetic Inference Using Sequence Data in BEAST

The phylogenetic tree for twelve diploid species in *Melampodium* sect. *Melampodium* was inferred using the multi-species coalescent in BEAST and the resulting MCC tree for the multi-species coalescent inference in BEAST is shown in **Figure 1A**. Posterior nodal support across the tree was generally quite high, with clade support ranging from 37 to 100%. The clades representing the different series in section *Melampodium* had very high support (97–100%). The least well-supported node (37%) was the split joining series *Cupulata* with *Leucantha*. Additionally, the phylogenetic position of *Melampodium glabribracteatum* as sister to series *Melampodium* only had 74% support. This low certainty left the backbone of the phylogeny of section *Melampodium* unresolved (**Figure 1A**).

Ancestral Genome Size Reconstruction

Three different models of genome size evolution were tested on all trees obtained from the BEAST phylogenetic analyses. The simple Brownian motion model only had a single parameter to estimate, namely the rate of genome size evolution. The other two models, the Ornstein–Uhlenbeck and relaxed rate model, had one additional parameter each, namely parameters for the strength of selection toward the optimum value and the standard deviation of the log-normal rate distribution across the tree, respectively. For all trees, these three models reduced to simple Brownian motion, where the aforementioned parameters were estimated

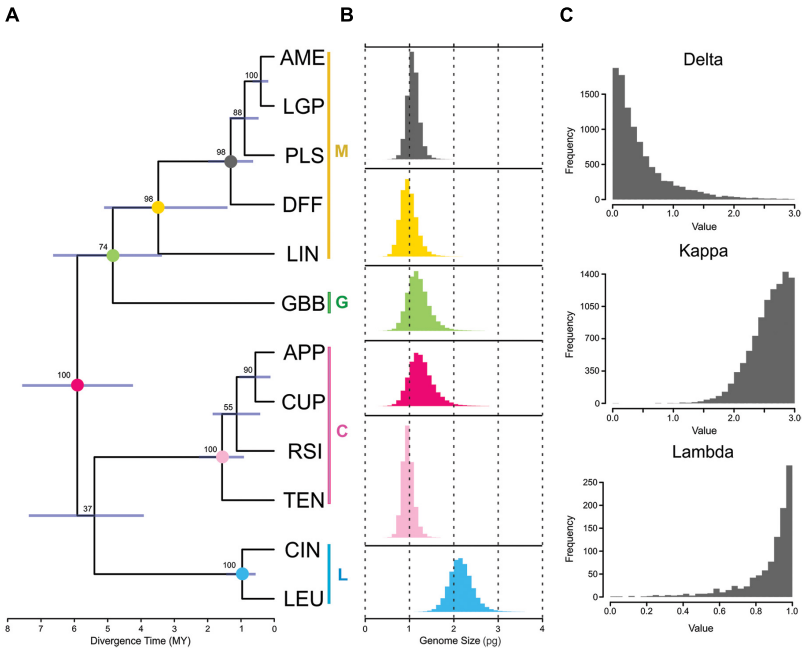

FIGURE 1 | Results of phylogenetic inference and ancestral genome size reconstruction in section *Melampodium*: **(A)** the maximum clade credibility tree from the BEAST analysis; Color vertical lines and letters next to species names indicate the taxonomic series the species belong to (C, series *Cupulata*, pink; G, series *Glabribracteata*, green; L, series *Leucantha*, blue; M, series *Melampodium*, yellow). **(B)** posterior distributions obtained from RevBayes analysis of ancestral genome size on important, well-supported nodes [annotated by color at nodes in **(A)**]; and **(C)** posterior distributions of commonly used tree transformation statistics (δ, κ, and λ) in BayesTraits. The light-blue bar at each node indicates uncertainty in the node height of the tree.

to be essentially equal to zero. Therefore, the simple Brownian motion model was considered the best evolutionary model for genome size evolution in section *Melampodium*.

The posterior distribution of ancestral genome sizes from the Brownian motion analysis for well-supported nodes in the maximum clade credibility tree are shown in **Figures 1A,B**. DNA sequence data from *Melampodium moctezumum* were not available for use in phylogenetic inference, and hence, it was not used for ancestral genome size inference. The mean genome size reconstructed for the ancestor of the entire group was 1.27 pg/1C with highest-posterior density (HPD) of 0.81–1.72. The mean values of genome size reconstructions for the individual series in section *Melampodium* were similar to the observed genome size ranges for the species they contained except for series *Melampodium*. This series contained *M. linearilobum*, which has a much smaller genome size than the other species in the group (**Table 1**). The mean of the posterior distribution for the whole group was 0.98 pg/1C, with HPD ranging from 0.68 to 1.28. The clade containing series *Melampodium* and the monotypic series *Glabribracteata* (1.89 pg/1C) had a mean of 1.20 pg/1C and HPD range of 0.79–1.59 (**Figures 1A,B**).

The program BayesTraits was used to estimate δ, κ, and λ, which are tree transformations that provide information about rates of evolution and inform if the particular trait appears to evolve independently of phylogenetic tree structure. The distribution of δ was strongly skewed toward 0 with a 95% HPD interval of 0.001–1.457 (**Figure 1C**). The distribution of κ, which has the same range of possible values as δ, was strongly skewed toward the other end of the range (3) with a 95% HPD of 1.976–3.0. The λ parameter, ranging between possible values of 0 and 1, had a 95% HPD interval of 0.543–1.0 (**Figure 1C**), indicating high phylogenetic signal in the genome size data.

Repetitive DNA Content of *Melampodium* Genomes

The *Melampodium* species analyzed using the RepeatExplorer pipeline are shown in **Table 1**. The genome sizes were published in previous studies (**Table 1**). Clustering of individual species reads was performed on a number of reads corresponding to between 0.40 and 2.13x coverage of the genome for each species (**Table 1**). Approximately 46–70% of the reads in each individual species analysis were found in medium to high copy number clusters comprising at least 0.05% of their genomes. Post-processing and further analysis allowed for the assignment of 90–96% of these clusters to specific repeat types, with the remainder left unannotated (**Table 2**). This amounted to approximately 40–64% of the total genome being annotated. Overall repeat composition in section *Melampodium* largely corresponded with the classification of the group at the series level. The genomic proportions of different repeat types observed in the analyzed genomes were summarized according to their taxonomic classification in **Table 2**.

Long-terminal-repeat-retrotransposons constituted the largest proportion of all analyzed *Melampodium* genomes, reaching up to nearly 60% in *M. glabribracteatum*. Nearly equal

TABLE 1 | General information about the analyzed species as well as the individual and comparative clustering analyses of high-throughput sequencing (HTS) reads using the RepeatExplorer bioinformatic pipeline.

Species	ID	Genome Size*	Individual clustering		Comparative clustering
		Gbp/1C	No. reads	Coverage	No. reads
Series *Melampodium*					
M. americanum	AME	1.11	7621416	0.89x	897117
M. diffusum	DFF	1.07	7797006	0.95x	888532
M. linearilobum	LIN	0.48	7957510	2.13x	388729
M. longipes	LGP	1.11	6736416	0.79x	880500
M. pilosum	PLS	1.07	7889718	0.96x	833622
Series *Glabribracteata*					
M. glabribracteatum	GBB	1.81	5555528	0.40x	1465723
Series *Cupulata*					
M. appendiculatum	APP	0.92	7522676	1.06x	753557
M. cupulatum	CUP	0.92	7327880	1.04x	738084
M. moctezumum	MZT	0.98	5946104	0.96x	737049
M. rosei	RSI	0.92	5787404	0.82x	721808
M. tenellum	TEN	0.92	4568954	0.65x	729206
Series *Leucantha*					
M. cinereum	CIN	2.17	7560312	0.45x	1761046
M. leucanthum	LEU	2.27	8062040	0.46x	1840574

** Genome sizes taken from Weiss-Schneeweiss et al. (2012) and Rebernig et al. (2012).*

TABLE 2 | Genome proportion estimates (%) from the individual RepeatExplorer analyses of repeats found in the genomes of the diploid species in *Melampodium* sect.

Type	Lineage	*Cupulata*	*Glabribracteata*	*Leucantha*	*Melampodium*
Retrotransposons	–*	32.5–48.6	59.7	54.1–55.1	38.5–49.8
Ty1-copia	–*	19.5–28.0	37.4	27.8–30.0	18.9–24.5
	SIRE	18.1–26.4	35.7	24.9–27.5	17.3–22.7
	Other	1.4–1.7	1.7	2.6–2.9	1.2–1.7
Ty3-gypsy	–*	11.7–17.4	21.0	20.2–23.4	19.0–23.5
	Athila	2.8–4.9	10.1	7.4–10.4	8.3–9.9
	Chromo	5.2–7.4	7.4	7.3–8.0	6.7–8.2
	Ogre/Tat	2.9–5.4	3.3	4.3–4.4	4.0–6.0
Other/Non-LTR	–	0.4–1.6	0.7	0.6–0.7	0.4–0.7
	Pararetrovirus	0.1–0.2	0.3	0.2–0.3	0.1–0.2
	LINE	0.2–0.7	0.2	0.2–0.2	0.0–0.1
	Other	0.1–0.7	0.2	0.2–0.2	0.3–0.4
DNA transposons	–	0.6–1.9	2.8	1.1–1.4	1.0–4.5
	CACTA	0.2–0.9	2.4	0.2–0.3	0.6–3.6
	Other	0.4–1.0	0.4	0.9–1.1	0.4–0.9
Tandem repeats	–	4.2–8.1	1.3	1.2–1.7	2.0–4.4
	35S rDNA	0.9–3.0	1.0	0.5–0.6	0.8–2.0
	5S rDNA	0.1–0.2	0.1	0.1–0.1	0.1–0.1
	Satellite DNAs	3.2–4.9	0.2	0.7–1.0	1.3–2.3
Unclassified	–	4.0–5.8	5.3	6.3–6.6	7.9–10.0
Low-Copy	–	37.9–54.1	29.8	34.7–34.8	35.4–43.5

*Melampodium grouped in series. Ranges are given for series with more than one species. * Total content of a given type including LTR-retrotransposons not assigned to either a superfamily or specific lineage within the superfamily.*

proportions of *Ty1-copia* and *Ty3-gypsy* elements were found in most species, with the exception of those in series *Glabribracteata* and *Leucantha*. These series had significantly higher amounts of *Ty1-copia* relative to *Ty3-gypsy* elements (nearly 3:1 in *M. glabribracteatum*) than the remaining series (**Table 2**). SIRE repeats made up the vast majority of the *Ty1-copia* elements (18–36%), while the other *Ty1-copia* lineages combined represented <3% of the analyzed genomes. The genomic

proportions of the three lineages of the *Ty3-gypsy* superfamily (Athila, Chromovirus, Ogre-Tat) were more balanced, although the Athila lineage was typically the most common (**Table 2**). Pararetroviruses, LINE, SINE, and MITE elements were found only in trace amounts in *Melampodium*.

Most of the lineages of DNA transposons and the Helitrons were represented in *Melampodium* genomes, albeit mostly in small amounts (**Table 2**). Of these lineages, *CACTA* elements were typically the most abundant, reaching genome proportions of 0.2 up to 6.6% of the genome in series *Cupulata* and in *M. linearilobum* of series *Melampodium*, respectively. The Helitrons and other DNA transposon lineages were found only in trace amounts across all species in section *Melampodium*.

Tandem repeats (5S rDNA, 35S rDNA, and satellite DNAs) were found in relatively low abundance in all species. Both the 35S and 5S rRNA genes ranged from 0.5 to 3% and 0.1 to 0.2% in the genome, respectively (**Table 2**). The number of satellite DNAs observed across all genomes was relatively low. Satellite DNAs were found in proportions as low as 0.5% in *M. glabribracteatum* and approaching 5% in some species of the *Cupulata* group.

Dynamics of Shared Repeats Across Section *Melampodium*

The comparative analysis of the repetitive DNA fraction entailed clustering of reads from different species to identify repeats that were shared or species/group specific across all genomes. This was performed using reads from all diploid species in section *Melampodium* listed in **Table 1**. The total number of reads analyzed per species corresponded to 0.1x coverage of each of their genomes (**Table 1**). These reads were clustered together and resulted in 438 clusters comprising at least 0.05% of the total number of reads analyzed.

The top 438 clusters exhibited wide variation in the proportion of reads from individual species, ranging from clusters containing reads from all species to clusters composed of reads from multiple species in a single taxonomic series (**Figure 2**). Overall, the majority of clusters contained reads from all species, where read abundance for each species was proportional to their genome sizes. The distribution of reads from LTR-retrotransposons mostly followed this trend, with the exception of clusters annotated as *Ty1-copia* SIRE lineage. These repeats exhibited both proportional distribution of species' reads along with several clusters being specific to either *M. glabribracteatum* or series *Leucantha*. On the other hand, the vast majority of repeat clusters in the *Ty3-gypsy* superfamily and *Ty1-copia* lineages other than SIRE were shared across all species. The DNA transposons displayed similar patterns to clusters of *Ty1-copia* type, with one lineage (*CACTA*) showing series-specificity and other lineages containing reads from all species.

Sequence similarity profiles (*Hs/Ho* distributions) were calculated for both within-series and between-series comparisons. The within-series comparisons revealed single peaks centered around zero for all major types of transposons (**Figure 3**). Between-series comparisons also showed single peaks for all *Ty3-gypsy* and *Ty1-copia* elements not derived from the SIRE lineage, albeit with means located around 1.5 (**Figures 3B,D**). The DNA transposons and *Ty1-copia* SIRE elements had between-series *Hs/Ho* distributions with secondary peaks shifted to the right (**Figures 3A,C**). The secondary peaks for the DNA transposons and *Ty1-copia* SIRE retroelements were centered around 2 and 6, respectively, indicating significant differences in similarities in this repeat type among groups.

Most of the satellite DNAs identified in the comparative analysis dataset exhibited a high degree of series-specificity in section *Melampodium* (**Figure 4B**). Eight satellite DNA clusters were specific to the series *Cupulata* alone (**Figure 4**). Seven of the eight satellite DNAs found in this series were highly similar both in monomer length and sequence composition, while the eighth (MelSat3, **Figure 4A**) was similar to a satellite found also in the *Leucantha* series (MelSat6, **Figure 4A**). Several unique, series-specific satellite DNA clusters were also found in series *Leucantha*, although, unlike in the *Cupulata* group, monomer sequences showed no similarity to one another. One 28 nt satellite monomer was recovered in all 13 analyzed species (MelSat8, **Figure 4B**). The majority of the monomer lengths in these species were approximately 180 nt in length, but ranged from 4 to 1200 nt. One cluster, which was recovered mostly from the *Leucantha* series (MelSat7, **Figure 4B**), contained perfect 7 nt telomeric repeats (TTTAGGG).

Phylogenetic Support for Independent Evolution of Repeat Types in Section *Melampodium*

Repetitive DNA evolution in *Melampodium* was further analyzed by reconstructing the phylogeny of the section using repeat abundances from the comparative analysis. These abundances were treated as continuous characters and used to construct phylogenies under several Brownian motion models of evolution with varying levels of complexity (i.e., number of rate parameters). Model selection was performed to determine which model best fits the data, thus providing information about which repeats evolve at similar rates.

Marginal likelihood estimates (MLE) were calculated using stepping stone sampling and ranged from −6954.64 to −42777 (**Table 3**). The simplest and most complex models had the lowest MLEs, while a model with seven rate parameters performed the best. This model applied a single unique rate to each of the following groups: DNA transposons, LINE, *Ty1-copia* SIRE, non-SIRE *Ty1-copia*, *Ty3-gypsy*, and tandem repeats (including rDNA repeats). The remaining repeat types were grouped together into a single rate class.

The relationships recovered from the phylogenetic analysis using repeat abundances (with the best supported model) as characters are shown in **Figure 5A**. In general, the posterior support in this analysis was higher than that from the BEAST analysis. The node posterior support across the entire tree was always greater than 90% and in most cases was 100% supported. The serial classification within section *Melampodium* was well-supported in this analysis. The maximum *a posteriori* (MAP) tree placed series *Melampodium* and *Cupulata* as sister groups, whereas *M. glabribracteatum* was recovered as sister to those groups. Series *Leucantha* was sister to all others.

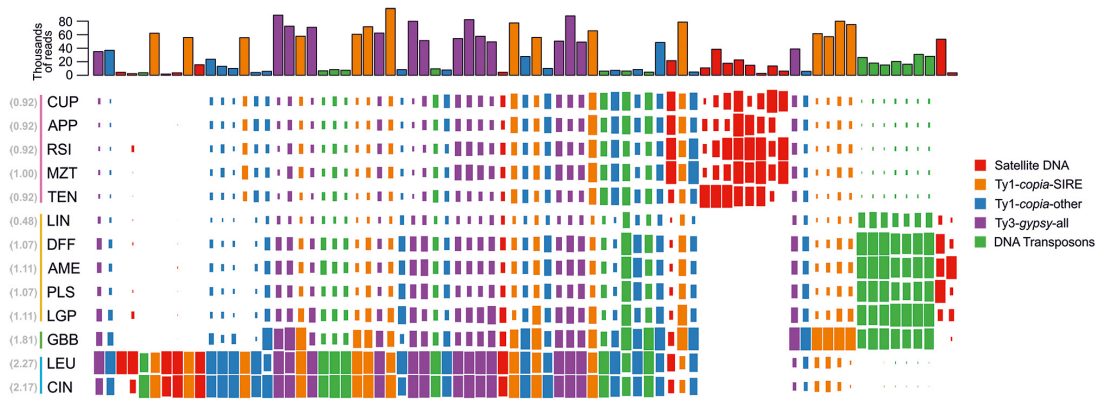

FIGURE 2 | Normalized genome representation of the top clusters of five major repeat groups found in the genomes of species analyzed in the comparative analysis. Each of the four major repeat types (Ty1-*copia*-SIRE; Ty1-*copia*-others; Ty3-*gypsy*-all and DNA transposon) is represented by 15 top clusters whereas satellite DNA is represented by all 17 clusters. The repeats are hierarchically clustered along the horizontal axis so that repeats with similar distributions across species are grouped together. Genome size (1C value) for each species is indicated next to species name abbreviation (in brackets, gray) and the color vertical lines to the left of the species names indicate the taxonomic series the species belong to (series *Cupulata*, pink; series *Glabribracteata*, green; series *Leucantha*, blue; series *Melampodium*, yellow).

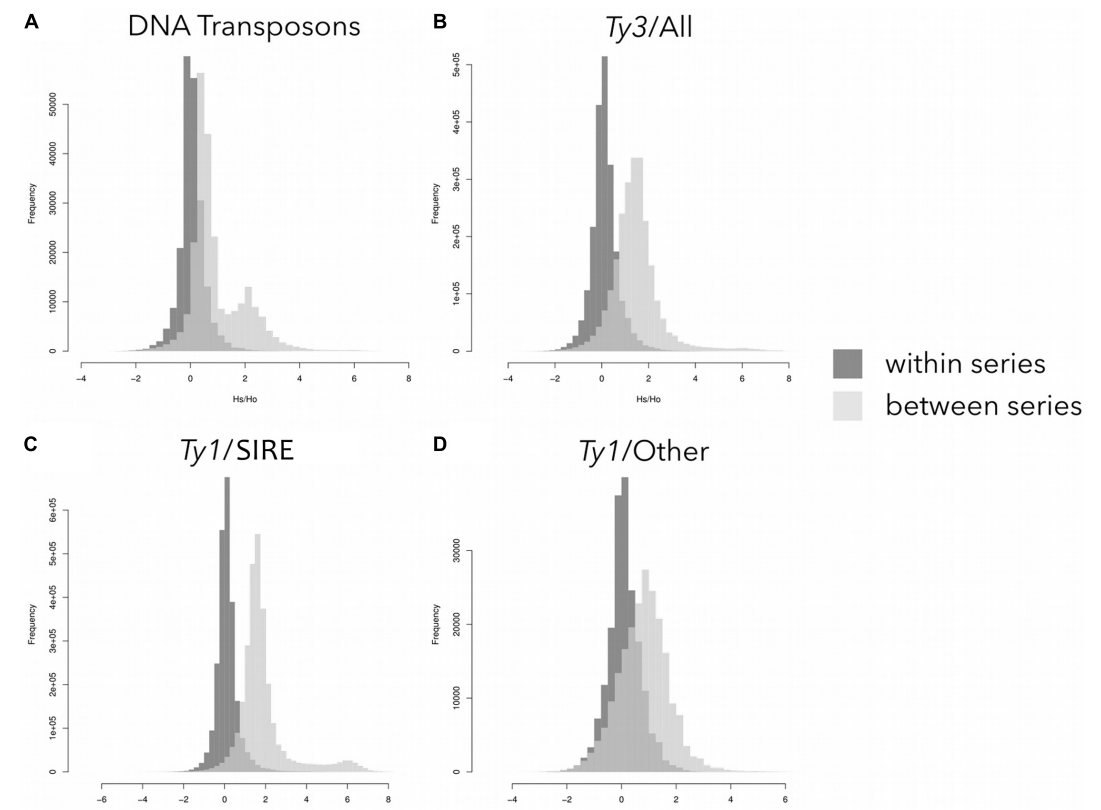

FIGURE 3 | Sequence similarity profiles (distributions of the logarithm of Hs/Ho) for different repeat types: **(A)** DNA transposons; **(B)** Ty3-*gypsy* elements; **(C)** Ty1-*copia* SIRE elements; and **(D)** other Ty1-*copia* elements except for the SIRE lineage. For each read the Hs/Ho ratio was calculated using either hits to reads from species in the same series (dark gray) or hits to reads from all other species (light gray).

To investigate differences in topology and support, phylogenetic analyses were performed for a few of the major repeat types separately. Maximum *a posteriori* (MAP) trees were constructed for each analysis (**Figures 5B–D**). The two MAP trees derived from analyses of *Ty1-copia* or *Ty1-gypsy* elements (**Figures 5B,C**, respectively) showed similar topology

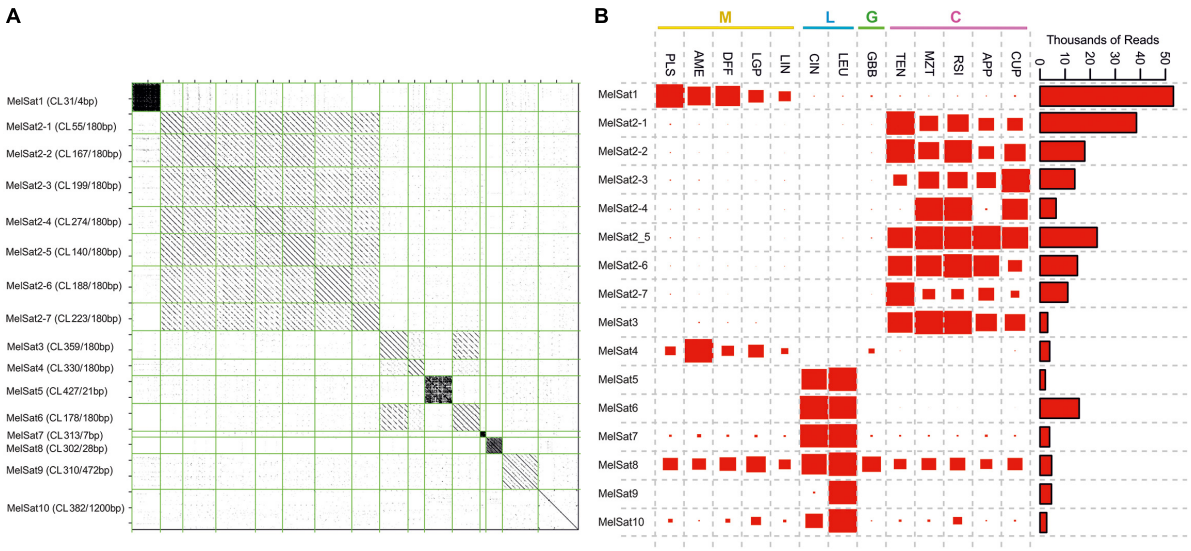

FIGURE 4 | A comparison of satellite DNAs identified in the comparative analysis of all species analyzed in this study: **(A)** Dot-plot comparison of identified satellite DNA clusters (each represented by one contig) and **(B)** the contribution of reads from different species to each satellite DNA cluster. The area of red rectangles is proportional to the number of reads (abundance) contributed by each species. Color bars and capital letters above species names indicate the taxonomic series the species belong to (C, series *Cupulata*, pink; G, series *Glabribracteata*, green; L, series *Leucantha*, blue; M, series *Melampodium*, yellow).

TABLE 3 | Marginal likelihood estimates (MLE) for all models used to construct repeat phylogenies.

Model	1	2	3	4	5	6	7	8	9	10
No. parameters	1	6	7	19	23	24	24	25	40	307
MLE	−7284.09	−7030.16	−6954.64	−7074.39	−6971.97	−7005.63	−7058.18	−6984.70	−7055.72	−42777

Full model specifications are available from the authors upon request.

including their branch lengths, with slightly weaker support in the *Ty3-gypsy* tree moving toward the tips of the tree. The satellite DNA tree topology (**Figure 5D**), however, had overall much weaker node support, larger differences in branch lengths and well-supported topological differences from both LTR-retrotransposon based trees and the tree based on all repeats.

DISCUSSION

The diversity of repetitive DNAs and the inherent propensity for their amplification and removal from the host genome greatly contributes to the variation of plant genome sizes and composition. This study addresses repeatome dynamics across the genomes of 13 closely related wild diploid species of genus *Melampodium* that differ nearly 4.5-fold in genome size despite having the same base chromosome number (x = 10). Evolutionary changes in repetitive DNA composition and genome size were complex, with independent patterns of genome up and downsizing throughout the evolution of the analyzed diploids mostly due to changes in abundances of the SIRE (*Ty1-copia*), Athila (*Ty3-gypsy*), and *CACTA* (DNA transposon) lineages. Evidence for strong phylogenetic signal and differential evolutionary rates of major lineages of repeats in the diploid genomes of *Melampodium* was also inferred.

Genome Size Evolution

Genome size values for diploid species in section *Melampodium* range from 0.49 to 2.27 pg/1C (Rebernig et al., 2012; Weiss-Schneeweiss et al., 2012). The range of genome sizes for all species in section *Melampodium* are below average relative to other species in Asteraceae (2.91 ± 1.22, Vallès et al., 2013). Patterns of genome size variation in section *Melampodium* follow the serial classification and are, thus, congruent with phylogenetic position of taxa. The only exception is *M. linearilobum*, which has reduced genome size relative to the other species in the same series (*Melampodium*).

The correlation between species-relatedness and genome size observed in section *Melampodium* suggests that this character contains phylogenetic signal. Statistical support for this finding was provided by the estimation of *Pagel's*λ, which demonstrated the dependence of genome size on the internal structure of the tree. The strong phylogenetic signal of genome size in section *Melampodium* and lack of support for the Ornstein–Uhlenbeck model provides evidence against processes that could erode this signal, such as selection toward a single optimum genome size (Bloomberg et al., 2003; Harmon et al., 2010). Estimates of the other two commonly used statistics, κ and δ, revealed that most evolution of genome size has occurred early on in the phylogeny and along longer branches (Pagel and Meade, 2004), supporting adaptive radiation and gradualism in

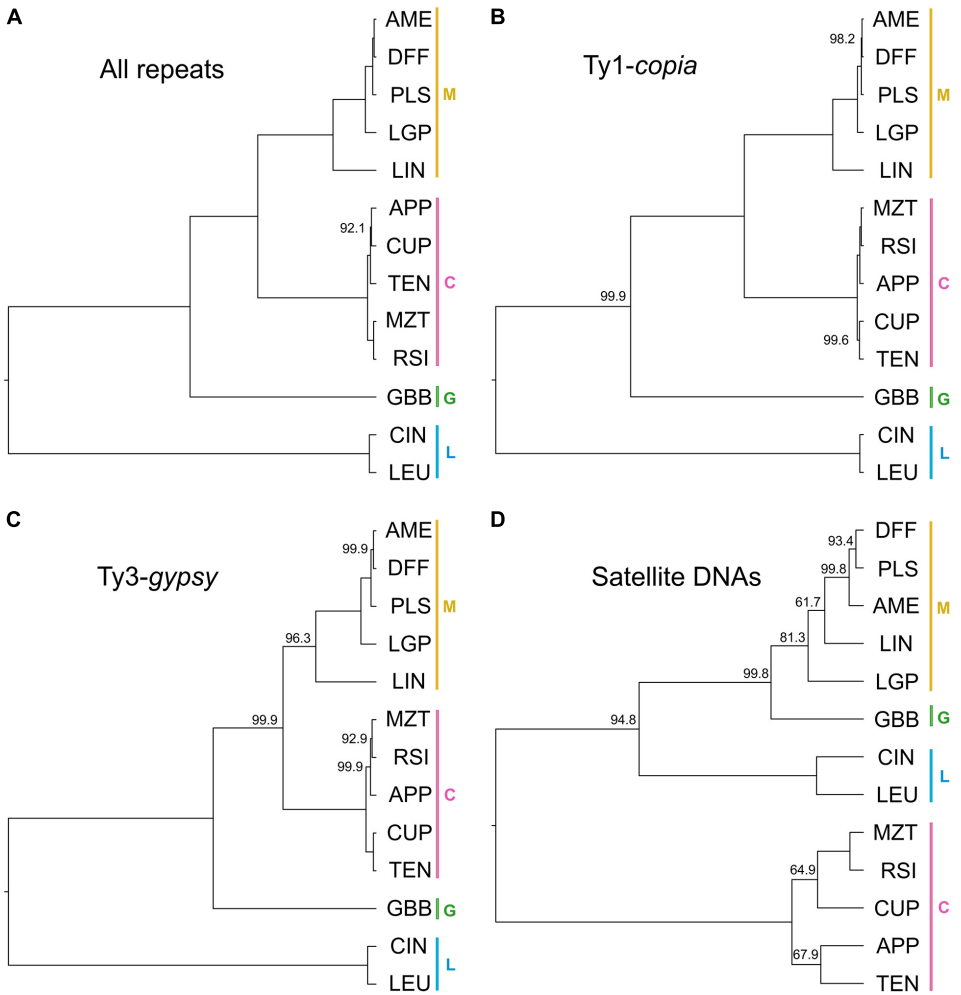

FIGURE 5 | Maximum *a posteriori* trees recovered from repeat-abundance based phylogenetic analysis with a single rate parameter and fixed root height of 1: **(A)** all repeats included with best model; **(B)** Ty1-*copia* elements only; **(C)** Ty3-*gypsy* elements only; and **(D)** satellite DNAs only. Color bars and capital letters next to species names indicate the taxonomic series the species belong to (C, series *Cupulata*, pink; G, series *Glabribracteata*, green; L, series *Leucantha*, blue; M, series *Melampodium*, yellow).

the group, respectively. A punctuational hypothesis of genome size evolution has been supported in other plant groups, through either arguments against the correlation of sequence substitution rates and genome size in evolution or via statistical approaches similar to that applied in this study (Albach and Greilhuber, 2004; Lysak et al., 2009). While trees with branch lengths proportional to average substitutions per site may not be the best models for inferring patterns of genome size evolution, this can be ameliorated through the use of ultrametric trees, where branch lengths are proportional to time. Such an alternative approach circumvents the need for a strong *a priori* assumption that genome size evolves either independent of branch length or correlated with substitution rates and can instead be tested in a statistical framework (Cusimano and Renner, 2014; McCann et al., 2016).

Independence of character evolution and branch length implies punctuational change along the phylogeny (Lysak et al., 2009). In plants, such changes in genome size can be the result of either polyploidization or transposable element activity (Bennetzen and Wang, 2014). In the present analysis, only diploid lineages were analyzed therefore eliminating the need to consider recent polyploidy as a potential explanation for the observed genome sizes. Estimates of κ for section *Melampodium* suggest gradual evolution of genome size on longer branches in the phylogeny, as has been reported in Brassicaceae (Lysak et al., 2009). Branches leading to lineages with the largest differences in genome size (i.e., *M. glabribracteatum, M. linearilobum* and the two species in series *Leucantha*) are also the longest in the tree, suggesting more gradual changes of genome size in these species over time. Alternatively, punctuational increase of transposable element activity early on along long branches could also explain the observed genome size patterns. However, a model allowing for rate heterogeneity among branches (relaxed-rate model)

in the tree was not supported over standard, constant-rate Brownian motion.

Evolution of Repetitive DNA

Quantification of the contribution of various elements to the repetitive portion of diploid genomes in section *Melampodium* was performed using the RepeatExplorer pipeline, the utility and accuracy of which has been demonstrated in multiple previous studies (Novák et al., 2010, 2014; Macas et al., 2015). All but two of the diploid species of section *Melampodium* have been analyzed in this study. Material for *M. sinuatum* was not available. *M. longipilum* (series *Longipila*) has been sequenced and was analyzed individually. This analysis revealed significant differences in repeat composition relative to the other species in section *Melampodium*, which is congruent with its phylogenetic position using nuclear markers, where it groups outside of section *Melampodium* in non-sister group relation (Blöch et al., 2009; Stuessy et al., 2011; McCann et al., 2018). Thus, it was not included in the comparative analysis.

The repetitive fraction of *Melampodium* genomes ranged from 46 to 70% with larger genomes having correspondingly higher amounts of repeats. This range is within the estimates published for other species including a close relative *Helianthus annuus* (>3.0 pg/1C and 81% repetitive; Natali et al., 2013). The abundance of the various repeat types in plant genomes is largely group-specific and responsible for genome size evolution. For example, *Ty3-gypsy* elements were found to be most abundant in both *Gossypium* and species in Fabeae (Hawkins et al., 2009; Macas et al., 2015), and were inferred to play a role in genome size differentiation. In *Melampodium*, the SIRE lineage of *Ty1-copia* superfamily was largely responsible for most of the observed genome size increases, as the larger genomes of *M. glabribracteatum* and the two species in series *Leucantha* have higher amounts of these elements. However, *Ty1-copia* SIRE elements do not play a significant role in the genome downsizing observed in *M. linearilobum*. The genome of *M. linearilobum*, although strongly reduced in size, contains repeats in similar proportions to those found in the other species in the same series. In conjunction with the ancestral genome size analysis, this suggests proportional decrease of all repeat types in *M. linearilobum*.

The comparative analysis conducted in this study is one of the few to have near complete sampling of a group of related species, which allows for a better understanding of repetitive DNA evolution in a closely related group of species. This analysis confirmed the differential amplification of *Ty1-copia* SIRE lineages by revealing repeat clusters of SIRE type with higher abundances in *M. glabribracteatum* and the two species in the *Leucantha* series. Sequence similarity profiles (*Hs*/*Ho* distributions) showed significant secondary peaks in read comparisons among taxonomic series for reads derived from SIRE repeats, suggesting that similarity among these elements differs between groups. The results of ancestral genome size reconstruction and repetitive DNA analyses together provide evidence for continuous amplification of various subtypes of SIRE retroelements over time in section *Melampodium*, whereas amplification of specific repeat types has been shown to be punctuated in other plant groups (Hawkins et al., 2009; Belyayev, 2014; Macas et al., 2015). The SIRE lineage, thus, likely represents a recently active retroelement type in section *Melampodium*.

DNA transposon composition also differed across the genomes of species in section *Melampodium*. However, the lower overall genome proportion of these elements, relative to retrotransposons, suggests a limited role in genome size differentiation. A secondary peak in the sequence similarity profile (*Hs*/*Ho* distribution) for DNA transposons was found, although it was not as pronounced as that of *Ty1-copia* SIRE elements. This secondary peak was likely the result of increased amount of DNA transposons in the species of series *Melampodium* and *Glabribracteata*.

Satellite DNAs can arise from any DNA sequence in the genome and such repeats typically exhibit high rates of evolution both in monomer sequence and overall abundance (Tek et al., 2005; Macas et al., 2010; Melters et al., 2013). Several satellite DNA repeats were recovered in the comparative analysis and these exhibited the highest taxonomic specificity. Monomer length of satellite DNA repeats found in section *Melampodium* ranged from a 4 nt microsatellite (ATTC) to 1200 nt. However, the most frequently occurring monomer length found among the satellite DNAs was around 180 bases. This fits a common pattern found in satellite DNAs in that monomer lengths around 165 nt appear to be favored, which may be a consequence of structural constraints imposed by the nucleosome (Macas et al., 2002; Garrido-Ramos, 2015). Higher copy numbers of telomeric sequences recovered in *M. leucanthum* and *M. cinereum* might be caused by technical constrains either during the DNA extraction or DNA sequencing or, alternatively, might suggest the presence of longer telomeres in these two species.

Several satellite DNA repeat clusters were shared by all species in series *Cupulata*, nearly all of which had the same monomer length and highly similar monomer sequences. There is no indication of species-specificity for monomers of individual clusters, as most are shared among all series *Cupulata* species. Given such a pattern, these different satellite DNA monomers are likely descendants of one common repeat. Inter-chromosomal divergence and specificity of centromeric satellites has been observed previously in other study systems such as *Oryza* (Macas et al., 2010), *Arabidopsis* (Heslop-Harrison et al., 1999), and in humans (Rudd et al., 2006), suggesting that homogenization of centromere-specific satellite DNA repeats occurs mostly within individual chromosomes with low rates of inter-chromosomal spread. The chromosomal localization of these satellite DNA repeats in series *Cupulata*, however, has not yet been determined.

Phylogenetic Signal in the Evolution of Repeats

The utility of repeat abundances in the genomes has been shown to allow reliable reconstruction of the phylogeny of a related group of species using both maximum parsimony and likelihood methods (Dodsworth et al., 2015). The program used for

maximum likelihood estimation of a phylogeny (Felsenstein, 1993) assumes that each character evolves independently. However, this may lead to model overfitting as repeat clusters stemming from the same lineage may have correlated evolutionary histories.

In this study, we relaxed the assumption of independent rates of evolution and performed model testing in a Bayesian framework to reduce the number of rate parameters in the model, and thus reduce overfitting the model to the data. Our results show that neither the simplest or most complex models (single rate or all independent rates, respectively) fit the data as well as models restricting rates based on repeat annotation, although models with fewer parameters generally perform better.

Repeat clusters of different origin can provide variable levels of phylogenetic signal, therefore affecting tree topology and resolution (Dodsworth et al., 2015; Weiss-Schneeweiss et al., 2015). Our comparative analysis and model testing, as well as analysis of repeats in many other genomes (Renny-Byfield et al., 2012; Dodsworth et al., 2015; Macas et al., 2015), show that the repetitive landscape of genome does not evolve in a uniform fashion. Satellite DNA repeats may be expected to resolve relationships closer to the tips of the tree due to their high rates of evolution. Other repeats, such as retroelements, may provide better resolution toward the root, as they have persisted in the genome over longer evolutionary time. Phylogeny reconstruction using repeats derived from the *Ty1-copia* and *Ty3-gypsy* superfamilies agreed in overall topology and branch lengths, although the *Ty1-copia* tree had higher posterior support. The phylogeny obtained from abundances of satellite DNAs had much lower support and a different topology. These findings demonstrate that the levels of phylogenetic signal are dependent on repeat type and can lead to different topologies with varying levels of support.

DATA AVAILABILITY STATEMENT

The data generated by this study can be found in the European Nucleotide Archive (ENA) using accession number PRJEB36721 (ERP119943).

AUTHOR CONTRIBUTIONS

HW-S, JMc, and TS conceived and coordinated the study. JMc performed the research. JMc, HW-S, JMa, and PN analyzed the data. HW-S, JMc, JMa, and PN interpreted the data and wrote the manuscript with input from TS and JV. All authors read and approved the final manuscript.

FUNDING

The authors acknowledge financial support of the Austrian Science Fund (FWF P25131 to HW-S) and computing resources provided by the Vienna Scientific Cluster (Vienna, Austria). Access to the computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum (Czechia) and ELIXIR CZ research infrastructure project (MEYS Grant No: LM2015047) are also appreciated.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00362/full#supplementary-material

REFERENCES

Albach, D. C., and Greilhuber, J. (2004). Genome size variation and evolution in *Veronica*. *Ann. Bot.* 94, 897–911. doi: 10.1093/aob/mch219

Barghini, E., Natali, L., Cossu, R. M., Giordani, T., Pindo, M., Cattonaro, F., et al. (2014). The peculiar landscape of repetitive sequences in the olive (*Olea europaea* L.) genome. *Genome Biol. Evol.* 6, 776–791. doi: 10.1093/gbe/evu058

Belyayev, A. (2014). Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* 27, 2573–2584. doi: 10.1111/jeb.12513

Bennetzen, J. L. (2005). Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 95, 127–132. doi: 10.1093/aob/mci008

Bennetzen, J. L., and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* 65, 505–530. doi: 10.1146/annurev-arplant-050213-035811

Blöch, C., Weiss-Schneeweiss, H., Schneeweiss, G. M., Barfuss, M. H., Rebernig, C. A., Villaseñor, J. L., et al. (2009). Molecular phylogenetic analyses of nuclear and plastid DNA sequences support dysploid and polyploid chromosome number changes and reticulate evolution in the diversification of *Melampodium* (*Millerieae*, *Asteraceae*). *Mol. Phyl. Evol.* 53, 220–233. doi: 10.1016/j.ympev.2009.02.021

Bloomberg, S. P., Garland, J. T., Ives, A. R., and Crespi, B. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x

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

Cusimano, N., and Renner, S. S. (2014). Ultrametric trees or phylograms for ancestral state reconstruction: does it matter? *Taxon* 63, 721–726. doi: 10.12705/634.14

Dodsworth, S., Chase, M. W., Kelly, L. J., Leitch, I. J., Macas, J., Novak, P., et al. (2015). Genomic repeat abundances contain phylogenetic signal. *Syst. Biol.* 64, 112–126. doi: 10.1093/sysbio/syu080

Dodsworth, S., Jang, T.-S., Struebig, M., Chase, M. W., Weiss-Schneeweiss, H., and Leitch, A. R. (2016). Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). *Plant Syst. Evol.* 303, 1013–1020. doi: 10.1007/s00606-016-1356-9

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

Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.* 125, 1–15. doi: 10.1086/286013

Felsenstein, J. (1993). *PHYLIP: Phylogenetic Inference Package, Version 3.5*. Available online at: http://evolution.genetics.washington.edu/phylip.html (accessed July 14, 2015).

Fleischmann, A., Michael, T. P., Rivadavia, F., Sousa, A., Wang, W., Temsch, E. M., et al. (2014). Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Ann. Bot.* 114, 1651–1663. doi: 10.1093/aob/mcu189

Garrido-Ramos, M. A. (2015). Satellite DNA in plants: more than just rubbish. *Cytogenet. Genome Res.* 146, 153–170. doi: 10.1159/000437008

Harmon, L. J., Losos, J. B., Davies, J. T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64, 2385–2396. doi: 10.1111/j.1558-5646.2010.01025.x

Hawkins, J. S., Proulx, S. R., Rapp, R. A., and Wendel, J. F. (2009). Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17811–17816. doi: 10.1073/pnas.0904339106

Heslop-Harrison, J. S., Murata, M., Ogura, Y., Schwarzacher, T., and Motoyoshi, F. (1999). Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. *Plant Cell* 11, 31–42. doi: 10.1105/tpc.11.1.31

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., et al. (2016). RevBayes: bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* 65, 726–736. doi: 10.1093/sysbio/syw021

Horvilleur, B., and Lartillot, N. (2014). Monte Carlo algorithms for Brownian phylogenetic models. *Bioinform.* 30, 3020–3028. doi: 10.1093/bioinformatics/btu485

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinform.* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kelly, L. J., Renny-Byfield, S., Pellicer, J., Macas, J., Novák, P., Neumann, P., et al. (2015). Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytol.* 208, 596–607. doi: 10.1111/nph.13471

Kovařík, A., Dadejova, M., Lim, Y. K., Chase, M. W., Clarkson, J. J., Knapp, S., et al. (2008). Evolution of rDNA in *Nicotiana* allopolyploids: a potential link between rDNA homogenization and epigenetics. *Ann. Bot.* 101, 815–823. doi: 10.1093/aob/mcn019

Lysak, M. A., Koch, M. A., Beaulieu, J. M., Meister, A., and Leitch, I. J. (2009). The dynamic ups and downs of genome size evolution in *Brassicaceae*. *Mol. Biol. Evol.* 26, 85–98. doi: 10.1093/molbev/msn223

Macas, J., Meszaros, T., and Nouzova, M. (2002). PlantSat: a specialized database for plant satellite repeats. *Bioinform* 18, 28–35. doi: 10.1093/bioinformatics/18.1.28

Macas, J., Neumann, P., Novák, P., and Jiang, J. (2010). Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in large-scale sequencing data. *Bioinform.* 26, 2101–2108. doi: 10.1093/bioinformatics/btq343

Macas, J., Novák, P., Pellicer, J., Čížková, J., Koblížková, A., Neumann, P., et al. (2015). In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe *Fabeae*. *PLoS One* 10:e0143424. doi: 10.1371/journal.pone.0143424

McCann, J., Jang, T.-S., Macas, J., Schneeweiss, G. M., Matzke, N. J., Novák, P., et al. (2018). Dating the species network: allopolyploidy and repetitive DNA evolution in American daisies (*Melampodium* sect. *Melampodium*, *Asteraceae*). Syst. Biol. 67, 1010–1024. doi: 10.1093/sysbio/syy024

McCann, J., Schneeweiss, G. M., Stuessy, T. F., Villaseñor, J. L., and Weiss-Schneeweiss, H. (2016). The impact of reconstruction methods, phylogenetic uncertainty and branch lengths on inference of chromosome number evolution in American daisies (*Melampodium*. *Asteraceae*). *PLoS One* 11:e0162299. doi: 10.1371/journal.pone.0162299

Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J. G., et al. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* 14:R10. doi: 10.1186/gb-2013-14-1-r10

Natali, L., Cossu, R. M., Barghini, E., Giordani, T., Buti, M., Mascagni, F., et al. (2013). The repetitive component of the sunflower genome as shown by different procedures for assembling next generation sequencing reads. *BMC Genom.* 14:686. doi: 10.1186/1471-2164-14-686

Novák, P., Hřibová, E., Neumann, P., Koblížková, A., Doležel, J., and Macas, J. (2014). Genome-wide analysis of repeat diversity across the family Musaceae. *PLoS One* 9:e98918. doi: 10.1371/journal.pone.0098918

Novák, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform* 11:378. doi: 10.1186/1471-2105-11-378

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinform* 29, 792–793. doi: 10.1093/bioinformatics/btt054

Ogilvie, H. A., and Drummond, A. J. (2016). StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *BioRxiv* [preprint]. doi: 10.1093/molbev/msx126

Pagel, M., and Meade, A. (2004). *BayesTraits Manual*. Available online at: www.evolution.rdg.ac.uk/BayesTraits.html (accessed December 12, 2015).

Pellicer, J., Fay, M. F., and Leitch, I. J. (2010). The largest eukaryotic genome of them all? *Bot. J. Linn. Soc.* 164, 10–15. doi: 10.1111/j.1095-8339.2010.01072.x

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News* 6, 7–11.

Rambaut, A. (2007). *Tracer Version 1.6*. Available online at: http://beast.bio.ed.ac.uk/Tracer (accessed February 22, 2016).

Rebernig, C. A., Weiss-Schneeweiss, H., Blöch, C., Turner, B., Stuessy, T. F., Obermayer, R., et al. (2012). The evolutionary history of the white-rayed species of *Melampodium* (*Asteraceae*) involved multiple cycles of hybridization and polyploidization. *Am. J. Bot.* 99, 1043–1057. doi: 10.3732/ajb.1100539

Renny-Byfield, S., Chester, M., Nichols, R. A., and Macas, J. (2012). Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. *PLoS One* 7:e36963. doi: 10.1371/journal.pone.0036963

Renny-Byfield, S., Kovařík, A., Kelly, L. J., Macas, J., Novak, P., Chase, M. W., et al. (2013). Diploidization and genome size change in allopolyploids is associated with differential dynamics of low-and high-copy sequences. *Plant J.* 74, 829–839. doi: 10.1111/tpj.12168

Rudd, M. K., Wray, G. A., and Willard, H. F. (2006). The evolutionary dynamics of α-satellite. *Genome Res.* 16, 88–96. doi: 10.1101/gr.3810906

Sonnhammer, E. L. L., and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167, 1–10. doi: 10.1016/0378-1119(95)00714-8

Stuessy, T. F., Blöch, C., Villaseñor, J. L., Rebernig, C. A., and Weiss-Schneeweiss, H. (2011). Phylogenetic analyses of DNA sequences with chromosomal and morphological data confirm and refine sectional and series classification within *Melampodi*um (*Asteraceae*. *Millerieae*). *Taxon* 60, 436–449. doi: 10.1002/tax.602013

Tek, A. L., Song, J., Macas, J., and Jiang, J. (2005). *Sobo*, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. *Genetics* 170, 1231–1238. doi: 10.1534/genetics.105.041087

Ugarković, Đ (2008). Satellite DNA libraries and centromere evolution. *Open Evol. J.* 2, 1–6. doi: 10.2174/1874404400802010001

Ugarković, Đ, and Plohl, M. (2002). Variation in satellite DNA profiles—causes and effects. *EMBO J.* 21, 5955–5959. doi: 10.1093/emboj/cdf612

Vallès, J, Canela, M. Á, Garcia, S., Hidalgo, O., Pellicer, J., Sánchez-Jiménez, I., et al. (2013). Genome size variation and evolution in the family *Asteraceae*. *Caryologia* 66, 221–235. doi: 10.1080/00087114.2013.829690

Vu, G. T. H., Schmutzer, T., Bull, F., Cao, H. X., Fuchs, J., Tran, T. D., et al. (2015). Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus. *Plant Genome* 8, 1–14. doi: 10.3835/plantgenome2015.04.0021

Weiss-Schneeweiss, H., Blöch, C., Turner, B., Villaseñor, J. L., Stuessy, T. F., and Schneeweiss, G. M. (2012). The promiscuous and the chaste: frequent allopolyploid speciation and its genomic consequences in American daisies (*Melampodium* sect. *Melampodium*; *Asteraceae*). Evolution 66, 211–228. doi: 10.1111/j.1558-5646.2011.01424.x

Weiss-Schneeweiss, H., Leitch, A. R., McCann, J., Jang, T.-S., and Macas, J. (2015). "Employing next generation sequencing to explore the repeat landscape of the plant genome," in *Next Generation Sequencing in Plant Systematics, Regnum Vegetabile 157*, eds E. Hörandl and M. Appelhans (Königstein: Koeltz Scientific Books), 155–180.

Weiss-Schneeweiss, H., Stuessy, T. F., and Villaseñor, J. L. (2009). Chromosome numbers, karyotypes, and evolution in *Melampodium* (*Asteraceae*). *Int. J. Plant Sci.* 170, 1168–1182. doi: 10.1086/605876

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nature Rev. Genet.* 8, 973–982. doi: 10.1038/nrg 2165

Zozomová-Lihová, J., Mandáková, T., Kovaříková, A., Mühlhausen, A., Mummenhoff, K., Lysak, M. A., et al. (2014). When fathers are instant losers: homogenization of rDNA loci in recently formed *Cardamine schulzii* trigenomic allopolyploid. *New Phytol.* 203, 1096–1108. doi: 10.1111/nph.12873

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 2020
doi: 10.3389/fpls.2020.00320

Divide to Conquer: Evolutionary History of Allioideae Tribes (Amaryllidaceae) Is Linked to Distinct Trends of Karyotype Evolution

*Lucas Costa[1], Horace Jimenez[2], Reginaldo Carvalho[2], Jefferson Carvalho-Sobrinho[2], Inelia Escobar[3] and Gustavo Souza[1]**

[1] Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Federal University of Pernambuco, Recife, Brazil,
[2] Laboratory of Plant Cytogenetics, Department of Biology, Federal Rural University of Pernambuco, Recife, Brazil,
[3] Department of Botany, University of Concepción, Concepción, Chile

Allioideae (e.g., chives, garlics, onions) comprises three mainly temperate tribes: Allieae (800 species from the northern hemisphere), Gilliesieae (80 South American species), and Tulbaghieae (26 Southern African species). We reconstructed the phylogeny of Allioideae (190 species plus 257 species from Agapanthoideae and Amaryllidoideae) based on ITS, *mat*K, *ndh*F, and *rbc*L to investigate its historical biogeography and karyotype evolution using newly generated cytomolecular data for Chilean Gilliesieae genera *Gethyum*, *Miersia*, *Solaria*, and *Speea*. The crown group of Allioideae diversified ∼62 Mya supporting a Gondwanic origin for the subfamily and vicariance as the cause of the intercontinental disjunction of the tribes. Our results support the hypothesis of the Indian tectonic plate carrying Allieae to northern hemisphere ('out-of-India' hypothesis). The colonization of the northern hemisphere (∼30 Mya) is correlated with a higher diversification rate in *Allium* associated to stable $x = 8$, increase of polyploidy and the geographic expansion in Europe and North America. Tulbaghieae presented $x = 6$, but with numerical stability ($2n = 12$). In contrast, the tribe Gilliesieae ($x = 6$) varied considerably in genome size (associated with Robertsonian translocations), rDNA sites distribution and chromosome number. Our data indicate that evolutionary history of Allioideae tribes is linked to distinct trends of karyotype evolution.

Keywords: Amaryllidaceae, BioGeoBEARS, biogeography, cytogenetics, rDNA sites, genome size, phylogenetic comparative methods (PCMs)

OPEN ACCESS

Edited by:
Martin A. Lysak,
Masaryk University, Czechia

Reviewed by:
Nikolai Friesen,
University of Osnabrück, Germany
Sònia Garcia,
Spanish National Research Council, Spain
Lorenzo Peruzzi,
University of Pisa, Italy

****Correspondence:***
Gustavo Souza
lgrsouza@hotmail.com

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

Received: *15 November 2019*
Accepted: *04 March 2020*
Published: *07 April 2020*

Citation:
Costa L, Jimenez H, Carvalho R, Carvalho-Sobrinho J, Escobar I and Souza G (2020) Divide to Conquer: Evolutionary History of Allioideae Tribes (Amaryllidaceae) Is Linked to Distinct Trends of Karyotype Evolution. Front. Plant Sci. 11:320. doi: 10.3389/fpls.2020.00320

INTRODUCTION

The search for the causes of species geographic distributions is notable for its lack of universally applicable rules (Stebbins, 1971; Lowry and Lester, 2006; Araújo et al., 2019; Liu et al., 2019). The current species distribution reflects their dispersal ability, environmental tolerance, niche breadth, population abundance, colonization–extinction dynamics, and character diversity (Brown et al., 1996; Lowry and Lester, 2006). Geographic distribution may impact evolution; however, in the case of karyotype data (e.g., number and morphology of chromosomes, ploidy level correlated with genome size, number of ribosomal DNA [rDNA] sites, genome size) the opposite also occurs and

chromosomal changes may lead to ability to colonize new environments (e.g., allopolyploids; Souza et al., 2012). In this context, cytogeography emerged as the analysis of the geographical distributions of polymorphic cytological markers, as polyploidy, inversions, Robertsonian translocations, increase/decrease of rDNA site number, etc. (Colombo and Confalonieri, 2004). Analyses of spatial distribution of karyotypes may indicate an adaptive value for certain types of chromosomal rearrangements and help to clarify processes that contributed to shape particular distribution patterns (Colombo and Confalonieri, 2004; Raskina et al., 2004; Van-Lume et al., 2017).

The classical cytogeographic analyses implemented by plotting cytotypes on maps now is furthered by incorporating modern approaches in a time-space interface by phylogenetic comparative methods (PCMs). Although PCMs are widely used in ecology (Legendre and Legendre, 2012) and cytogenetics (Glick and Mayrose, 2014; Kolano et al., 2015; Van-Lume et al., 2017; Carta et al., 2018; Serbin et al., 2019), few papers have demonstrated their applicability in a geographic perspective (Carta and Peruzzi, 2016; Souza et al., 2019a). Thus, biogeographic analyses of ancestral area reconstruction (Matzke, 2012) using dated molecular phylogenies along with diversification rate analyses (see Menezes et al., 2017), may help to elucidate processes associated with karyotype diversification. This may be especially interesting in plant groups with ancient origin, intercontinental disjunct distribution and marked cytogenetic variability.

The subfamily Allioideae of Amaryllidaceae (sensu Angiosperm Phylogeny Group [APG], 2016) is an excellent model to assess the links between karyotypes and environmental variables by presenting large chromosomes with remarkable cytogenetic variability (Vosa, 2000; Souza et al., 2016; Pellicer et al., 2017; Peruzzi et al., 2017; Sassone et al., 2018), associated with an intriguing discontinuous geographical distribution (**Figure 1**). Allioideae is phylogenetically composed of three tribes (Angiosperm Phylogeny Group [APG], 2016): Allieae, Gilliesieae, and Tulbaghieae, which are exclusively distributed in North Hemisphere, South America (Peru to Chile), and Southern Africa, respectively (**Figure 1**). The subfamily is composed of rhizomatous or bulbous geophytes, widely known for ornamental, medicinal or food use (e.g., chives, garlics, onions) (Pellicer et al., 2017). Allieae is represented by only *Allium* L. with ∼800 species (Friesen et al., 2006; Li et al., 2010; Hauenschild et al., 2017). South American tribe Gilliesieae with 80 species have traditionally been classified into two main groups, namely subtribes Gilliesiinae and Leucocoryninae, characterized by zygomorphic and actinomorphic flowers, respectively (Rudall et al., 2002; Chase et al., 2009; Escobar, 2012). Tulbaghieae is composed only of the monotypic *Tulbaghia* L. which comprises 26 Southern African species (Vosa, 2000; Chase et al., 2009; Stafford et al., 2016). A second genus, *Prototulbaghia*, has been proposed (Vosa, 2007), but it is deeply nested within the *Tulbaghia* clade (Stafford et al., 2016).

Most of the Allioideae genera can be distinguished by distinct and easily recognized karyological characteristics, making a cytotaxonomic approach particularly useful in this subfamily. They may have bimodal karyotypes (especially South American species), with large metacentric (M) and acrocentric (A) chromosomes, which are considered to be classical examples of karyotype evolution by Robertsonian translocations (RTs) or centric fusions/fissions (Jones, 1998). In general Allieae and Tulbaghieae are characterized by stability of the basic chromosome number, $x = 8$ and $x = 6$, respectively [except for few *Allium* species of the subgenera *Amerallium* Traub and *Melanocrommyum* (Webb & Berthel.) Rouy which present different chromosome numbers] (Peruzzi et al., 2017). Conversely, the South American Gilliesieae present the highest karyotype variability in Allioideae: it is represented by both different basic chromosome numbers ($x = 4$, 5, 6, 7, and 10) and chromosome numbers ($2n = 8$, 10, 12, 14, 15, 16, 18, 19, 20, 24, 26, and 32) resulted from intense Robertsonian translocation and polyploidy events (Escobar, 2012; Souza et al., 2016; Pellicer et al., 2017; Sassone et al., 2018). In addition, analyses of heterochromatic bands and distribution of 5S/35S rDNA sites by fluorescence *in situ* hybridization (FISH) have revealed extensive cytomolecular variability in the subfamily (Souza et al., 2016, 2019b). The different patterns of karyotype evolution and species-richness (i.e., high karyotype variation in low diversified lineages or stable karyotypes in high diversified groups) of Allioideae tribes (**Figure 1**) represent an interesting case study for the implications of intercontinental disjunction.

The current distribution of the three tribes of Allioideae indicates a few possible scenarios to explain its intercontinental disjunction based on the timing and place of their separation. A late Cretaceous split coupled with a Gondwanic origin would imply separation via continental drift, with an "Out-of-India" arrival of tribe Allieae in the Northern hemisphere (Briggs, 2003; Bossuyt et al., 2006; Datta-Roy and Karanth, 2009). A Laurasian origin following a mid-Eocene expansion through the Northern Hemisphere via the Boreotropical belt with late arrival in Africa and South America could also explain these disjunctions, as is the case with a number of angiosperms (Meng et al., 2014; Wei et al., 2015). Nevertheless, long distance dispersal has also often been appointed as a mechanism for intercontinental disjunction, especially in case of relatively young splits (Thiv et al., 2011; Crayn et al., 2014). In addition, if the two split events that formed the three Allioideae tribes are significantly apart, a combination of more than one of these processes could also be a likely scenario (Janssens et al., 2016; Yang et al., 2018). The reconstruction of the biogeographic history of Allioideae using a dated phylogeny may provide an adequate framework to discuss these different hypotheses and shed light on how the disjunction may have contributed to shape karyotype and species-richness patterns in the group.

We reconstructed the phylogeny of Amaryllidaceae, with a focus in Allioideae to assess its biogeographic and karyotype patterns. Additionally, we newly generate cytomolecular data (number and morphology of chromosomes and distribution of 5S and 35S rDNA sites) for Chilean genera of Gilliesieae: *Gethyum* Phil., *Miersia* Lindl., *Solaria* Phil., and *Speea* Loes. A molecular clock analysis was implemented to assess biogeographic hypotheses that might explain the disjunctions of the three Allioideae tribes. We specifically addressed three questions: (1) Why was karyotype evolution of Gilliesieae so variable

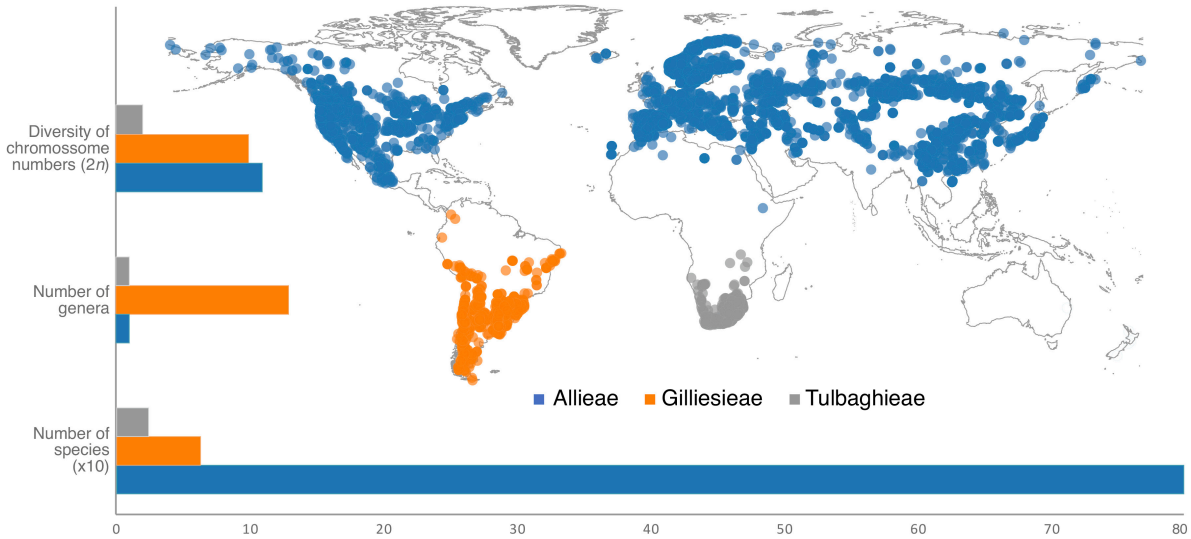

FIGURE 1 | Geographic distribution and diversification patterns of the three tribes of the subfamily Allioideae (Amaryllidaceae). Geographic distribution was obtained from http://www.gbif.org.

(in terms of chromosome numbers and number/position of rDNA sites) compared to the other tribes of Allioideae? (2) Is Allioideae a Gondwanan or Laurasian group? (3) Can historical biogeographical events be related to the distinct karyotypic patterns of the Allioideae tribes?

MATERIALS AND METHODS

Taxa Sampling

Karyotype and molecular data were collected from representatives of the subfamily Allioideae (Amaryllidaceae). Comparative cytogenetic analysis sampled 451 species from the three subfamilies of Amaryllidaceae: Amaryllidoideae (257 spp.) Agapanthoideae (4 spp.), and Allioideae [including the tribes Allieae (150 spp.), Gilliesieae (37 spp.), and Tulbaghieae (3 spp.)] (**Supplementary Table S1**). Newly generated cytomolecular data for seven species of Gilliesieae were included: *Gethyum atropurpureum* Phil., *Gethyum cuspidatum* (Harv. ex Baker) Muñoz-Schick, *Gilliesia graminea* Lindl., *Gilliesia montana* Poepp. & Endl., *Miersia chilensis* Lindl., *Solaria miersioides* Phil., and *Speea humilis* Loes. In addition, new genome size estimates were undertaken for genera *Ipheion* Raf., *Leucocoryne* Lindl., *Nothoscordum* Kunth, and *Zoellnerallium* Crosa [vouchers and collection locations presented in Souza et al. (2010, 2016)]. Collection locations, voucher numbers and karyotype data are presented in **Table 1**. The vouchers specimens were deposited at the herbarium CONC of the University of Concepción (Chile).

Cytogenetic Analysis

Root tips from bulbs were pretreated with 0.05% colchicine for 24 h at 10°C, fixed in ethanol:acetic acid (3:1; v/v) for 2–24 h at room temperature and stored at −20°C. The fixed root samples were washed in distilled water and digested in 2% cellulase (Onozuka) and 20% pectinase (Sigma) at 37°C for 90 min. Subsequently, the apical meristem was squashed in 45% acetic acid under a coverslip. The material was fixed to the slide by deep freezing with liquid nitrogen, and the coverslip was removed perhaps with a razor blade. The FISH technique was used to located 5S and 35S rDNA sites following Souza et al. (2019b). The slides were mounted with DAPI (4 μg mL^{-1}) diluted in Vectashield (Vector) 1:1 (v/v) and analyzed under an epifluorescence microscope (Leica DMLB). Images were recorded using a Cohu CCD camera and software Leica QFISH before editing with the software Adobe Photoshop CS3 v.10.0.

Flow Cytometry

Absolute nuclear DNA contents were determined by flow cytometry according to Doležel et al. (2007). Fresh leaves from the specimens were collected to prepare the samples of 25–50 mg each. The material was chopped together with fresh leaf tissue of the internal standard (*Vicia faba* L. subsp. *faba* 'Inovec' 2C = 26.9 pg/2C DNA; Doležel et al., 1992) with a razor blade on a Petri dish (kept on ice) containing 1 mL of WPB isolation buffer (Loureiro et al., 2007). The solution was filtered through a 30 μm mesh filter and mixed with 50 μg/mL of propidium iodide (1 mg/mL).

Flow cytometry measurements were taken using a Partec Cyflow Space (Müster, Germany) equipped with a 488 nm laser canon. The relative fluorescence histograms were analyzed on FloMax program version 2.3. The coefficient of variation of obtained peaks was assessed at half of the peak height (H.P.C.V.), discarding peaks with a H.P.C.V. > 5%. The genome size (ρg) of the samples were calculated using the following equation: "sample DNA = (sample G1/standard G1) × standard DNA," where sample G1 is the peak position

TABLE 1 | Original data from cytogenetically analyzed species of the tribe Gilliesieae with voucher number, collect location, haploid chromosome number (*n*), Karyotypic formulae, Fundamental number (FN), and number of 5S and 35S rDNA sites number.

						rDNA sites	
Species	**Voucher**	**Collect location**	***2n***	**Karyotypic formulae**	**FN**	**5S**	**35S**
Speea humilis Loes	CONC 30	Parque Nacional La Campana, Región de Valparaíso, Chile	6	4M + 1SM + 1A	11	2	2
Gethyum cuspidatum (Harv. ex Baker) Muñoz Schick	CONC 12	Parque Nacional Fray Jorge, Región de Coquimbo	7	4M + 1SM + 1A	11	2	6
Gilliesia graminea Lindl	CONC 18	Parque Nacional La Campana, Región de Valparaíso, Chile	7	2M + 2SM + 3A	11	2	6
Gilliesia montana Poepp & Endl	CONC 56	Reserva Los Ruiles de Empedrado, Región del Maule, Chile	7	2M + 2SM + 3A	11	2	6
Miersia chilensis Lindl	CONC 27	Parque Nacional La Campana, Región de Valparaíso, Chile	6	4M + 1SM + 1A	11	2	4
	CONC 95	Alhué, Región Metropolitana, Chile	10	1SM + 9A	11	6	26
Solaria atropurpurea (Phil.) Ravenna	CONC 4160	Quebrada Nido de Águila, Región Metropolitana, Chile	7	4M + 1SM + 1A	11	2	6
Solaria miersioides Phil	CONC 137	Los Álamos-Cipreses, Región del Maule, Chile	7	2M + 2SM + 3A	11	2	6
Speea humilis Loes	CONC 30	Parque Nacional La Campana, Región de Valparaíso, Chile	6	4M + 1SM + 1A	11	2	2

M, metacentric; SM, Submetacentric; A, acrocentric.

(G1) of the sample; standard G1 is the peak position (G1) of the standard, and standard DNA is the nuclear DNA (ρg) of the standard used in each measure. Three independent DNA estimations were performed on different days for each sample. Measurements were exhausted with at least 1,500 events per fluorescence peak.

Cytogenetic Data Survey

We surveyed additional data for chromosome number, genome size and 35S and 5S rDNA sites for species of Allioideae, Agapanthoideae and Amaryllidoideae (Plant rDNA database[1]). The chromosome number of 448 species was obtained from the Chromosome Count Database v. 1.46[2] (Rice et al., 2015) (**Table 1** and **Supplementary Table S1**). The ratio of chromosome arms (AR = the long arm length/short arm length) was used to classify the chromosomes as metacentric (M; AR = 1–1.4), submetacentric (SM; AR = 1.5–2.9), or acrocentric (A; AR > 3.0), following Guerra (1986).

For genome size, we compiled data for 189 species from the Kew Gardens C-Value Database v. 6.0[3] (Bennett and Leitch, 2012) along with 62 records from literature that were not in the database and 13 new estimates (**Table 1** and **Supplementary Table S1**). We plotted genome sizes and chromosome numbers for Allioideae in dispersion plots using the package *stats* of the software R (R Core Team, 2013) (see **Figure 2**).

We also surveyed number and position of 35S and 5S rDNA sites for 55 and 59 species, respectively (**Supplementary Table S1**). Metaphasis pictures and/or original idiograms with scale information (when available) were used to construct a simplified idiogram based on Lima-de-Faria (1976) containing only the site-bearing chromosomes of the species surveyed. All chromosomes measurements were made using Adobe Photoshop CS3 and the idiogram was drawn on CorelDraw X7.

Phylogenetic Analyses

To provide a robust phylogenetic framework for the subsequent analyses, we reconstructed a phylogenetic tree sampling taxa from the three subfamilies of Amaryllidaceae, of which 190 were Allioideae species and 261 species were from Agapanthoideae (4 spp.) and Amaryllidoideae (257 spp.). *Aloe vera* (L.) Burm.f. (Xanthorrhoeaceae) was used as outgroup (**Supplementary Table S1**). We used available data for one nuclear (ITS) and three plastids (*mat*K, *ndh*F, and *rbc*L) loci from GenBank (see accession numbers in **Table 1** and **Supplementary Table S1**), with each species having at least one locus sampled. Missing data were coded as gaps and accounted for 20.5% of the matrix.

An aligned matrix including data from the four markers was obtained using MUSCLE as a plugin implemented in Geneious v.7.1.9 (Kearse et al., 2012). We used jModelTest v.2.1.6 to assess the best-fit model of DNA substitution for each marker (Darriba et al., 2012) through the Akaike Information Criterion (AIC; Akaike, 1974). The selected models were SYM + I + G for ITS, HKY + G for *rbc*L, and GTR + G for *mat*K and *ndh*F.

Phylogenetic relationships were inferred using Bayesian Inference (BI) implemented in MrBayes v.3.2.6 (Ronquist et al., 2012). The analyses were performed on the combined data set, specifying the substitution model for each marker. Four independent runs with four Markov Chain Monte Carlo (MCMC) runs were conducted, sampling every 1,000 generations for 100,000,000 generations. Each run was evaluated in TRACER

[1]http://www.plantrdnadatabase.com

[2]http://ccdb.tau.ac.il/

[3]https://cvalues.science.kew.org/

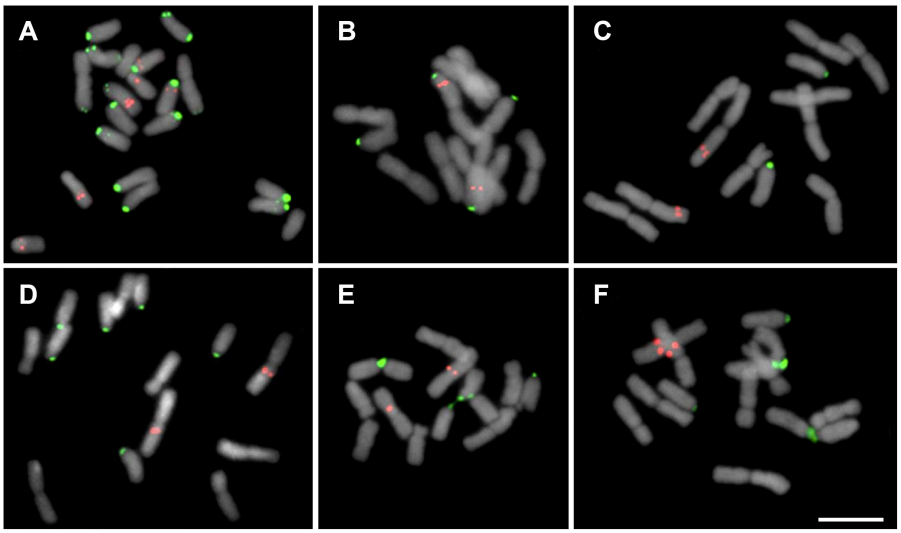

FIGURE 2 | Mitotic cells showing the distribution of 5S (red) and 35S (green) rDNA sites in Chilean Gilliesieae species. **(A,B)** Individuals of *Miersia chilensis* com $2n = 20$ (2SM + 18A) **(A)** and $2n = 12$ (8M + 2SM + 2A) **(B)**. **(C)** *Speea humilis* $2n = 12$ (8M + 2SM + 2A). **(D–F)** Species with $2n = 14$ (4M + 4SM + 6A): *Gilliesia graminea* **(D)**, *Gethyum atropurpureum* **(E)**, and *Solaria miersioides* **(F)**. Scale bar in **(F)** = 10 μm.

v.1.6 (Rambaut et al., 2018) to determine that the estimated sample size (ESS) for each relevant parameter was higher than 200 and a burn-in of 25% was applied. We then obtained the consensus phylogeny and clade posterior probabilities with the "sumt" command (contype = allcompat). The tree was visualized and edited in FigTree v.1.4.2 (Rambaut, 2014). All BI and jModelTest analyses were performed on the CIPRES Science Gateway (Miller et al., 2015).

Divergence Time Estimates

A molecular clock analysis was performed to explore the karyotype and biogeographic evolution in Allioideae. Divergence times were estimated on BEAST v.1.8.3 (Drummond and Rambaut, 2007) through CIPRES Science Gateway fixing the tree topology from the BI. Uncorrelated relaxed lognormal clock (Drummond and Rambaut, 2007) and Birth-Death speciation model (Gernhard, 2008) were applied. Two independent runs of 100,000,000 generations were performed, sampling every 10,000 generations. After removing 25% of samples as burn-in, the independent runs were combined and a maximum clade credibility (MCC) tree was constructed using TreeAnnotator v.1.8.2 (Rambaut and Drummond, 2013). In order to verify the effective sampling of all parameters and assess convergence of independent chains, we examined their posterior distributions in TRACER. The MCMC sampling was considered sufficient at effective sampling sizes (ESS) equal to or higher than 200.

The phylogeny was dated using one macrofossil of Amaryllidaceae and one secondary calibration from published dated phylogenies. The first fossil calibration was based on a fossil leaf attributed to Amaryllidaceae from the Cerrejón Formation, Colombia, estimated at 58 Mya (Wing et al., 2009). This fossil calibration was used to set the minimum age for Amaryllidaceae diversification. The second calibration point was based on a *rbc*L phylogeny of 554 angiosperms genera, which estimated the Amaryllidaceae crown group to be 59.6 Mya (BEAST analysis; Hermant et al., 2012). From the age variability suggested for the secondary calibration (33–59.6 Mya) we followed Hermant et al. (2012) due to their broadly sampled analysis with an age estimate more consistent with the fossil age.

Ancestral Range and Biogeographic Events Estimation

To investigate the historic biogeography of Amaryllidaceae, we employed a model-based likelihood approach implemented in the R package BioGeoBEARS (Matzke, 2013, 2014). The sampled species from the MCC tree yielded by BEAST were coded as present or absent in nine discrete areas around the globe: Africa, Andean region, Asia (India, west and east Asia), Europe + Siberia, Mediterranean region, Mesoamerica (Mexico + Central America), North America (minus Mexico), Oceania, and South America (minus Andean region). The regions were coded based on a search on the Global Biodiversity Information Facility website[4]. For ancestral range estimation, we used the MCC tree to test likelihood implementations of three different biogeographic models in BioGeoBEARS. In order to better reflect geological events through time, we stratified the tree in five time periods based on important events: (i) 70 to 55 Mya: period when the Indian subcontinent is completely separated of the African continent and migrates to the Laurasia, which is still in early separation; (ii) 55 to 40 Mya: Both Gondwana and Laurasia completed separation in two smaller land masses, the Indian subcontinent is completely

[4]gbif.org

connected with Asia and the Himalayans and Andean uplifts are in motion; (iii) 40 to 18 Mya: The Andes and the Himalayans continue to grow and change the South American and Asian landscape, Africa and Europe are closer each other; (iv) 18 to 3 Mya: Africa and Europe are connected by the Gibraltar strait and the Mediterranean region is fully established with its own characteristic climate; (v) 3 Mya to the present: The Isthmus of Panama connects the Americas. Both the dispersal probability and connection between areas in each of these periods was adjusted accordingly. Based on this, we informed different dispersion probabilities between areas for each time slice, following Buerki et al. (2011): low dispersal = 0.01; medium dispersal = 0.5; high dispersal (including areas adjacent or very close) = 1.0. We compared the results of models with and without the parameter j using likelihood ratio tests and the model weights were calculated under the AIC. To illustrate the geological state of the earth in different time periods, we generated paleomaps using the web tool available at http://www.odsn.de/odsn/services/paleomap/paleomap.html and edited the maps with the software CorelDraw X7.

Ancestral Character Reconstruction

Literature and newly generated chromosome number data were used to reconstruct the chromosome number evolution of the family along the MCC tree. To assess the events and processes (for instance RTs) that may have fostered the karyotype diversity across Amaryllidaceae phylogeny, we employed a statistical framework. We applied ChromEvol to test whether karyotypes evolved by polyploidy or dysploidy (Glick and Mayrose, 2014). The best fitting model was assessed using the AIC (Glick and Mayrose, 2014). The best fitted model was used to reconstruct the chromosome number along the MCC tree of Amaryllidaceae and two simplified trees with key events and ancestral numbers were drawn on CorelDraw X7.

In order to investigate the mode of evolution for chromosome number in each Allioideae tribe, we used the function *fitContiuous* implemented in the R package *geiger* (Harmon et al., 2008). Individual trees for each lineage (Gilliesieae, Tulbaghieae, *Allium* I, II, and III) were obtained by pruning the MCC tree with the function *drop.tip* implemented in the R package *phytools* (Revell, 2012). We fitted nine different likelihood models of continuous character evolution for each lineage and compared the results using AIC: (i) Brownian motion model - BM (Felsenstein, 1973); (ii) Ornstein-Uhlenbeck model - OU (Butler and King, 2004); (iii) Early-burst model - EB (Harmon et al., 2010); (iv) trend model; (v) lambda model (Pagel, 1999); (vi) kappa model (Pagel, 1999); (vii) delta model; (viii) drift model; (ix) white model (for details[5]).

We also reconstructed the 35S rDNA sites number as a continuous character along a simplified phylogeny (**Figure 1B**). Species without information for this site were pruned off from the tree with the function *drop.tip* implemented in the package *phytools* on R and the ancestral reconstruction was made with the function *cont.map*, also in *phytools*.

Diversification Rate Analysis

Shifts in diversification rates were calculated using speciation/extinction model type analysis in BAMM (Rabosky et al., 2014). To work with incomplete phylogenetic datasets in BAMM, it is necessary to input the percentage of sampled species for each major clade. This percentage was estimated according to the total number of accepted names reported for each subtribe (WFO, 2019). For this, tribe Allieae was divided in the three different lineages proposed by Friesen et al. (2006). Percentage of sampled species per tribe as informed on BAMMtools is presented in the **Supplementary Table S2**. Priors for the BAMM control file were generated using the dated phylogenetic tree input into the function *set BAMM priors* in the package BAMMtools v. 2.5.0 implemented in R. The control file was set for 10,000,000 generations and the analysis was run twice as recommended, returning similar results. Resulting MCMC log likelihoods were tested against generation number using the *CODA* package (Plummer et al., 2006) implemented in R. All remaining outputs contained in the event data file were analyzed using BAMMtools. BAMMtools was then used to produce a figure showing the best rate shift configuration as well as graphics of diversification through time for Gilliesieae + Tulbaghieae and each of the three evolutionary lineages of Allieae.

RESULTS

Cytomolecular Characterization of Chilean Gilliesieae

The analysis of the seven Chilean Gilliesieae species revealed large (5.5–14.9 μm) chromosomes which were metacentric (M) or submetacentric (SM) or acrocentric (A). Three distinct chromosome complements with the same number of chromosome arms or fundamental number NF = 11 were observed: $2n = 12$ (8M + 2SM + 2A), $2n = 14$ (4M + 4SM + 6A), and $2n = 20$ (2SM + 18A). FISH with the 35S rDNA probe revealed signals on the short arm of the acrocentric chromosomes (**Figure 1**).

Two different cytotypes were observed in *Miersia chilensis*. The individual collected in the municipality of Santiago (Chile) presented $2n = 20$ (2SM + 18A) with a large 5S rDNA site in the interstitial region of one acrocentric pair and a smaller extra 5S rDNA site in the proximal region of two additional acrocentric pairs (**Figure 1A**). However, the individual collected in the municipality of Valparaiso (Chile) presented $2n = 12$ (8M + 2SM + 2A) with 5S rDNA sites in the interstitial region of the long arm of the largest metacentric pair (**Figure 1B**), a karyotype similar to that observed in *Speea humilis* (**Figure 1C**). *Gethyum atropurpureum*, *Gethyum cuspidatum*, *Gilliesia graminea*, *Gilliesia montana*, and *Solaria miersioides* presented $2n = 14$ (4M + 4SM + 6A) and very similar karyotypes with 5S rDNA sites near the centromere of a pair of metacentric chromosomes (**Figures 1D–F**).

[5]http://www.webpages.uidaho.edu/lukeh/software.html

Cytogenetic Variability of Amaryllidaceae

The haploid chromosome number (*n*) varied in Amaryllidaceae from *n* = 4 on *Nothoscordum pulchellum*, *Tristagma bivalve* and *Tristagma graminifolium*, to *n* = 68 in *Eucharis amazonica* (**Supplementary Table S1** and **Figure 3**). Gilliesieae was represented by only 36 records and ten different chromosome numbers (2*n*). On the other hand, 148 records were obtained to Allieae (11 different 2*n*, most of which polyploid series). For Tulbaghieae only three different 2*n* were recovered, with stability 2*n* = 12.

For genome size we accomplished 13 new estimates for species of Gilliesieae (**Supplementary Table S1**). In the subtribe Gilliesiinae, both *Speea humilis* and *Gilliesia gramina* presented large genome sizes (2C = 56.65 and 36.04 pg, respectively) as observed in most of Allioideae. The estimates for *Ipheion* presented two of the smallest genome sizes in the subfamily [*I. recurvifolium* (2C = 18.1 pg) and *I. uniflorum* (2C = 18.06 pg)]. The estimates for genera *Leucocoryne* and *Nothoscordum* showed large genomes as frequently observed for these two genera (**Supplementary Table S1**). For the subsequent analysis, the monoploid genome size value (1C*x*) was obtained by dividing the 2C value by the ploidy level, also informed on the 2C value database. The 1C*x* value varied from 1C*x* = 4.52 in *Ipheion uniflorum* to 1C*x* = 65.45 in *Sprekelia formosissima*. In this case, the genome size varied 6.3-fold among 25 records of Gilliesieae, whereas in Allieae the variation was of 4.5-fold among 126 records (**Supplementary Table S1** and **Figure 3**).

According to our Chromevol analysis, all these three events were almost equally important to karyotype evolution (*f* = 93 for chromosome gains, 83.1 for chromosome losses, and 86.5 for duplications). The ancestral haploid number for Amaryllidaceae was *n* = 8 (pp = 0.50), with the loss of one chromosome originating an ancestral haploid number *n* = 7 (pp = 0.45) for Allioideae (**Figure 4**). The Allieae was characterized by gain of one chromosome (*n* = 8, pp = 0.61), while Tulbaghieae and Gilliesieae derived from a shared ancestor with *n* = 6 (pp = 0.94). The MRCA of Tulbaghieae retained *n* = 6 (pp = 0.99). The MRCA of the Chilean Gilliesieae also retained *n* = 6, though with a high variability despite its fewer number of extant taxa (**Figure 1**). The other Gilliesieae species were marked by the loss of one chromosome in its MRCA (*n* = 5, pp = 0.56) and also a high number of karyotype events in more recent splits. Among the lineages of *Allium*, only the first presented a change on the haploid chromosome number of the ancestral node, showing *n* = 7 (pp = 0.99), with this number being conserved on most of its taxa. Both the second and third taxa retained *n* = 8 (pp = 0.99), with the former being extremely conserved, while the latter presented high incidence of polyploidy. Our analysis of continuous character evolution revealed that chromosome number evolution likely followed distinct modes in each lineage (**Supplementary Table S3**). The best-fitted model for tribe Gilliesieae was a model based on Pagel's 'Lambda' (Pagel, 1999) which assumes that trait variation is associated with phylogenetic relatedness. For Tribe Tulbaghieae, the best-fitted model was based on the time-dependent parameter 'Delta' (Pagel, 1999). The delta model fits the relative contributions of early versus late evolution in the tree to the covariance of species trait values. On Tribe Allieae, chromosome evolution of *Allium* I and II was better explained by the 'White' model, which implies that trait variation has no phylogenetic meaning. Meanwhile, the best-fitted model for *Allium* III was the 'Ornstein-Uhlenbeck' model (Butler and King, 2004), which implies that trait variation fits a random walk toward an "evolutionary optimum" state in different lineages.

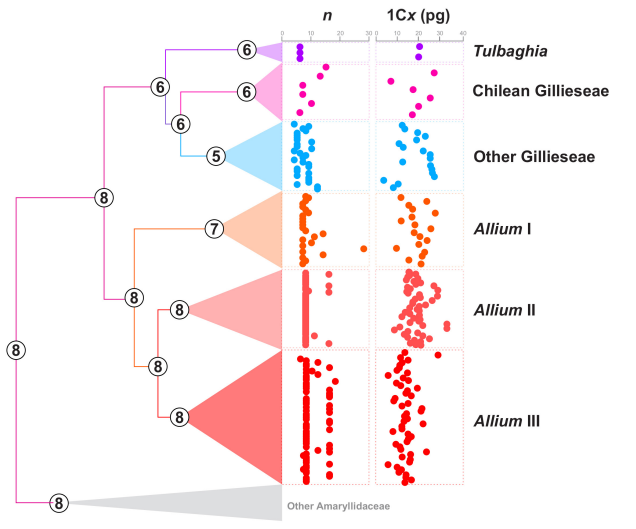

FIGURE 3 | Ancestral chromosome number reconstruction inferred with ChromEvol of Amaryllidaceae species. Numbers at nodes represent the most probable ancestral chromosome number (*n*). Dispersion plots at the right of the phylogeny represent the chromosome numbers and genome sizes ($1C_x$ values) of the different Allioideae groups.

For the rDNA survey, 60 records (including the seven original FISH results reported here) were observed for 35S rDNA ranging from 2 to 26 sites, while 64 records were observed for 5S sites number, varying from 2 to 16 (**Supplementary Table S1**). The 5S rDNA sites were scattered along different regions of the chromosomes. Duplicate sites in the interstitial region of the same chromosomal arm were consistently observed for the three subfamilies of Amaryllidaceae (**Figure 4**). In turn, the 35S rDNA sites showed a tendency to appear on the short arm of acrocentric chromosomes (**Figure 4**).

To assess the evolution of 35S rDNA, we treated it as a continuous character (**Figure 4**). Agapanthoideae, Amaryllidoideae, Allieae, and Tulbaghieae presented little variation on number of sites, with similar inferred ancestral numbers (four to five sites). In contrast, Gilliesieae had great variability on 35S site number and a noticeable increase was observed on the inferred ancestral of the tribe, presenting six to seven sites (**Figure 4**).

Phylogenetic Relationships and Historical Biogeography of Amaryllidaceae

The most recent common ancestors (MRCAs) of Amaryllidaceae, the three subfamilies and the three Allioideae tribes presented

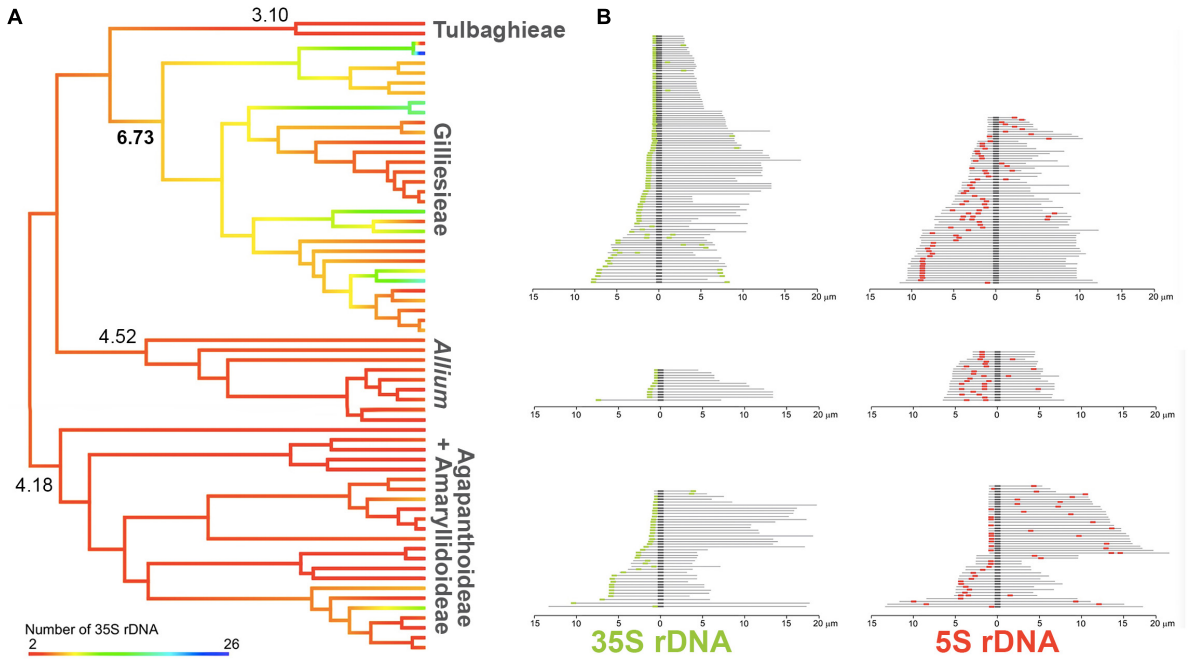

FIGURE 4 | Evolution of rDNA sites across Amaryllidaceae phylogeny. **(A)** Ancestral number of 35S rDNA sites reconstruction on a pruned phylogeny of Amaryllidaceae. Hot branch colors indicate higher number of sites while cold colors represent lower number of sites. **(B)** Schematic idiograms showing the position of 5S (red) and 35S (green) rDNA sites on the chromosomes of species of the subfamily Allioideae. Darker regions represent the centromere.

high support values (pp > 0.95) while most internal nodes presented moderate to low support. The DEC model with the addition of the free parameter j presented the most likely biogeographic scenario for the family (LnL = −723.55). According to our data, the crown node of Amaryllidaceae appeared 67.9 Mya (77.2–58.5 Mya: 95% HPD), with a probable Gondwanan stem age and distribution in parts of South America, Africa, and India (**Figure 5A**). From there, the three main subfamilies followed different evolutionary paths. The MRCA of Agapanthoideae and Amaryllidoideae diversified from Africa approximately 62.7 Mya (68.4–55.8 Mya: 95% HPD), with the former remaining in Africa, while the latter colonized regions of Europe, Asia and South America. Allioideae splitted shortly after separating from the other subfamilies 63.2 Mya (67.5–53.7 Mya: 95% HPD) (**Figure 5B**). One of the lineages became Allieae 52.2 Mya (58.1–44.4 Mya: 95% HPD), rapidly colonizing parts of Asia and North America after arriving presumably via the Indian Subcontinent (**Figure 5C**). From there, Allieae colonized most regions of the Northern hemisphere (**Figure 5D**). The other lineage splitted 54.1 Mya (65.1–37.11 Mya: 95% HPD) into Tulbaghieae and Gilliesieae. Tulbaghieae did not expand from Africa while Gilliesieae diversified in South America, splitting further into the more widespread Gilliesiinae and the Chilean Andean Leucocorynae approximately 45 Mya (61.2–32.2 Mya: 95% HPD).

Diversification Rate Shifts

The 95% credible set of rate shift configurations yielded by BAMM showed seven possible shift configurations, always with one shift on a different early node of the third evolutionary lineage of *Allium*. For better visualization, a mean phylorate was obtained, showing a continuous increase on diversifications rate on this lineage (**Figure 6**). To further explore the difference in diversification rate, four different density plots of speciation through time were obtained. It was evidenced a steady increase on diversification in the third evolutionary lineage of *Allium* (III) since its origin circa 42 Mya. Meanwhile, Gilliesieae + Tulbaghieae and *Allium* I + II presented a more constant speciation rate through time (**Figure 6**).

DISCUSSION

Centric Fissions Are the Key Mechanism to Explain the High Karyotype Diversity in Gilliesieae

The maintenance of the number of chromosome arms accompanied by chromosomal number ($2n$) changes in Chilean Gilliesieae species analyzed here clearly indicates a karyotype evolution by Robertsonian translocations Escobar, 2012). This type of chromosomal rearrangement was also observed other genera of Gilliesieae (Crosa, 1972; Jones, 1998; Souza et al., 2010, 2012, 2015, 2016, 2019b), however, it is rarely reported in species of Allieae or Tulbaghieae (Vosa, 2000; Peruzzi et al., 2017). Interestingly, our data also suggests that chromosome number evolution is more associated with phylogenetic relatedness in Gilliesieae than in the other tribes (**Supplementary Table S1**), which reinforces the impact of

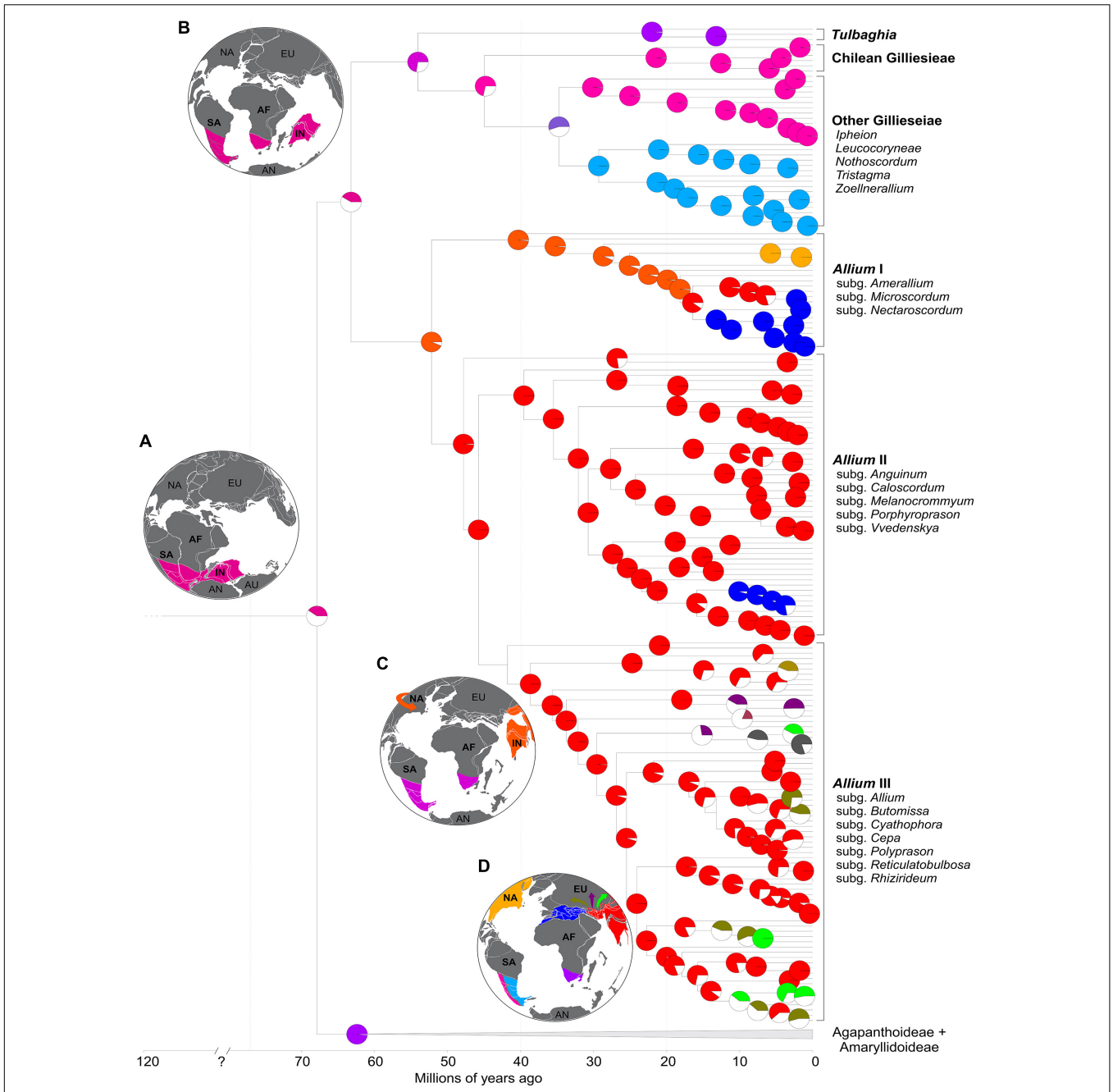

FIGURE 5 | Ancestral range reconstruction for Amaryllidaceae (under the DEC + J model) focusing on subfamily Allioideae. Pie charts at the nodes represent the probability of ancestral range. Colored slices represent the most probable ancestral range while white slices represent other possible ranges. Panels **(A–D)** are paleomaps representing the geological state of the Earth at 120, 65, 40, and 20 Mya, respectively.

chromosome number change for the evolution of this tribe. Morphoanatomic and phylogenetic analyzes suggest that *Miersia* and *Speea* are the first diverging lineages of Gilliesieae (Rudall et al., 2002; Escobar, 2012; Pellicer et al., 2017) corroborating our phylogenetic hypothesis (see **Figure 5**). Thus, the $2n = 12$ karyotype observed in these genera may represent a plesiomorphic condition for Gilliesieae, shared with *Thulbaghia* (Vosa, 2000).

Across the Allioideae subfamily karyotypes are predominantly formed by metacentric chromosomes as reported for *Allium* (Peruzzi et al., 2017), *Leucocoryne* (Souza et al., 2015), *Nothoscordum* (Souza et al., 2019b), and *Tristagma* (Crosa, 1981). Only *Ipheion* (three species; Souza et al., 2010) and *Zoellnerallium* (two species; Souza et al., 2016) have acrocentric-rich karyotypes, suggesting that this is a derived condition. In this sense, *Miersia chilensis* samples with $2n = 20$ and $2n = 21$

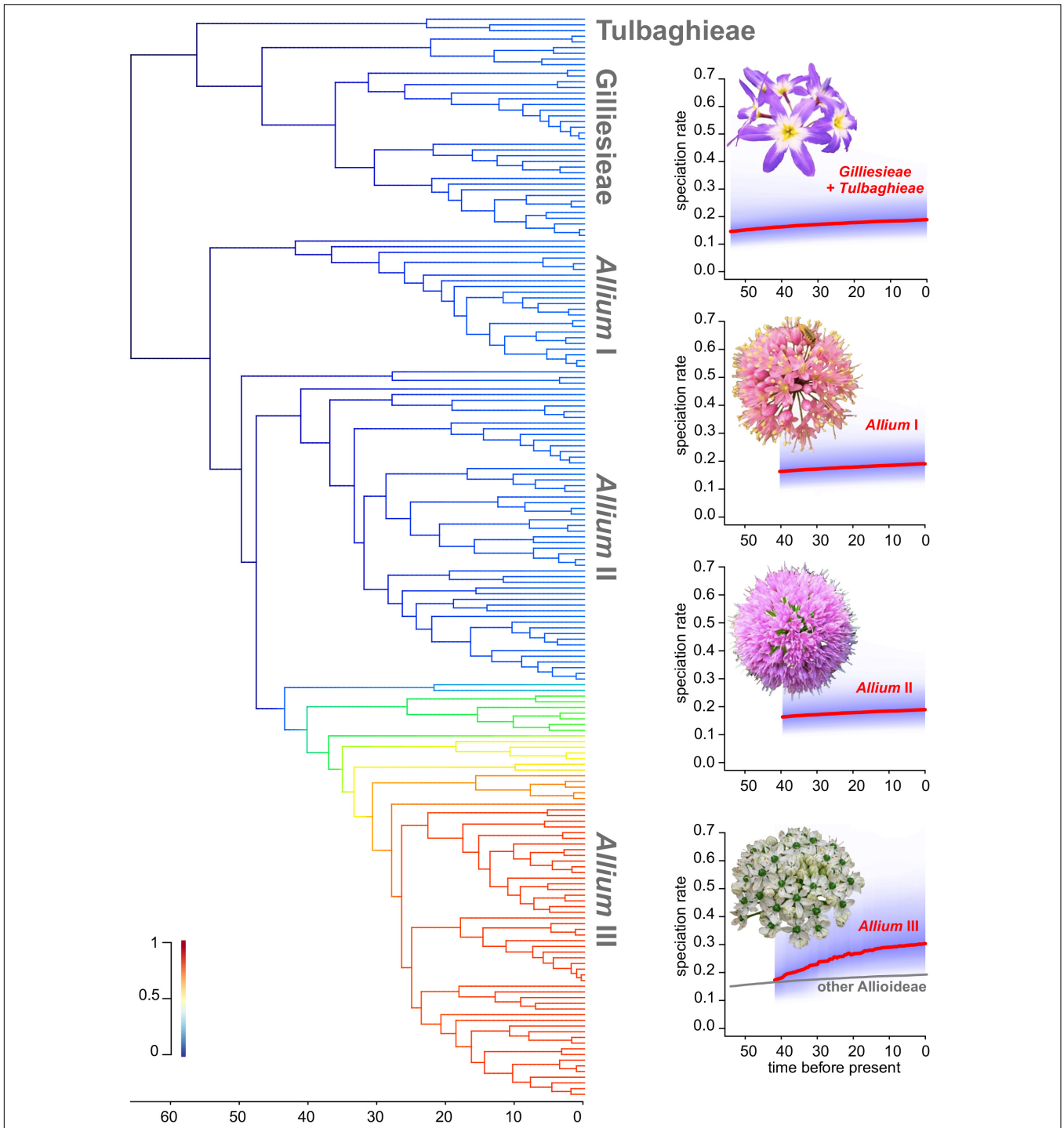

FIGURE 6 | Rate of diversification through time on the subfamily Allioideae. On the left, the dated phylogeny of the subfamily, with colored branches indicating the rate of diversification. Colder colors represent low rates, while hot colors represent high rates. On the right, plots of diversification rate through time for each Allioideae lineage. Strong red lines represent the consensus correlation, while purple shade represents the uncertainty of the analysis.

(Cave and Bradley, 1943) may represent recent events of multiple centric fissions. Pellicer et al. (2017) analyzing Gilliesieae species reported the impact of multiple Robertsonian translocations on the reductions in the overall genome size. However, it is unclear how often such centric fissions lead to genomic expansions or contractions (Pellicer et al., 2017). This explains the high variability in genome sizes observed here in South American Gilliesieae.

Interestingly, the Robertsonian translocations seem do not affect the number of 5S rDNA sites, with one site per monoploid assembly being observed in most species of the tribe (Souza et al., 2012, 2016). Conversely, the number and position of 35S rDNA sites were directly affected by Robertsonian translocations, with the formation of new 35S rDNA sites in the short arms of the acrocentric (see Hall and Parker, 1995). This trend led to an increase in the number of rDNA sites in Gilliesieae when compared to other tribes of Allioideae (see **Figure 3**). This correlation between centric fissions and increase in the number of 35S rDNA sites in the acrocentric short arms has also been reported in other plant genera (Tagashira and Kondo, 2001), mollusks (Pascoe et al., 1996) and insects (Nguyen et al., 2010), suggesting that this may be an inherent feature of the mechanism of centric fission in eukaryotes.

Historical Biogeography of Allioideae Support the Role of Indian Plate Carrying Allieae to Northern Hemisphere

Our data support that the current intercontinental disjunction of the Allioideae tribes (see **Figure 1**) may have been the result of vicariance after Gondwanan breakup (Givnish and Renner, 2004; Bartish et al., 2011). The fact that Allioideae are geophyte plants and usually without specialization for long-range dispersion and the old age of the phylogenetic splits that formed the three tribes reinforces the hypothesis of vicariance. This is corroborated by BioGeoBEARS analysis that revealed a predominance of vicariant events compared to few long-range dispersal events (concentrated in the *Allium* III clade). Our estimate Amaryllidaceae crown age 67.9 Mya (77.2–58.5 Mya: 95% HPD) suggests that the family may be much older than the secondary calibration previous estimates (∼33 Mya: Bremer, 2000; ∼50 Mya: Chen et al., 2013). This hypothesis of old-aged Amaryllidaceae is corroborated by their only macrofossil collected in the Cerrejón Formation, Colombia, estimated at 58 Mya (Wing et al., 2009) and by an extensive molecular clock analysis covering 800 monocots, which concluded that Allioideae has crown age 87 Mya and stem age 91 Mya (Janssen and Bremer, 2004). Specifically in *Allium*, the divergence times shown here (∼52 Mya) are also much older than reported in the literature (11 Mya to 34.25 Mya: Li et al., 2010; Chen et al., 2013; Hauenschild et al., 2017) based on secondary calibrations.

Our molecular clock and ancestral area reconstruction analysis, as well as the presence of the only Amaryllidaceae fossil in South America (Wing et al., 2009), suggest a Gondwanic origin of Allioideae. All Gondwanan breakup models suggest that the physical separation between Africa, South America, and India occurred sometime during the end of the Early Cretaceous or earliest Late Cretaceous (∼110–70 Mya). Our median crown age estimate for Amaryllidaceae is 67.89 Mya, with a variance from 77.26–58.51 Mya (95% HPD). Because the upper age estimate is situated within the 110–70 Mya range assumed for the Gondwanan continents (**Figure 5**), we cannot reject the hypothesis that the split 'Gilliesieae + Tulbaghieae → Allieae' and 'Gilliesieae → Tulbaghieae' resulted from the rifting of Africa, South America and India tectonic plates.

After the Gondwana breakup the Indian plate supposedly underwent a period of isolation [30–40 Mya] moving north, before colliding with the Eurasian plate around 40–50 Mya (Datta-Roy and Karanth, 2009). Consequently, the "Biotic ferry model" was proposed, according to which the rafting Indian plate carried ancient Gondwanan forms to Asia (Briggs, 2003; Bossuyt et al., 2006). After India collided with the Asian continent in the Early Tertiary, a few surviving Gondwanan elements dispersed out of India into South and Southeast Asia, which at the time lay in the same latitudinal and climatic zone (Morley, 2000). The out-of-India hypothesis adjusts to the arrival of *Allium* in the northern hemisphere in view of the age of the group (∼52 Mya compatible with the collision of the Indian and Eurasian plate) and by the center of origin of *Allium* in eastern Asia identified here. Friesen et al. (2006) identify also three main clades in *Allium*, with the first diverging lineages *Nectaroscordum* and *Microscodum* with center of origin Mediterranean and eastern Asia, respectively. Similarly, Li et al. (2010) proposed an *Allium* origin in eastern Asia (Northwest China), a geographic region of high species diversity for the genus, which corroborates our results. Interestingly, the genus *Allium* is not currently distributed in peninsular India (except Himalaya), a pattern similar to that seen in other 'out of Indian' groups (Datta-Roy and Karanth, 2009). It is argued that the extinction of these lines in India due to aridification and drastic climate change that occurred in India upon collision with Eurasia (Karanth, 2003).

Evolutionary History of Each Allioideae Lineage Impacts in Tribe-Specific Trends of Diversification

Our data suggests that the evolutionary history of each Allioideae lineage impacts in tribe-specific trends of diversification. The long time of origin, stable diversification rates, and relatively low number of species in Gilliesieae (80 species) and Tulbaghieae (26 species) may suggest that these are relictual lineages. In this sense, the scenario of multiple geomorphological changes in South America, mainly caused by the Andean uplift (Antonelli et al., 2009; Pennington et al., 2010), may have been responsible for phylogenetic, karyotype and morphological differentiation in Gilliesieae species (Rudall et al., 2002; Pellicer et al., 2017; Sassone et al., 2018). The origin of the Andes is related here to the isolation of the Chilean clade with strong differentiation in floral morphology (Rudall et al., 2002; Escobar, 2012) and karyotypes. On the other hand, southern Africa experienced a scenario of more recent tectonic stability (Hälbich, 1992), which might be related to low morphological and karyotype diversification of *Tulbaghia*.

Similarly, diversification patterns in *Allium* seem to reflect coherence between biogeography and karyotype evolution. Accordingly, our continuous character evolution analysis suggests very different modes of karyotype evolution for the three main tribes of Allioideae (**Supplementary Table S3**). The colonization of northern hemisphere should have favored a higher diversification rate in *Allium* (**Figure 6**) associated with increased polyploidy and territorial expansion to Europe and North America. This geographic expanding trend is especially

pronounced in the most recent *Allium* clade (III), which in light of the mid-Eocene date of the crown node could present an example of range expansion through the Boreotropical belt. Although the relationship between polyploidy and geographic expansion, especially colonization of new environments, is widely reported (Souza et al., 2012), the impact of genomic duplication on the rate of diversification is controversial, as some analyses have shown that the increase of polyploidy does not led necessarily to increased diversification rate (Sader et al., 2019). This suggests that colonization of the northern hemisphere by *Allium* was a complex and long-time process, accompanied by intense morphological diversification, which resulted in a species-rich genus with a complex taxonomic delimitation (Friesen et al., 2006).

CONCLUSION

Historical biogeographic analysis including ancestral area reconstruction, dated molecular phylogeny and diversification rate analysis, were used here to unravel the karyotypic evolution of Allioideae (Amaryllidaceae). Our data support that the current intercontinental disjunction between the three tribes of Allioideae may have been the result of vicariance due to Gondwanan breakup. The results point to the possibility that the Indian plate carried Allieae to northern hemisphere ('out-of-India' hypothesis). From there, the genus *Allium* diversified through polyploidy and geographic expansion in North Hemisphere. Interestingly, karyotype stability in Allieae (predominantly $2n = 16$) and Tulbaghieae (predominantly $2n = 12$) are probably results of two distinct processes: recent colonization in North Hemisphere and relictual distribution in south Africa, respectively. On the other hand, the South American tribe Gilliesieae ($x = 6$) varied widely in genome size, chromosome number and rDNA sites distribution mainly related to Robertsonian translocations.

DATA AVAILABILITY STATEMENT

All datasets generated and analyzed for this study are included and cited in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

LC, HJ, and GS designed the study. GS and LC carried out FISH and flow cytometry experiments, respectively. JC-S and IE reviewed phylogenetic analyses. IE collected and identified Chilean Gilliesieae species. GS, LC, JC-S, and RC performed data analysis and wrote the manuscript. All authors read and approved the manuscript.

FUNDING

This study was financed in part by the CAPES (Brazil) – Finance Code 001.

ACKNOWLEDGMENTS

The authors wish to thank the Brazilian agencies Conselho Nacional de Desenvolvimento Cientiífico e Tecnológico (CNPq) for a productivity fellowship for GS (process number PQ - 310693/2018-7) and Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE - APQ- 0970- 2.03/15) for financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00320/full#supplementary-material

FIGURE S1 | Ancestral haploid chromosome number reconstruction in Amaryllidaceae. Pies charts at the nodes represent the probabilities of the inferred numbers with the most probable number shown inside the pie. Color coding is explained at the left side captions. Numbers above branches represent the posterior probability of different chromosome number events as explained in the botton left caption.

TABLE S1 | List of Amaryllidaceae species, haploid chromosome numbers (*n*), genome size of the diploid chromosome complement (2C) in picograms (pg), number of 35S and 5S rDNA sites, references for both 2C and rDNA sites number and GenBank accession numbers for the four regions used in this work. * = 2C values obtained from the Royal Botanic Gardens Kew C-value database (https://cvalues.science.kew.org/, Bennett and Leitch, 2012).

TABLE S2 | Number of species sampled for BAMM analysis and percentage relative to the total number of species reported for tribes Tulbaghieae, Gilliesieae (divided in subtribes Leucocoryninae and Gilliesiinae) and Allieae (divided in the three *Allium* evolutionary lineages).

TABLE S3 | Results from the evolutionary model analysis for chromosome number and genome size with Log-Likelihood (Log-Lk) and AIC scores for each of the nine models. The analysis were repeated seven times for different combinations of taxa. Most likely models are highlighted in bold.

REFERENCES

Akaike, H. A. (1974). New look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Angiosperm Phylogeny Group [APG] (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385

Antonelli, A., Nylander, J. A., Persson, C., and Sanmartín, I. (2009). Tracing the impact of the Andean uplift on neotropical plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9749–9754. doi: 10.1073/pnas.0811421106

Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., et al. (2019). Standards for distribution models in biodiversity assessments. *Sci. Adv.* 5:eaat4858. doi: 10.1126/sciadv.aat4858

Bartish, I. V., Antonelli, A., Richardson, J. E., and Swenson, U. (2011). Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily chrysophylloideae (Sapotaceae). *J. Biogeogr.* 38, 177–190. doi: 10.1111/j.1365-2699.2010.02389.x

Bennett, M. D., and Leitch, I. J. (2012). *Plant DNA C-Values Database (Release 6.0)*. Available at: https://cvalues.science.kew.org/

Bossuyt, F., Brown, R. M., Hillis, D. M., Cannatella, D. C., and Milinkovitch, M. C. (2006). Phylogeny and biogeography of a cosmopolitan frog radiation: late

Cretaceous diversification resulted in continent-scale endemism in the family Ranidae. *Syst. Biol.* 55, 579–594. doi: 10.1080/10635150600812551

Bremer, K. (2000). Early Cretaceous lineages of monocot flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4707–4711. doi: 10.1073/pnas.080421597

Briggs, J. C. (2003). The biogeographic and tectonic history of India. *J. Biogeogr.* 30, 381–388. doi: 10.1046/j.1365-2699.2003.00809.x

Brown, J. H., Stevens, G. C., and Kaufman, D. M. (1996). The geographic range: size, shape, boundaries, and internal structure. *Annu. Rev. Ecol. Evol. Syst.* 27, 597–623. doi: 10.1146/annurev.ecolsys.27.1.597

Buerki, S., Forest, F., Alvarez, N., Nylander, J. A. A., Arrigo, N., and Sanmartín, I. (2011). An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *J. Biogeogr.* 38, 531–550. doi: 10.1111/j.1365-2699.2010.02432.x

Butler, M. A., and King, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164, 683–695. doi: 10.1086/426002

Carta, A., Bedini, G., and Peruzzi, L. (2018). Unscrambling phylogenetic effects and ecological determinants of chromosome numbers in major angiosperms clades. *Sci. Rep.* 8:14258. doi: 10.1038/s41598-018-32515-x

Carta, A., and Peruzzi, L. (2016). Testing the large genome constraint hypothesis: plant traits, habitat and climate seasonality in Liliaceae. *New Phytol.* 210, 709–716. doi: 10.1111/nph.13769

Cave, M. S., and Bradley, M. V. (1943). Alterations of chromosome numbers in *Miersia chilensis*. *Am. J. Bot.* 30, 142–149. doi: 10.1002/j.1537-2197.1943.tb14741.x

Chase, M. W., Reveal, J. L., and Fay, M. F. (2009). A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Bot. J. Linn. Soc.* 161, 132–136. doi: 10.1111/j.1095-8339.2009.00999.x

Chen, S., Kim, D. K., Chase, M. W., and Kim, J. H. (2013). Networks in a large-scale phylogenetic analysis: reconstructing evolutionary history of Asparagales (Lilianae) based on four plastid genes. *PLoS One* 8:e59472. doi: 10.1371/journal.pone.0059472

Colombo, P., and Confalonieri, V. (2004). Cytogeography and the evolutionary significance of B chromosomes in relation to inverted rearrangements in a grasshopper species. *Cytogenet. Gen. Res.* 106, 351–358. doi: 10.1159/000079312

Crayn, D. M., Costion, C., and Harrington, M. G. (2014). The Sahul–Sunda floristic exchange: dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. *J. Biogeogr.* 42, 11–24. doi: 10.1111/jbi.12405

Crosa, O. (1972). Estudios cariología en el género *Nothoscordum* (Liliaceae). *Bol. Fac. Agron. Urug.* 122, 3–8.

Crosa, O. (1981). Los cromosomas de cinco especies del género *Tristagma* (Liliaceae). *Darwiniana* 23, 361–366.

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

Datta-Roy, A., and Karanth, K. P. (2009). The out-of-India hypothesis: what do molecules suggest? *J. Biosci.* 34, 687–697. doi: 10.1007/s12038-009-0057-8

Doležel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2, 2233–2244. doi: 10.1038/nprot.2007.310

Doležel, J., Sgorbati, S., and Lucretti, S. (1992). Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plant.* 85, 625–631. doi: 10.1111/j.1399-3054.1992.tb04764.x

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

Escobar, I. (2012). *Systematics of the Tribe Gilliesieae Lindl. (Alliaceae), Based on Morpho-Anatomical, Cytological and Molecular Evidence*. Biology thesis, Universidad deConcepción, Bio Bio.

Felsenstein, J. (1973). Maximum likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* 25, 471–492.

Friesen, N., Fritsch, R. M., and Blattner, F. R. (2006). Phylogeny and new intrageneric classification of *Allium* (Alliaceae) based on nuclear ribosomal DNA ITS sequences. *Aliso* 22, 372–395. doi: 10.3732/ajb.1200641

Gernhard, T. (2008). New analytic results for speciation times in neutral models. *Bull. Math. Biol.* 70, 1082–1097. doi: 10.1007/s11538-007-9291-0

Givnish, T. J., and Renner, S. S. (2004). Tropical intercontinental disjunctions: Gondwana breakup, immigration from the boreotropics, and transoceanic dispersal. *Int. J. Plant Sci.* 165, S1–S6. doi: 10.1086/424022

Glick, L., and Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* 31, 1914–1922. doi: 10.1093/molbev/msu122

Guerra, M. S. (1986). Reviewing the chromosome nomenclature of Levan et al. *Rev. Brasil. de Genética* 9, 741–743.

Hälbich, I. W. (1992). "The cape fold belt orogeny: state of the art 1970s– 1980s," in *Inversion Tectonics of the Cape Fold Belt, Karoo and Cretaceous Basins of Southern Africa*, eds M. J. de Wit and I. G. D. Ransome (Rotterdam: A.A. Balkema), 141–158.

Hall, K. J., and Parker, J. S. (1995). Stable chromosome fission associated with rDNA mobility. *Chromosome Res.* 3, 417–422. doi: 10.1007/BF00713891

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64, 2385–2396. doi: 10.1111/j.1558-5646.2010.01025.x

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., and Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics* 24, 129–131. doi: 10.1093/bioinformatics/btm538

Hauenschild, F., Favre, A., Schnitzler, J., Michalaka, I., Freiberg, M., and Muellner-Riehla, A. N. (2017). Spatio-temporal evolution of *Allium* L. in the Qinghai–Tibet-Plateau region: immigration and in situ radiation. *Plant Divers.* 39, 167–179. doi: 10.1016/j.pld.2017.05.010

Hermant, M., Hennion, F., Bartish, I. V., Yguel, B., and Prinzing, A. (2012). Disparate relatives: life histories vary more in genera occupying intermediate environments. *Perspect. Plant Ecol. Evol. Syst.* 4, 283–301. doi: 10.1016/j.ppees.2012.02.001

Janssen, T., and Bremer, K. (2004). The age of major monocot groups inferred from 800+ *rbc*L sequences. *Bot. J. Linn. Soc.* 146, 385–398. doi: 10.1111/j.1095-8339.2004.00345.x

Janssens, S. B., Groeninckx, T., De Block, P. J., Verstraete, B., Smets, E. F., and Dessein, S. (2016). Dispersing towards Madagascar: biogeography and evolution of the Madagascan endemics of the Spermacoceae tribe (Rubiaceae). *Mol. Phylogenet. Evol.* 95, 58–66. doi: 10.1016/j.ympev.2015.10.024

Jones, K. (1998). Robertsonian fusion and centric fission in karyotype evolution of higher plants. *Bot. Rev.* 64, 273–289. doi: 10.1007/BF02856567

Karanth, K. P. (2003). Evolution of disjunct distributions among wet-zone species of the Indian subcontinent: testing various hypotheses using a phylogenetic approach. *Curr. Sci.* 85, 1276–1283.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kolano, B., Siwinska, D., McCann, J., and Weiss-Schneeweiss, H. (2015). The evolution of genome size and rDNA in diploid species of *Chenopodium* sl (Amaranthaceae). *Bot. J. Linn. Soc.* 179, 218–235. doi: 10.1111/boj.12321

Legendre, P., and Legendre, L. (2012). *Numerical Ecology*, 3rd Edn. Amsterdam: Elsevier.

Li, Q.-Q., Zhou, S.-D., He, X.-J., Yu, Y., Zhang, Y.-C., and Wei, X.-Q. (2010). Phylogeny and biogeography of *Allium* (Amaryllidaceae: Allieae) based on nuclear ribosomal internal transcribed spacer and chloroplast *rps16* sequences, focusing on the inclusion of species endemic to China. *Ann. Bot.* 106, 709–733. doi: 10.1093/aob/mcq177

Lima-de-Faria, A. (1976). The chromosome field: I. Prediction of the location of ribosomal cistrons. *Hereditas* 83, 1–22. doi: 10.1111/j.1601-5223.1976.tb01565.x

Liu, K., Meng, W., Zheng, L., Wang, L., and Zhou, S. (2019). Cytogeography and chromosomal variation of the endemic East Asian herb *Lycoris radiata*. *Ecol. Evol.* 9, 6849–6859. doi: 10.1002/ece3.5252

Loureiro, J., Rodriguez, E., Doležel, J., and Santos, C. (2007). Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Ann. Bot.* 100, 875–888. doi: 10.1093/aob/mcm152

Lowry, E., and Lester, S. E. (2006). The biogeography of plant reproduction: potential determinants of species' range sizes. *J. Biogeogr.* 33, 1975–1982. doi: 10.1111/j.1365-2699.2006.01562.x

Matzke, N. J. (2012). Founder-event speciation in BioGeoBEARS package dramatically improves likelihoods and alters parameter inference in Dispersal-Extinction-Cladogenesis (DEC) analyses. *Front. Biogeogr.* 4:210.

Matzke, N. J. (2013). *BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R Package, version 0.2.1*, Available online at: http://CRAN.R-project.org/package=BioGeoBEARS (accessed November 12, 2019).

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst. Biol.* 63, 951–970. doi: 10.1093/sysbio/syu056

Menezes, R. S., Brady, S. G., Carvalho, A. F., Del Lama, M. A., and Costa, M. A. (2017). The roles of barriers, refugia, and chromosomal clines underlying diversification in Atlantic Forest social wasps. *Sci. Rep.* 7:7689. doi: 10.1038/s41598-017-07776-7

Meng, H. H., Jacques, F. M., Su, T., Huang, Y. J., Zhang, S. T., Ma, H. J., et al. (2014). New biogeographic insight into *Bauhinia* sl (Leguminosae): integration from fossil records and molecular analyses. *BMC Evol. Biol.* 14:181. doi: 10.1186/s12862-014-0181-4

Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., et al. (2015). A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol. Bioinform.* 11, 43–48. doi: 10.4137/EBO.S21501

Morley, R. J. (2000). *Origin and Evolution of Tropical Rain Forests*. Hoboken, NJ: John Wiley & Sons.

Nguyen, P., Sahara, K., Yoshido, A., and Marec, F. (2010). Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica* 138, 343–354. doi: 10.1007/s10709-009-9424-5

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature* 401, 877–884. doi: 10.1038/44766

Pascoe, P. L., Patton, S. J., Critcher, R., and Dixon, D. R. (1996). Robertsonian polymorphism in the marine gastropod, *Nucella lapillus*: advances in karyology using rDNA loci and NORs. *Chromosoma* 104, 455–460. doi: 10.1007/BF00352269

Pellicer, J., Hidalgo, O., Walker, J., Chase, M. W., Christenhusz, M. J., Shackelford, G., et al. (2017). Genome size dynamics in tribe Gilliesieae (Amaryllidaceae, subfamily Allioideae) in the context of polyploidy and unusual incidence of Robertsonian translocations. *Bot. J. Linn. Soc.* 184, 16–31. doi: 10.1093/botlinnean/box016

Pennington, R. T., Lavin, M., Särkinen, T., Lewis, G. P., Klitgaard, B. B., and Hughes, C. E. (2010). Contrasting plant diversification histories within the Andean biodiversity hotspot. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13783–13787. doi: 10.1073/pnas.1001317107

Peruzzi, L., Carta, A., and Altinordu, F. (2017). Chromosome diversity and evolution in *Allium* (Allioideae, Amaryllidaceae). *Plant Biosyst.* 151, 212–220. doi: 10.1080/11263504.2016.1149123

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News* 6, 7–11.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for statistical Computing.

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., et al. (2014). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5, 701–707. doi: 10.1111/2041-210X.12199

Rambaut, A. (2014). *FigTree 1.4. 2 Software*. Edinburgh: The University of Edinburgh.

Rambaut, A., and Drummond, A. J. (2013). *TreeAnnotator v1. 7.0. Available as Part of the BEAST package*. Available at: http://beast.bio.ed.ac.uk

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Raskina, O., Belyayev, A., and Nevo, E. (2004). Quantum speciation in *Aegilops*: molecular cytogenetic evidence from rDNA cluster variability in natural populations. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14818–14823. doi: 10.1073/pnas.0405817101

Revell, L. J. (2012). *phytools*: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., et al. (2015). The chromosome counts database (CCDB) – a community resource of plant chromosome numbers. *New Phytol.* 206, 19–26. doi: 10.1111/nph.13191

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

Rudall, P. J., Bateman, R. M., Fay, M. F., and Eastman, A. (2002). Floral anatomy and systematics of Alliaceae with particular reference to *Gilliesia*, a presumed insect mimic with strongly zygomorphic flowers. *Am. J. Bot.* 89, 1867–1883. doi: 10.3732/ajb.89.12.1867

Sader, M. A., Amorim, B. S., Costa, L., Souza, G., and Pedrosa-Harand, A. (2019). The role of chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). *Syst. Biodivers.* 17, 7–21. doi: 10.1080/14772000.2018.1546777

Sassone, A. B., López, A., Hojsgaard, D. H., and Giussani, L. M. (2018). A novel indicator of karyotype evolution in the tribe Leucocoryneae (Allioideae, Amaryllidaceae). *J. Plant Res.* 131, 211–223. doi: 10.1007/s10265-017-0987-4

Serbin, G. M., Machado, R. M., Pinto, R. B., Diniz Filho, J. A. F., de Azevedo Tozzi, A. M. G., Forni-Martins, E. R., et al. (2019). Karyological traits related to phylogenetic signal and environmental conditions within the *Hymenaea* clade (Leguminosae, Detarioideae). *Perspect. Plant Ecol. Evol. Syst.* 39:125462. doi: 10.1016/j.ppees.2019.125462

Souza, G., Costa, L., Guignard, M. S., Van-Lume, B., Pellicer, J., Gagnon, E., et al. (2019a). Do tropical plants have smaller genomes? Correlation between genome size and climatic variables in the *Caesalpinia* Group (Caesalpinioideae, Leguminosae). *Perspect. Plant Ecol. Evol. Syst.* 38, 13–23. doi: 10.1016/j.ppees.2019.03.002

Souza, G., Marques, A., Ribeiro, T., Dantas, L. G., Speranza, P., Guerra, M., et al. (2019b). Allopolyploidy and extensive rDNA site variation underlie rapid karyotype evolution in *Nothoscordum* section *Nothoscordum* (Amaryllidaceae). *Bot. J. Linn. Soc.* 190, 215–228. doi: 10.1093/botlinnean/boz008

Souza, G., Crosa, O., and Guerra, M. (2015). Karyological, morphological, and phylogenetic diversification in *Leucocoryne* Lindl (Allioideae, Amaryllidaceae). *Plant Syst. Evol.* 301, 2013–2023. doi: 10.1007/s00606-015-1216-z

Souza, G., Crosa, O., Speranza, P., and Guerra, M. (2016). Phylogenetic relations in tribe Leucocoryneae (Amaryllidaceae, Allioideae) and the validation of *Zoellnerallium* based on DNA sequences and cytomolecular data. *Bot. J. Linn. Soc.* 182, 811–824. doi: 10.1111/boj.12484

Souza, L. G. R., Crosa, O., and Guerra, M. (2010). Karyological circumscription of *Ipheion* Rafinesque (Gilliesioideae, Alliaceae). *Plant Syst. Evol.* 287, 119–127. doi: 10.1007/s00606-010-0304-3

Souza, L. G. R., Crosa, O., Speranza, P., and Guerra, M. (2012). Cytogenetic and molecular evidence suggest multiple origins and geographical parthenogenesis in *Nothoscordum gracile* (Alliaceae). *Ann. Bot.* 109, 987–999. doi: 10.1093/aob/mcs020

Stafford, G. I., Wikkelsø, M. J., Nancke, L., Jäger, A. K., Möller, M., and Rønsted, N. (2016). The first phylogenetic hypothesis for the southern African endemic genus *Tulbaghia* (Amaryllidaceae, Allioideae) based on plastid and nuclear DNA sequences. *Bot. J. Linn. Soc.* 181, 156–170. doi: 10.1111/boj.12417

Stebbins, G. L. (1971). *Chromosomal Evolution in Higher Plants*. London: Edward Arnold, 216.

Tagashira, N., and Kondo, K. (2001). Chromosome phylogeny of *Zamia* and *Ceratozamia* by means of Robertsonian changes detected by fluorescence *in situ* hybridization (FISH) technique of rDNA. *Plant Syst. Evol.* 227, 145–155. doi: 10.1007/s006060170045

Thiv, M., Van Der Niet, T., Rutschmann, F., Thulin, M., Brune, T., and Linder, H. P. (2011). Old–new world and trans-African disjunctions of *Thamnosma* (Rutaceae): intercontinental long-distance dispersal and local differentiation in the succulent biome. *Am. J. Bot.* 98, 76–87. doi: 10.3732/ajb.1000339

Van-Lume, B., Esposito, T., Diniz-Filho, J. A. F., Gagnon, E., Lewis, G. P., and Souza, G. (2017). Heterochromatic and cytomolecular diversification in the *Caesalpinia* group (Leguminosae): relationships between phylogenetic and cytogeographical data. *Perspect. Plant Ecol. Evol. Syst.* 29, 51–63. doi: 10.1016/j.ppees.2017.11.004

Vosa, C. G. (2000). A revised cytotaxonomy of the genus *Tulbaghia* (Alliaceae). *Caryologia* 53, 83–112. doi: 10.1080/00087114.2000.10589184

Vosa, C. G. (2007). *Prototulbaghia*, a new genus of the Alliaceae family from the Leolo Mountains in Sekhukhuneland, South Africa. *Caryologia* 60, 273–278. doi: 10.1080/00087114.2007.10797948

Wei, R., Xiang, Q., Schneider, H., Sundue, M. A., Kessler, M., Kamau, P. W., et al. (2015). Eurasian origin, boreotropical migration and transoceanic dispersal in the pantropical fern genus *Diplazium* (Athyriaceae). *J. Biogeogr.* 42, 1809–1819. doi: 10.1111/jbi.12551

WFO (2019). *World Flora Online*. Available at: http://www.worldfloraonline.org (accessed November 12, 2019).

Wing, S. L., Herrera, F., Jaramillo, C. A., Gómez-Navarro, C., Wilf, P., and Labandeira, C. C. (2009). Late paleocene fossils from the cerrejón formation, Colombia, are the earliest record of neotropical rainforest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18627–18632. doi: 10.1073/pnas.0905130106

Yang, T., Lu, L.-M., Wang, W., Li, J.-H., Manchester, S. R., Wen, J., et al. (2018). Boreotropical range expansion and long-distance dispersal explain two amphi-Pacific tropical disjunctions in Sabiaceae. *Mol. Phylogenet. Evol.* 124, 181–191. doi: 10.1016/j.ympev.2018.03.005

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: 20 May 2020
doi: 10.3389/fpls.2020.00644

What Can Long Terminal Repeats Tell Us About the Age of LTR Retrotransposons, Gene Conversion and Ectopic Recombination?

*Pavel Jedlicka[1], Matej Lexa[2] and Eduard Kejnovsky[1]**

[1] Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czechia,
[2] Faculty of Informatics, Masaryk University, Brno, Czechia

LTR retrotransposons constitute a significant part of plant genomes and their evolutionary dynamics play an important role in genome size changes. Current methods of LTR retrotransposon age estimation are based only on LTR (long terminal repeat) divergence. This has prompted us to analyze sequence similarity of LTRs in 25,144 LTR retrotransposons from fifteen plant species as well as formation of solo LTRs. We found that approximately one fourth of nested retrotransposons showed a higher LTR divergence than the pre-existing retrotransposons into which they had been inserted. Moreover, LTR similarity was correlated with LTR length. We propose that gene conversion can contribute to this phenomenon. Gene conversion prediction in LTRs showed potential converted regions in 25% of LTR pairs. Gene conversion was higher in species with smaller genomes while the proportion of solo LTRs did not change with genome size in analyzed species. The negative correlation between the extent of gene conversion and the abundance of solo LTRs suggests interference between gene conversion and ectopic recombination. Since such phenomena limit the traditional methods of LTR retrotransposon age estimation, we recommend an improved approach based on the exclusion of regions affected by gene conversion.

Keywords: transposable elements, LTR retrotransposons, nesting, age estimation, gene conversion, ectopic recombination, plants

OPEN ACCESS

Edited by:
Hanna Weiss-Schneeweiss, University of Vienna, Austria

Reviewed by:
Ruslan Kalendar, University of Helsinki, Finland
Tony Heitkam, Technische Universität Dresden, Germany
Carlos M. Vicient, Centre for Research in Agricultural Genomics (CRAG), Spain

****Correspondence:***
Eduard Kejnovsky
kejnovsk@ibp.cz

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

Received: *14 November 2019*
Accepted: *27 April 2020*
Published: *20 May 2020*

Citation:
Jedlicka P, Lexa M and Kejnovsky E (2020) What Can Long Terminal Repeats Tell Us About the Age of LTR Retrotransposons, Gene Conversion and Ectopic Recombination? Front. Plant Sci. 11:644. doi: 10.3389/fpls.2020.00644

INTRODUCTION

Transposable elements (TEs) are abundant structural and functional genome components inhabiting genomes throughout the course of life evolution. They have evolved into many different types distinguished by structure, mechanisms of spreading and effect on cell functioning. The activity of transposable elements is dependent on the developmental stage, is tissue-specific, epigenetically regulated and often induced by stress. This is evident especially in plants (that are sessile) where TEs represent in large genomes like maize, barley or wheat often more than 85% of the genome (Charles et al., 2008; Schnable et al., 2009; Wicker et al., 2018).

LTR retrotransposons are ancient genome inhabitants present in the genomes of all major taxonomic groups, being abundant especially in plants (Feschotte et al., 2002; Kejnovsky et al., 2012). They exhibit waves of explosive amplification during the evolution of host species that often predate the speciation events (Kim et al., 2004). Since retrotransposon activation is caused

by stress (Hirochika, 1997; Grandbastien, 1998), such amplification waves probably corresponded to major environmental challenges such as climate change or pathogen attack. The generation of new retrotransposon copies is balanced by deletions resulting from ectopic recombination and the formation of solo LTRs, leading to genome size either increasing or decreasing (Devos et al., 2002; Ma et al., 2004; Bennetzen et al., 2005; Vitte and Panaud, 2005).

The removal of LTR retrotransposons is caused by two unrelated, ectopic processes: (i) homologous unequal recombination, producing solo LTRs with or without TSDs, intact elements without TSDs and/or recombined elements with LTRs flanked by both PBS and PPT (Shirazu et al., 2000; Vitte and Panaud, 2003), and by (ii) illegitimate recombination, using a mechanism of mis-repair of double strand breaks, as was shown in Arabidopsis (Devos et al., 2002) and wheat (Wicker et al., 2003) and resulting in whole or partial deletion of LTRs (Devos et al., 2002; Ma et al., 2004).

Retrotransposon activity during the course of evolution differs between families and plant species. Some retrotransposon families have short bursts of intense activity for a few 100,000 years while other families have only moderate activity over long periods of time e.g., 1–2 million years (Wicker and Keller, 2007). Such amplifications are more visible in animal genomes (Kim et al., 2004) than in plants because plant genomes are more dynamic and intermingled (Kejnovsky et al., 2009). While in animals endogenous retroviral integrations older than 100 million can be identified (Martins and Villesen, 2011), the high turnover of retrotransposons (birth and decay of elements) in plant genomes prevents the detection of insertions more than tens of million years old (Maumus and Quesneville, 2014). Seminal papers from the beginning of this millenium, analyzing a number of plant species, showed that the majority of LTR retrotransposons were inserted less than three million years ago (Devos et al., 2002; Ma et al., 2004; Bennetzen et al., 2005).

Studies of the evolutionary dynamics of various LTR retrotransposon families are based on the estimation of relative and absolute age (Kijima and Innan, 2009). The age of LTR retrotransposons is mostly estimated using the divergence of 5′ and 3′ LTRs (Gaut et al., 1996; SanMiguel et al., 1998, 2002). However, recent studies have shown that this traditional age estimation method is not absolute, namely because (i) the differences in substitution rates between species (Ma and Bennetzen, 2004) and (ii) the effect of homogenizing processes such as gene conversion (Kijima and Innan, 2009; Cossu et al., 2017).

The absolute age of LTR retrotransposons is calculated using the formula $T = K/2 \times r$, where T = time of divergence, K = divergence and r = substitution rate (Bowen and McDonald, 2001). However, substitution rates vary between species e.g., 1.6×10^{-8} substitutions per site per year in drosophila (Li, 1997), 1.5×10^{-8} in Arabidopsis (Koch et al., 2000) and 1.3×10^{-8} in grasses (Ma and Bennetzen, 2004). The weakness of the traditional method for LTR retrotransposon and retrovirus integration time estimation using only LTR divergence has been highlighted previously by Martins and Villesen (2011) who developed an improved approach using phylogenetic data. These authors showed that 5′ and 3′ LTR have distinct evolutionary rates.

The need for other approaches for LTR retrotransposon age estimation has led to the development of an alternative method based on the comparison of intra-specific versus interspecific differences in repeats (species-specific elements are younger than conservative elements). This method has been used to date a variety of repeats (not only LTR retrotransposons) in Arabidopsis (Maumus and Quesneville, 2014) and the legume tribe Fabeae (Macas et al., 2015).

Here we measured the LTR divergence of thousands of LTR retrotransposons coming from fifteen plant species to determine their age and thus study their evolutionary dynamics. We found that LTR divergence depends not only on the element age but also on e.g., LTR length. We propose gene conversion as the process complicating age estimation from LTR similarity. In addition, we measured the extent of gene conversion in LTRs as well as its relation to other processes such as solo LTR formation by ectopic recombination.

MATERIALS AND METHODS

Genomic Sequence Sources and TE Annotation

Plant genomes covering diverse taxons of higher plants were downloaded from Phytozome 12.0 (Goodstein et al., 2012). The fifteen species included *Arabidopsis thaliana* (Lamesch et al., 2012), *Arabidopsis lyrata* (Rawat et al., 2015), *Brachypodium distachyon* (International Brachypodium Initiative, 2010), *Chlamydomonas reinhardtii* (Merchant et al., 2007), *Glycine max* (Schmutz et al., 2010), *Lotus japonicus* (Sato et al., 2008)[1], *Medicago truncatula* (Tang et al., 2014), *Musa acuminata* (D'Hont et al., 2012), *Oryza sativa* (Ouyang et al., 2007), *Physcomitrella patens* (Lang et al., 2018), *Populus trichocarpa* (Tuskan et al., 2006), *Selaginella moellendorffii* (Banks et al., 2011), *Sorghum bicolor* (McCormick et al., 2017), *Solanum lycopersicum* (Tomato Genome Consortium, 2012), and *Solanum tuberosum* (Sharma et al., 2013). The complete workflow of our analysis is visualized as a step-by-step flowchart in **Supplementary Figure S1**. Unmasked sequences were analyzed with TE-greedy-nester (Lexa et al., 2018). TE-greedy-nester in its latest version relies upon LTR Finder (Xu and Wang, 2007) to identify full-length LTR retroelements. It recursively removes the identified elements from the analyzed genomes so that other full-length copies fragmented by nesting can be identified with the same tools. The annotations were saved as GFF3 files for visualization and downstream analysis. They contained information on the positions of entire elements as well as their structural components [LTR, PBS, PPT, *gag* and *pol* gene protein domain sequences, target site duplications (TSD)]. Subsequences of interest (LTR, RT domain) were extracted from downloaded genome sequences using the bedtools package (Quinlan and Hall, 2010).

[1] ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r2.5/Lj2.5_genome_contigs.fna.gz

The elements, retrieved by TE-greedy-nester, which contain detected LTR retroelement protein domains are also automatically annotated using recent classification by Neumann et al. (2019). The annotation process is based on homology (BLASTX; Altschul et al., 1990) with a custom database consisting of a combination of Cores Seq. from Gypsy Database (Llorens et al., 2011) and polyprotein sequences recently present by Neumann et al. (2019). Therefore, GFF3 outputs were filtered for the presence of at least one protein domain. Further, based on the mutual position of annotated LTR retrotransposons within the genomic sequence the TEs with boundaries present within the coordinates of another TE were simply considered as "nested" and "original," respectively. *Vice versa*, the solitary TE was labeled as "non-nested." Finally, in order to minimize the amount of false positive elements detected by TE-greedy-nester, all the non-nested and original elements were filtered for the presence of TSD. The respective counts of LTR retrotransposons used in this study are given in **Table 1** and corresponding GFF files of filtered retroelements are provided in the **Supplementary Material**. The plant species presented in table and all figures are ordered by their genome size in **Table 1** and by their taxonomic affiliation in Figures. LTR retrotransposon families labels in **Supplementary Figure S3** are presented as a combination of superfamily (i.e., Ty1/Copia and Ty3/Gypsy as "copia" and "gypsy," respectively) and given families concatenated by underscores (e.g., "gypsy_Athila").

LTR Divergence

The LTR divergence of elements in individual families was obtained from global alignment by STRETCHER tool (Emboss 6.6.0; Rice et al., 2000), expressed as percentage of identical bases in the alignment (LTR similarity). These values served for visualization of LTR similarity and length relationship and subtraction of LTR similarity within each pair of nested and original (pre-existing) element ("delta LTR similarity"). Furthermore, in order to exclude the possibility that the observed negative delta LTR similarity was simply a result of random mutations, we simulated a pair of LTRs subject to mutations with BBMap mutate.sh[2] and subsequently generated 1000 independent mutations of that pair. For each pair of sequences we calculated the similarity of their global alignment and plotted the distribution of these values as simulated delta LTR similarity.

Insertion Time Estimation

The nucleotide divergence between aligned sequences (CLUSTALW tool with -output = PHYLIP command; Larkin et al., 2007) was calculated using PhyML (Guindon and Gascuel, 2003) with substitution model K80. Subsequent steps were adopted from Pereira (2004). In order to minimize errors from poor quality alignments retrieved by CLUSTALW, alignments shorter than eighty nucleotides and LTR pairs with divergence (K) value greater than 0.2 were discarded (207 out of 25,144; i.e., less than 1%). Subsequently insertion time was estimated using the formula $T = K / 2r$, with substitution rate of 1.3×10^{-8} per site per year (Bowen and McDonald, 2001; Ma and Bennetzen, 2004).

Solo LTR Detection

In addition to the GFFs files with information on full-length LTR retrotransposons, the TE-greedy-nester also retrieves respective chromosome sequence remainder after full-length elements removal in FASTA format. These sequences were used for solo LTRs detection, which was conducted in two subsequent steps: (i) LTR_retriever was employed to process split sequences using the default arguments setup (Ou and Jiang, 2018); and (ii) obtained outputs were passed to REannotate software (Pereira, 2008), which clearly distinguish solo LTRs from truncated retroelements containing also uncoupled LTRs or their remnants.

Determination of Gene Conversion in LTRs and Removal of Converted Parts

In order to estimate the extent of potential gene conversion along the long terminal repeats of LTR retrotransposons we employed GENECONV (Sawyer, 1999) which was shown to be precise and reliable compared with other software (Mansai and Innan, 2010). Moreover this tool has already been used for this specific task in plant LTR retrotransposons (Cossu et al., 2017). GENECONV uses permutation analysis of sequence alignment to determine a probability that two LTR subregions have a common origin due to gene conversion. This is based on the density of nucleotide substitutions in these regions, compared to the background in other parts of the input sequences. Consequently, we are aware that alongside gene conversion, the sequence identities retrieved by GENECONV could be caused also by random processes (for instance, a low overall rate of mutation or multiple testing). We consider our results as "upper limits" and interpret the results as "possible gene conversion" on that account. The LTR pair sequences of all elements from each specific family and plant species were collected within one fasta file. Then all possible pairs of LTRs from two different elements were generated (i.e., 5′ and 3′ LTRs from two elements – four sequences per one fasta file). Thereafter each set of LTRs was aligned using CLUSTALW (Larkin et al., 2007) and subjected to GENECONV using parameters: /w123 /lp /f /eb /g1 -nolog. Because of the extraordinary number of pairs (over 100,000 files) generated in some overrepresented retrotransposon families, the GENECONV run was stopped when LTRs of each element were analyzed with those of at least ten other elements. Pairwise inner fragments from GENECONV output were evaluated and filtered. The first filter was conducted in order to avoid getting false positive results due to multiple comparisons of all possible sequences. Thus the *p*-value retrieved by GENECONV was multiplied by the number of all sequences in the original plant- and LTR retrotransposon family specific multifasta file, and only records with p-value < 0.05 were accepted for following steps. Another filter was used in cases where gene conversion fragments overlapped each other and the best candidate was chosen based on the lowest *p*-value and number of mismatches. Further, since the minimal length of gene conversion fragments varied among different organisms (Mansai et al., 2011), we set this value to

[2]https://github.com/BioInfoTools/BBMap

TABLE 1 | Summary table of LTR retrotransposon counts and mean age obtained from fifteen plant species.

Species	Label	Class	Family	LTR retrotransposons				
				Genome size [Mbp]	Nested and Original	Non-nested	Sum	Mean age [± SD]
Glycine max	Gmax	Eudicots	Fabaceae	978.5	789	2876	3665	1.33 ± 1.28
Solanum lycopersicum	Slyc	Eudicots	Solanaceae	823.9	300	1436	1736	2.26 ± 1.65
Solanum tuberosum	Stub	Eudicots	Solanaceae	773.0	253	1140	1393	2.17 ± 1.6
Sorghum bicolor	Sbic	Monocots	Poaceae	732.2	2881	4591	7472	0.89 ± 0.95
Physcomitrella patens	Ppat	Bryopsida	Funariaceae	473.2	633	2478	3111	1.15 ± 1.13
Lotus japonicus	Ljap	Eudicots	Fabaceae	462.5	96	656	752	0.63 ± 1.04
Populus trichocarpa	Ptri	Eudicots	Salicaceae	422.9	88	726	814	1.19 ± 1.39
Medicago truncatula	Mtru	Eudicots	Fabaceae	411.8	139	330	469	2.62 ± 1.77
Musa acuminata	Macu	Monocots	Musaceae	390.6	66	572	638	0.56 ± 0.97
Oryza sativa japonica	Osat	Monocots	Poaceae	374.5	661	1750	2411	0.9 ± 1.07
Brachypodium distachyon	Bdis	Monocots	Poaceae	271.2	137	608	745	1.83 ± 1.23
Selaginella moellendorffii	Smoe	Isoetopsida	Selaginellaceae	212.7	111	648	759	1.58 ± 1.45
Arabidopsis lyrata	Alyr	Eudicots	Brassicaceae	206.7	155	837	992	0.58 ± 0.8
Arabidopsis thaliana	Atha	Eudicots	Brassicaceae	119.1	32	130	162	1.21 ± 1.09
Chlamydomonas reinhardtii	Crei	Chlorophyceae	Chlamydomonadaceae	107.1	0	25	25	0.25 ± 0.89
Total					6341	18,803	25,144	

Average time from their insertion is indicated (million years ago ± Standard deviation).

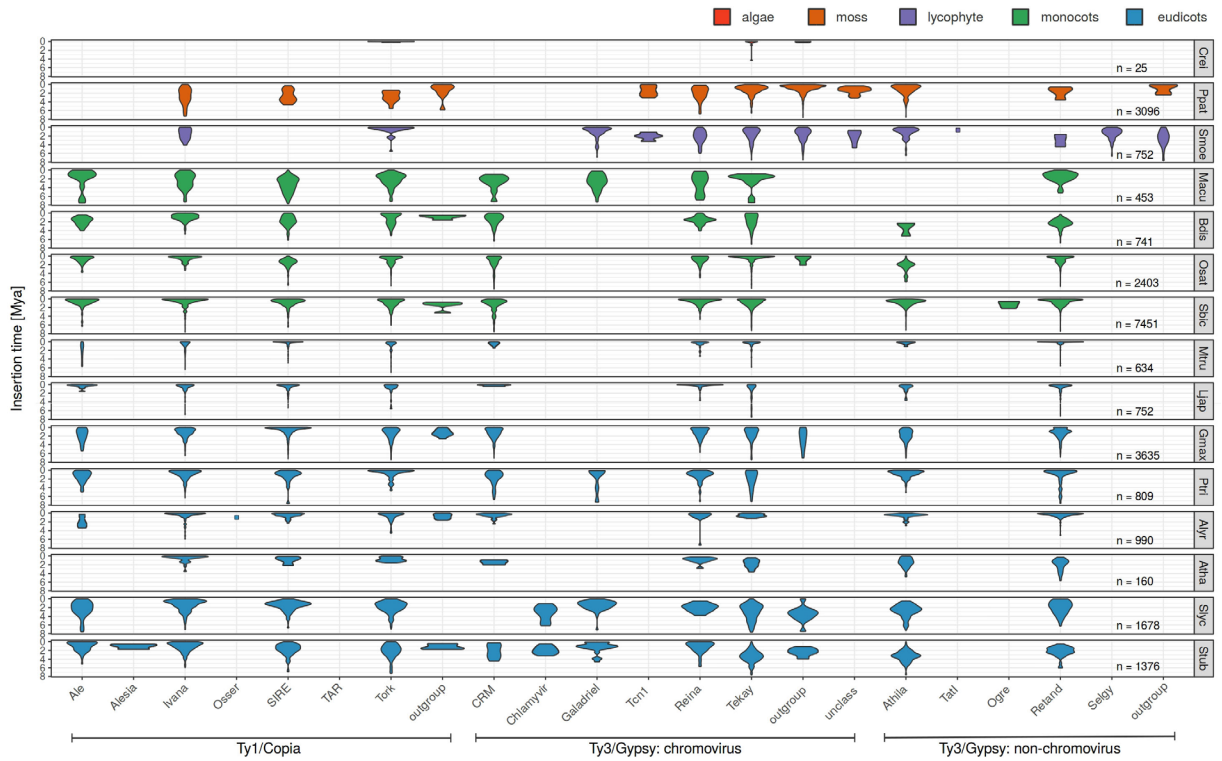

FIGURE 1 | Evolutionary dynamics of LTR retrotransposons of fifteen plant species using LTR divergence method for LTR retrotransposon age estimation. For each species and each family, we measured the abundance of elements having specific LTR divergence to reveal evolutionary expansions and contractions occurring within each family. Nested, original and non-nested LTR retrotransposons were analyzed together.

50 bp to avoid overestimation of our findings. Finally, for the determination of LTR similarity of original elements prior to gene conversion, the maximal length of a converted fragment was limited up to 80% of given LTR length. The converted part was then clipped, flanking parts joined and LTR similarity determined using global alignment by STRETCHER.

Effect of Whole Genome Mutation on LTR Similarity – *In silico* Simulation

Changes in the similarities of LTRs with different lengths were additionally analyzed by the following simulation. We took LTRs of different lengths deposited in the Gypsy database[3] (Llorens et al., 2011) and each LTR sequence was duplicated, the space between two LTRs filled by random sequence with length nine times longer than two respective LTRs (since LTRs constitute 10% of full-length LTR retrotransposon in average). This pseudoelement was then randomly inserted into a generated DNA sequence (1 Mbp long) which represented an artificial genome. Such a genome was subjected to mutation at level ranging from 0.7 to 1.0 (with step of 0.01) using BBMap mutate.sh. The similarity of LTRs were counted by emboss stretcher and plotted against LTR lengths. Because outcomes of all the mutation levels revealed the same pattern, only results at mutation level 0.99 were used for our visualization (**Supplementary Figure S5**).

[3]http://gydb.org/images/9/94/LTRs_and_TIRs.zip

RESULTS

Evolutionary Dynamics of LTR Retrotransposons in Plants

The LTR similarity in individual families of 25,144 LTR retrotransposons in fifteen plant species (**Table 1**) was measured and their age determined using the above mentioned formula and substitution rate of 1.3×10^{-8} per site per year (**Figure 1** and **Supplementary Figure S2**). This constant was estimated and until now is widely used in grasses (Bowen and McDonald, 2001; Ma and Bennetzen, 2004; Choulet et al., 2010; Zhang and Gao, 2017). In addition, this rate was employed also in Solanum (Xu and Du, 2014) and is close to that established for *A. thaliana* (1.5×10^{-8}). The overall average insertion time ranges from 0.25 to 2.62 Mya in green alga *Chlamydomonas reinhardtii* and barrel clover *Medicago truncatula*, respectively (**Table 1** and **Supplementary Figure S2**). In **Figure 1** LTR retrotransposons were sorted according to a recent LTR retrotransposon classification (Neumann et al., 2019) and plant species were sorted according to phylogeny. The patterns of family expansions differed between retrotransposons as well as between plant species. The age distribution of LTR

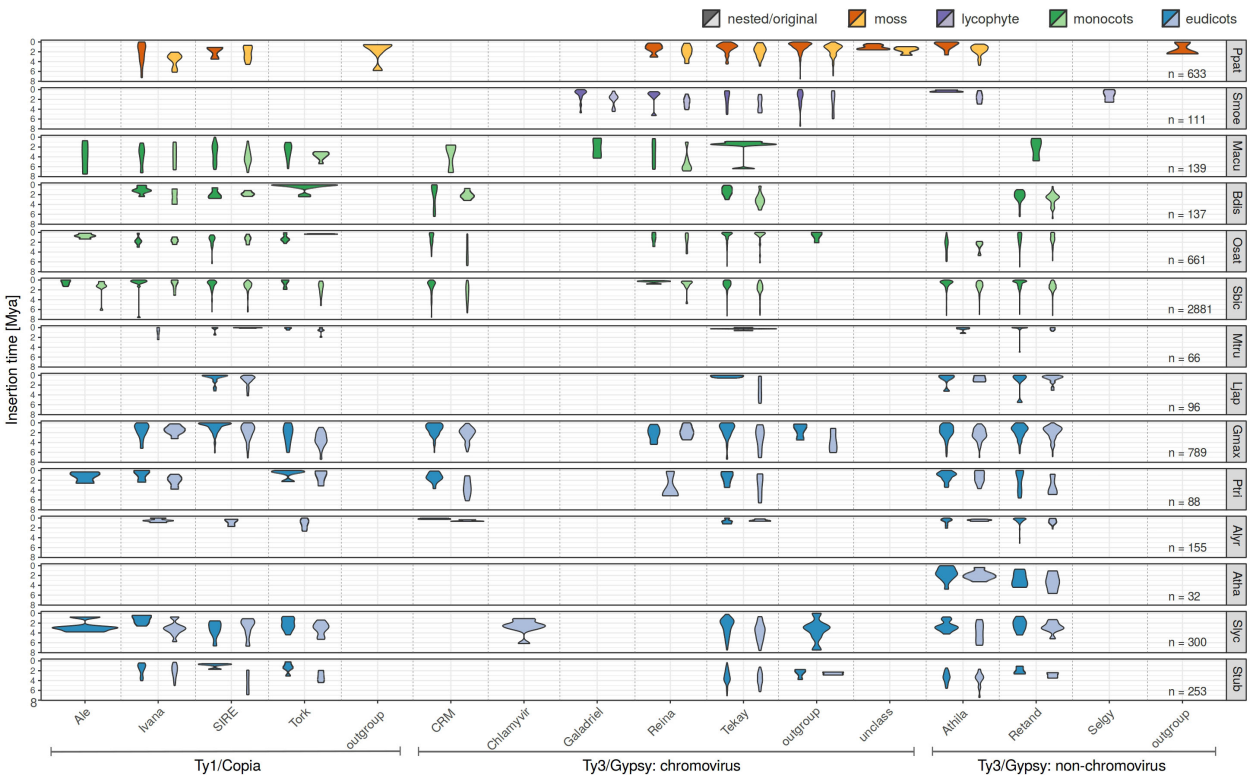

FIGURE 2 | Evolutionary dynamics of nested and original LTR retrotransposons of studied plant species using complex approach for LTR retrotransposon age estimation. For each species and each family, we measured the abundance of elements having specific LTR divergence to reveal evolutionary expansions and contractions occurring within each family. Nested (dark colors), original (light colors) LTR retrotransposons were analyzed separately. Low abundant families were excluded from this visualization.

retrotransposons persisting in one plant species often had similar patterns, despite some visible differences. On the other hand, the evolutionarily dynamics of the same LTR retrotransposon family varied in a number of plant species - some families showed short recent expansion in one species while in another species it had continual moderate activity (**Figure 1**). Specifically, in rice most of the dominant retroelement families showed recent insertions (Ty1/Copia: Ale, Ivan, Tork; Ty3/Gypsy: CRM, Reina, Retand and Tekay), with the exception of Ty1/Copia SIRE and Ty3/Gypsy Athila (**Figure 1**). Similarly, in *Sorghum bicolor* all the abundant families were inserted recently. Contrastingly, in the tomato, potato and soybean we found earlier insertions of most LTR retrotransposons families.

Separate visualization of nested and original (pre-existing) LTR retrotransposons (**Figure 2**) showed that (i) nested retrotransposons are, as expected, mostly younger compared to the original ones (see e.g., SanMiguel et al., 1998 for comparison) and (ii) nested elements showed recent expansion in many families.

Ectopic (unequal) recombination contributes, together with illegitimate recombination, to element removal and genome contraction. In order to detail how this process is related to the expansion of individual retrotransposon families we measured the ratio of solo LTR to full length elements (solo LTR/FL). We found that in the analyzed species the ratio of solo LTR/FL did not change significantly in dependence on genome size (Pearson's $r = 0.1038$ with p-value = 0.2363; **Figure 3**), indicating the similar removal of an LTR retrotransposon by ectopic recombination in large and small genomes. This trend was observed in a wide range of species. The proportion of solo LTR significantly differed between individual chromosomes of the same plant species (**Figure 3**).

LTRs of Nested Elements Are Often More Diverged Than Original (Pre-existing) Elements

To assess the factors contributing to the similarity of 5′ and 3′ LTRs of the same retrotransposon we compared LTRs in 4126 pairs of nested and original (pre-existing) LTR retrotransposons. Nesting is an absolute measure of relative age – the nested element is always younger than the original and thus the similarity of the nested (younger) element should always be higher than the original (older) element. We named the difference of LTR similarity of nested and original elements as "delta LTR similarity" and expected it to always be positive. Negative delta LTR similarity can be a result of processes that affect the LTRs after insertion, such as the homology-driven form of recombination reshaping LTRs - gene conversion. By filtering the original LTR retrotransposons for the presence of TSDs we

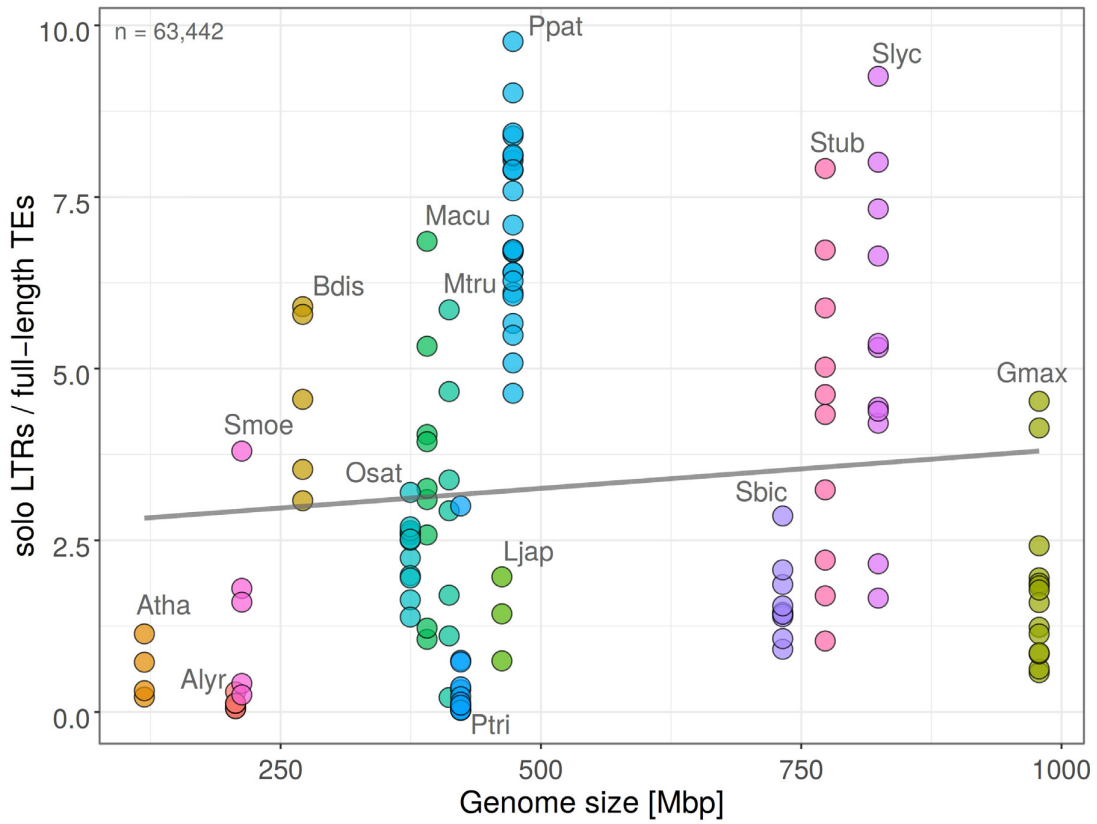

FIGURE 3 | Ratio of solo LTR to full-length retrotransposons (solo LTR/FL) plotted against genome size in fifteen plant species. Each filled circle corresponds to one chromosome, plant species are labeled by different colors ($y = 2.69 + 0.00114x$; $R^2 = 0.011$; Pearson's $r = 0.1038$ with p-value = 0.2363). Total count of solo LTRs is indicated. Higher solo LTR/FL ratio is observed in larger genomes and corresponds to LTR retrotransposon removal by ectopic recombination.

minimized the possibility of improper element delineation by TE-greedy-nester.

We performed this analysis on fifteen plant species and, surprisingly, we found that the delta LTR similarity was often negative i.e., the similarity of nested elements was lower compared to the similarity of original elements (**Figure 4**). The proportion of pairs with negative delta LTR similarity (higher similarity of original than nested elements) was 25% (1042 of 4126) and varied in individual species (**Figure 4**). To rule out the possibility that the observed negative results were simply due to random mutations, we simulated a pair of LTRs with BBMap mutate.sh[4] generating 1000 independent mutations. For each pair of sequences we calculated the similarity of their global alignment and plotted the distribution of these values as simulated delta LTR similarity (gray area, **Figure 4**).

Longer LTRs Have a Higher 5′-3′ LTR Similarity Than Shorter Ones

The age of LTR retrotransposons is mostly determined by a traditional method measuring LTR similarity, based on the fact that the 5′ and 3′ LTRs are identical at the time of insertion and accumulate mutations and diverge as an element gets older. However, during our analyses we found that LTR similarity surprisingly positively correlated with the LTR length (**Figure 5**). The LTRs longer than the median (552 bp) comprised 57 and 65% of the LTRs with 95 and 99% similarity, respectively. This suggests that factors other than age have contributed to the similarity of the LTRs. LTR length density of the most abundant retrotransposon families (Ty1/copia: Ivana, SIRE and Tork; Ty3/gypsy: Athila, CRM, Reina, Retand and Tekay) culminated twice, around 300 and 1000 bp (**Supplementary Figure S3**).

The Extent of Gene Conversion

In order to find a possible explanation for the anomalies described above, we analyzed the extent of potential gene conversion along the long terminal repeats of LTR retrotransposons using GENECONV software. Pairwise inner fragments from GENECONV output were evaluated and filtered for gene conversion length and overlaps, e-value and number of mismatches (see section "Materials and Methods"). After quality filtering we calculated (i) the number of LTR retrotransposon containing gene converted regions in dependence on genome size of host species (**Figure 6A**) and measured (ii) the length of converted regions (**Figure 6B**). Both the number of elements with converted regions and

[4]https://github.com/BioInfoTools/BBMap

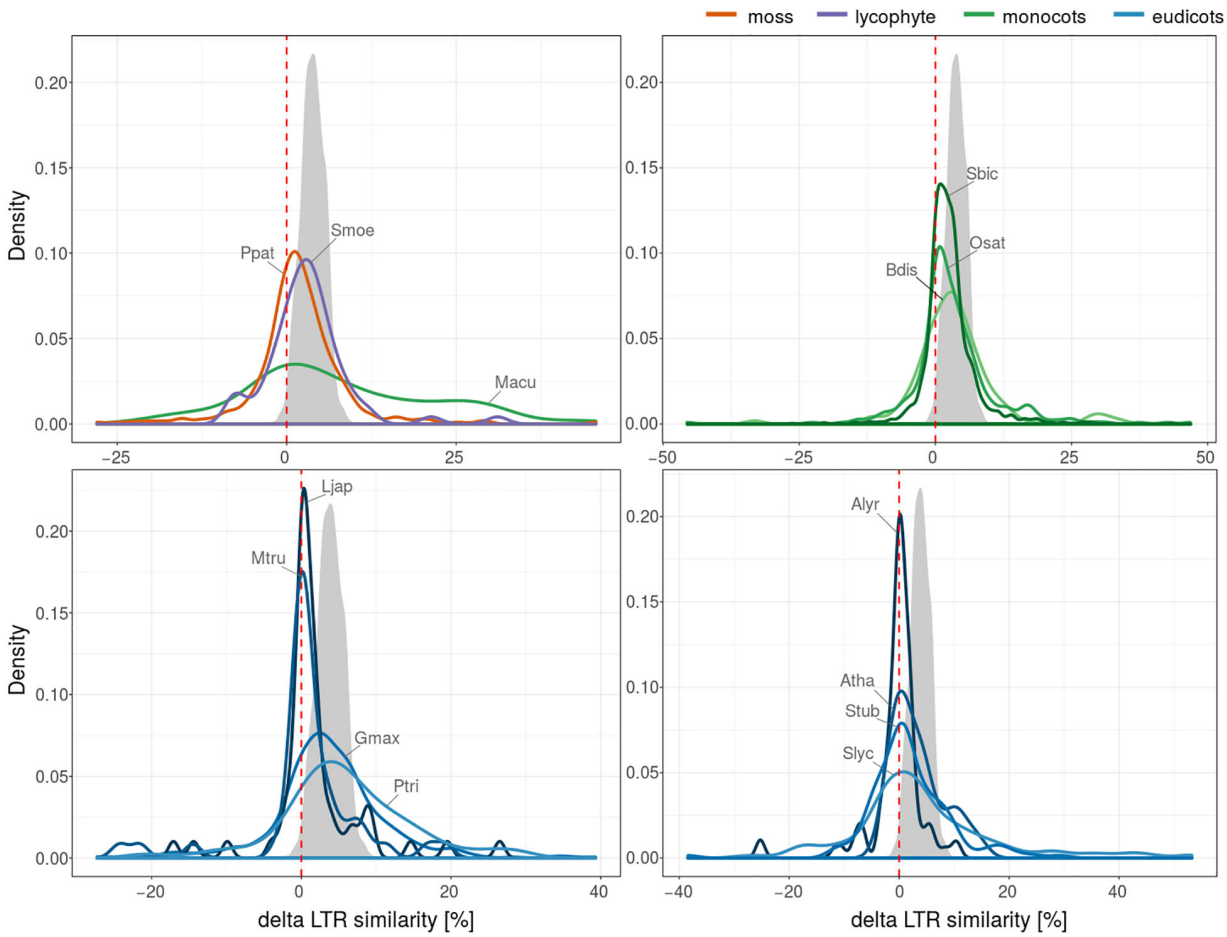

FIGURE 4 | 5′-3′ LTR similarity in nested and original LTR retrotransposons. Plant species were divided into four subfigures for better readability. Plotted values represent probability density function based on kernel density estimation. Number of LTR retrotransposons with values of delta LTR similarity (positive values correspond to higher LTR similarity of nested than original elements) for LTR retrotransposons in studied plant species (4126 nested/original pairs). The gray area shows simulated delta LTR similarity distribution under the assumption that only randomly distributed point mutations affected a pair of nested LTRs inserted into another pair with 97% LTR similarity at the time of insertion. The simulated nested structure was then further mutated with mutate.sh at 10% of positions, on average and delta LTR similarity was calculated.

the length of converted region differed among plant species. Gene conversion negatively correlated with genome size (Pearson's $r = -0.2420$ with p-value = 0.005175; **Figure 6A**). The length of converted regions (i) varied most often between 100 and 1000 bp and (ii) was higher in the case of gene conversion between LTRs of the same element (intra-element conversion) than for conversion between LTRs of different elements (inter-element conversion). The highest lengths of converted regions were found in *O. sativa*, *P. trichocarpa*, and *S. bicolor* (**Figure 6B**).

When we removed converted regions (predicted by GENECONV) from the LTRs, we found that the curve showing dependence of LTR similarities on LTR length was shifted to the left. This indicates that LTR similarities have decreased, leading to an increase in the LTR retrotransposon age estimates (**Figure 7**). When linear trendline was used, the slope after the removal of converted regions decreased (**Supplementary Figure S4**). However, the strong increase of LTR at the highest LTR similarities was not affected by the removal of converted regions. This possibly suggests that the increase of LTR similarity with length can be caused by other factors or by an unknown technical issue.

In order to better assess the strong increase of LTR similarity in the longer LTRs (even after the removal of converted regions), we performed the following simulation: we took set of LTRs with different length deposited in the Gypsy database (Gydb; $n = 413$), separately inserted the pairs of LTRs (imitating 5′ and 3′ LTR of retrotransposon) into the artificial genomes (always 1 Mb long) and mutated these genomes to a level ranging from 0.7 to 1.0. For each mutation level we found that the distribution of the longer LTRs were always more homogenous than the shorter ones (**Supplementary Figure S5** demonstrated mutation level 0.99). Such a finding suggests that this technical phenomenon, in addition to gene conversion, can explain the increase of LTR similarity in longer LTRs as observed in **Figure 5**.

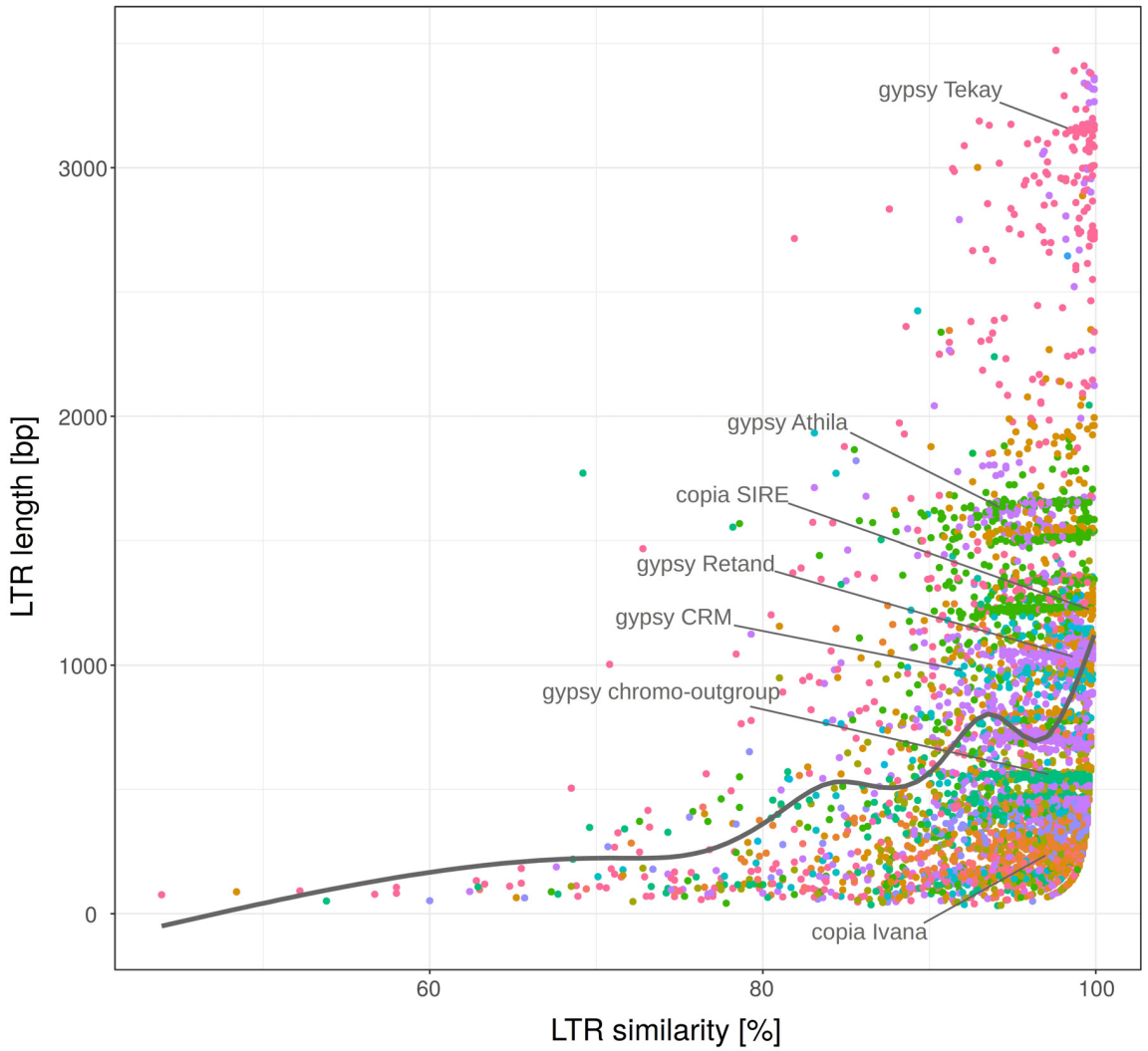

FIGURE 5 | LTR length plotted against 5′-3′ LTR similarity. LTR retrotransposon families (labeled with different colors) of fifteen plant species. Nested, original and non-nested LTR retrotransposons were analyzed together. Full set of 25,144 elements was randomly sampled to subset with n = 5000. Most abundant families are labeled within the plot.

The Relationship Between Gene Conversion and Ectopic Recombination

Our further analysis was motivated by the speculation that homogenization of retrotransposon families by gene conversion could accelerate ectopic recombination. Such a process would respond to family expansion threatening the host. Therefore, we measured in fifteen plant species the correlation between the intensity of gene conversion predicted by GENECONV and the ratio of solo LTR/FL. We found that the number of LTR retrotransposons exhibiting signs of gene conversion negatively correlated with the proportion of solo LTRs i.e., families exhibiting stronger signs of gene conversion had a lower proportion of solo LTRs (Pearson's $r = -0.5428$ with p-value = 1.784e-11; **Figure 8**). The remarkable position in the plot showed genomes of *Physcomitrella patens*, *Solanum lycopersicum*, and *S. tuberosum* hosting elements with high values of solo LTR/FL and low proportion of gene conversion (up to 20%). On the other hand, the genome of *Chlamydomonas reinhardtii* contained LTR retrotransposon strongly affected by gene conversion but having very low proportion of solo LTRs. Both extremes support the view that gene conversion and ectopic recombination interfere.

DISCUSSION

Our results show that (i) evolutionary dynamics of individual LTR retrotransposons differ among retrotransposon families and plant species, (ii) the commonly used LTR retrotransposon age estimation method based on LTR divergence is not absolute, probably due to the influence of gene conversion, (iii) families exhibiting signs of gene conversion less readily form solo LTRs,

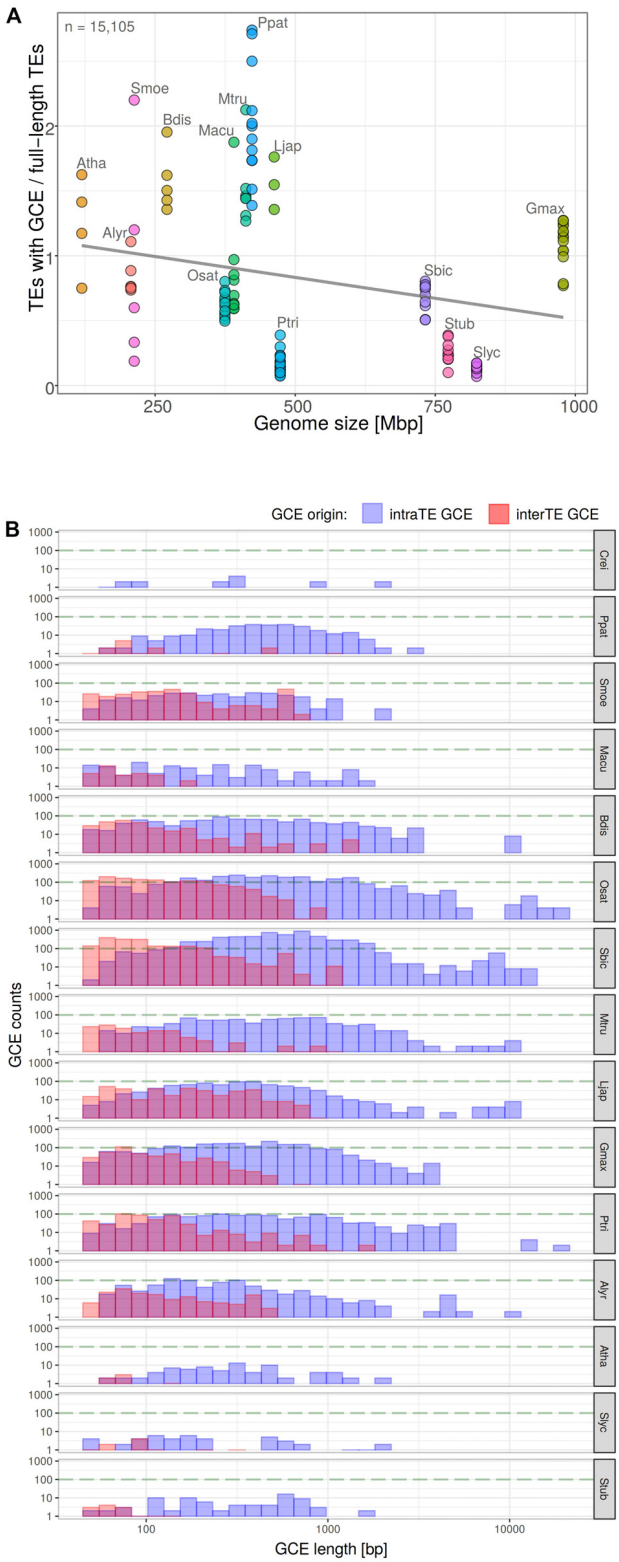

FIGURE 6 | Measurement of gene conversion events (GCE) along the LTR retrotransposons by GENECONV software. Proportion of LTR retrotransposons with GCE plotted against genome size in plant genomes **(A)** ($y = 1.16 - 0.000644x$; $R^2 = 0.059$; Pearson's $r = -0.2420$ with p-value = 0.005175). Total count of elements with GCE is indicated. Each filled circle corresponds to one chromosome, plant species are labeled by different colors. The GCE length distribution with respect to the origin of GCE donor LTR i.e., from the same element or from the other one **(B)**.

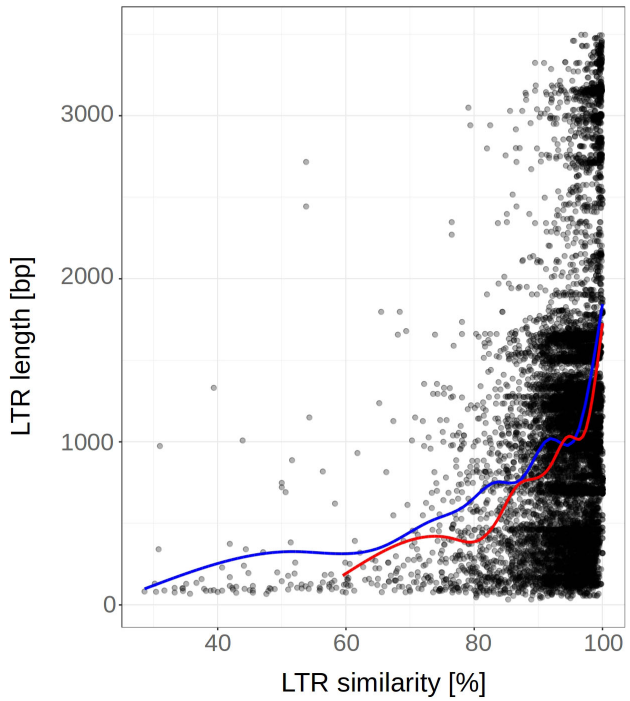

FIGURE 7 | LTR length plotted against the 5′-3′ LTR similarity before and after removal of gene converted regions (predicted by GENECONV). LTR retrotransposon of fifteen plant species (n = 5812). Smooth curves (fitted using "loess" method) are plotted for LTRs before and after GCE removal (red and blue color, respectively). Nested, original and non-nested LTR retrotransposons were analyzed together. The removal of converted regions from the LTRs has shifted the curve to the left resulting in an increase of LTR retrotransposon age estimates.

and (iv) the proportion of solo LTRs did not change with genome size, indicating a similar intensity of ectopic recombination in small and large genomes.

Our LTR retrotransposon age estimates were lower than estimates published by Bennetzen et al. (2005). This difference can be explained by the fact that (i) we used a much higher number of elements (hundreds and thousands compared to tens of elements in most species used by Bennetzen et al., 2005) and (ii) we used constant (1.3×10^{-8} in grasses) derived from grasses while Bennetzen et al. (2005) used the constant (6.5×10^{-9}) originating from maize (SanMiguel et al., 1998).

The age distribution of a range of LTR retrotransposon families in fifteen plant species indicates that retrotransposon activity differed among families, probably as a result of an interplay of various genomic and environmental factors. Such an observation is in accordance with the concept of the genome as an ecosystem of varied elements exhibiting a spectrum of interactions from parasitism via competition to collaboration. Nevertheless, despite the differences in age distribution patterns, some similarities of the expansion profiles in several LTR retrotransposon families of the same species were evident and could reflect stresses that a species underwent when selected retrotransposon families were simultaneously activated.

Some of our results are necessarily affected by technical issues. While we used reasonable settings of TE-greedy-nester and subsequent filtering for minimal full-length TE structure and TSDs as evidence of real insertions, these settings and filtering steps are currently notoriously error-prone and could affect our results. Also, the age estimates (**Table 1**) could be affected by the quality of genome assembly. Namely, the average age of LTR retrotransposons in Solanum species (tomato and potato plants) was higher compared to other analyzed species here. High number of phylogenetically older retroelements (e.g., Ty3/gypsy: chromo outgroup and Galadriel; **Figure 1**) was found also in genomes of algae and mosses (Neumann et al., 2019). This putatively false (higher) age determination could be explained by the worse quality of LTR retrotransposon assembly (e.g., when chimeric elements are assembled from different families resulting in their higher distance from the consensus). Our assumption is supported by recent report on lower quality of tomato assembly (Hosmani et al., 2019).

Our finding that LTR similarity depends not only on the retrotransposon age but also on the LTR length (**Figure 5**) could be partially explained by absence of older longer LTRs, since they are more prone to unequal recombination (Du et al., 2012). The potential involvement of other factors affecting LTR retrotransposon age estimation is also supported by the lower LTR similarity of nested elements compared to the pre-existing ones. Our analysis using GENECONV software predicting the presence of gene conversion indicates that this process is probably responsible for the limitations of the LTR divergence method.

Our results are in accordance with the finding of Cossu et al. (2017) who reported that the length of LTR and the whole

FIGURE 8 | Relationship between gene conversion and solo LTR formation. Dependence of gene conversion events (GCE) predicted by GENECONV plotted against the ratio of solo LTR/FL ($y = 5.11–2.23x$; $R^2 = 0.29$; Pearson's $r = –0.5428$ with p-value = 1.784e-11). Chromosomes with no records of GCE and/or solo LTRs were excluded. Total counts of solo LTRs and elements with GCE are indicated (n_1 and n_2, respectively). Each filled circle corresponds to one chromosome, plant species are labeled by different colors. The graph shows that chromosomes containing high proportions of gene converted element LTRs contain low proportions of solo LTRs.

LTR retrotransposons (rather than sequence similarity) appears to be a major determinant of the gene conversion frequency. We also showed that gene conversion negatively correlates with the formation of solo LTRs. Compared to Cossu et al. (2017) here we analyzed more plant genomes and more elements and used the whole elements retrieved from the complete genome instead of Illumina reads. The importance of LTR length in an intensity of gene conversion was previously proposed by Du et al. (2012) who showed that the ratio of solo LTR to complete LTR retrotransposons correlates with a number of element features, such as LTR length. The potential role of gene conversion in homogenization of transposable elements was suggested decades ago for yeast Ty elements (Roeder and Fink, 1982), primate SINE elements (Kass et al., 1995), and human Alu elements (Roy et al., 2000). Gene conversion of LTR retrotransposons was proposed to be stronger on non-recombining Y chromosomes than on other chromosomes (Kejnovsky et al., 2007). Gene conversion has also been observed in satellite DNA (Krzywinski et al., 2005) and ribosomal genes (Lim et al., 2000).

The non-allelic gene conversion among long terminal repeats has been studied in human endogenous retroviruses recently (Trombetta et al., 2016). The authors suggest that ectopic recombination among LTRs is rather common and could also take place between elements occupying different chromosomes. Here we show that gene conversion between intra-element LTRs is much more frequent than between LTRs of two different elements in plant LTR retrotransposons.

The negative correlation between gene conversion and solo LTR formation indicates that gene conversion does not accelerate ectopic recombination by homogenizing LTRs of the same elements, as we expected, but rather that both processes (gene conversion and ectopic recombination) probably are influential. Therefore, homologous LTRs susceptible to recombination events, are responsive to either ectopic recombination or gene conversion. Both processes are homology-driven and differ in whether or not they resolve in crossing-over.

The presence of gene conversion has almost certainly led to underestimations of LTR retrotransposon age in many studies using the LTR divergence method. Recently, Maumus and Quesneville (2014) cast doubt on the popular dating approach that only assesses the LTR divergence widely applied in plants and stressed the need to use alternative methods based on e.g., reconstruction of ancestral/consensus repeats established from several related species. These authors evidenced such an approach by providing a higher age estimation of TEs in *A. thaliana* (Maumus and Quesneville, 2014) compared to the LTR divergence method. Similarly, Giordano et al. (2007) recommended the use of the genome-wide defragmentation approach for the estimation of TE age providing chronological order of elements rather than the use of an older method based on divergence from a derived consensus (Jurka, 1998). Retrotransposon age underestimation obtained by the LTR divergence method also agrees with the conclusion that LTR retrotransposons in Drosophila are much younger than the host

species in which they reside (Bowen and McDonald, 2001). Taken together, the optimization of methods for LTR retrotransposon age estimation should be a subject of further research.

The extent of gene conversion can be affected not only by the LTR length but also possibly by the distance between LTRs, as was shown for duplicated genes (Ezawa et al., 2006), and especially by epigenetic factors such as e.g., chromatin structure (Cummings et al., 2007). Since reversely transcribed cDNA molecules are often used as templates in gene conversion (Doolittle, 1985; Derr and Strathern, 1993; Benovoy and Drouin, 2009), and RNA molecules participate in gene conversion (Doolittle, 1985; Derr et al., 1991; Derr, 1998), then even transcriptional activity of specific LTR retrotransposon families could contribute to such homogenization. Thus, the expression of genome, induced by environmental or endogenous factors, can change the genome structure by homogenization of repetitive DNA.

The interplay between gene conversion and ectopic recombination can oppose LTR retrotransposon amplifications and lead to genome size reduction. This way, gene conversion can fulfill an important regulatory role in genome repeat expansions and contractions as well as related genome rearrangements. Since the activity of transposable elements is epigenetically regulated (Fedoroff, 2012), both gene conversion and ectopic recombination may respond to environmental challenges and thus contribute to eukaryotic evolvability and a higher genome dynamism in plants (Kejnovsky et al., 2009).

CONCLUSION AND PERSPECTIVES

LTR retrotransposons have colonized plant genomes throughout the whole course of evolution. Estimation of LTR retrotransposon age is thus of great importance for the study of plant genome evolution as well as for understanding transposable element biology. Recent research indicates that the traditional age estimation method based on the LTR divergence has some limits, mostly due to the action of gene conversion. Here, we have extended the available knowledge and showed that (i) LTR similarity depends on LTR length and (ii) nested elements often have lower LTR similarity that pre-existing original ones. We have found regions in LTR with signs of gene conversion responsible for both phenomena. Negative correlation between the extent of gene conversion and the abundance of solo LTRs indicates that gene conversion probably interferes with the ectopic recombination between LTRs. Our findings demonstrate that the LTR divergence method should be used carefully keeping in mind the effect of other factors such as gene conversion. We conclude that more methods should be combined for a more reliable LTR retrotransposon age estimation, using e.g., retrotransposon family variability or mutual nesting of elements in order to achieve absolute chronology.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Phytozome 12.0 unmasked genomes; https://phytozome.jgi.doe.gov/pz/portal.html.

AUTHOR CONTRIBUTIONS

PJ and ML analyzed the data. EK and ML conceived the study. All authors wrote the manuscript, read and approved the final manuscript.

FUNDING

This research was supported by the Czech Science Foundation (grant 18-00258S).

ACKNOWLEDGMENTS

We thank Christopher Johnson for critical reading of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00644/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

Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., dePamphilis, C., et al. (2011). The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332, 960–963. doi: 10.1126/science.1203810

Bennetzen, J. L., Ma, J., and Devos, K. M. (2005). Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 95, 127–132. doi: 10.1093/aob/mci008

Benovoy, D., and Drouin, G. (2009). Ectopic gene conversions in the human genome. *Genomics* 93, 27–32. doi: 10.1016/j.ygeno.2008.09.007

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

Charles, M., Belcram, H., Just, J., Huneau, C., Viollet, A., Couloux, A., et al. (2008). Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics* 180, 1071–1086. doi: 10.1534/genetics.108.092304

Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., et al. (2010). Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell* 22, 1686–1701. doi: 10.1105/tpc.110.074187

Cossu, R. M., Casola, C., Giacomello, S., Vidalis, A., Scofield, D. G., and Zuccolo, A. (2017). LTR retrotransposons show low levels of unequal recombination and high rates of intraelement gene conversion in large plant genomes. *Genome Biol. Evol.* 9, 3449–3462. doi: 10.1093/gbe/evx260

Cummings, W. J., Yabuki, M., Ordinario, E. C., Bednarski, D. W., Quay, S., and Maizels, N. (2007). Chromatin structure regulates gene conversion. *PLoS Biol.* 5:e246. doi: 10.1371/journal.pbio.0050246

Derr, L. K. (1998). The involvement of cellular recombination and repair genes in RNA-mediated recombination in Saccharomyces cerevisiae. *Genetics* 148, 937–945.

Derr, L. K., and Strathern, J. N. (1993). A role of reverse transcription in gene conversion. *Nature* 361, 170–173. doi: 10.1038/361170a0

Derr, L. K., Strathern, J. N., and Garfinkel, D. J. (1991). RNA-mediated recombination in *S. cerevisiae*. *Cell* 67, 355–364. doi: 10.1016/0092-8674(91)90187-4

Devos, K. M., Brown, J. K. M., and Bennetzen, J. L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 12, 1075–1079. doi: 10.1101/gr.132102

D'Hont, A., Denoeud, F., and Aury, J. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217. doi: 10.1038/nature11241

Doolittle, W. F. (1985). RNA-mediated gene conversion? *Trends Genet.* 1, 64–65. doi: 10.1016/0168-9525(85)90028-9

Du, J., Tian, Z., Hans, C. S., Laten, H. M., Cannon, S. B., Jackson, S. A., et al. (2012). Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insight from genome-wide analysis and multi-specific comparison. *Plant J.* 63, 584–598. doi: 10.1111/j.1365-313X.2010.04263.x

Ezawa, K., OOta, S., and Saitou, N. (2006). Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol. Biol. Evol.* 23, 927–940. doi: 10.1093/molbev/msj093

Fedoroff, N. V. (2012). Transposable elements, epigenetics, and genome evolution. *Science* 338, 758–768. doi: 10.1126/science.338.6108.758

Feschotte, C., Jiang, N., and Wessler, S. R. (2002). Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 3, 329–341. doi: 10.1038/nrg793

Gaut, B., Morton, B. R., McCaig, B. C., and Clegg, M. T. (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10274–10279. doi: 10.1073/pnas.93.19.10274

Giordano, J., Ge, Y., Gelfand, Y., Abrusan, G., Benson, G., and Warburton, P. E. (2007). Evolutionary history of mammalian transposons determined by genome-wide defragmentation. *PLoS Comp. Biol.* 3:e137. doi: 10.1371/journal.pcbi.0030137

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

Grandbastien, M.-A. (1998). Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* 3, 181–187. doi: 10.1016/S1360-1385(98)01232-1

Guindon, S., and Gascuel, O. (2003). A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Hirochika, H. (1997). Retrotransposons of rice: their regulation and use for genome analysis. *Plant Mol. Biol.* 35, 231–240. doi: 10.1023/A:100577470

Hosmani, P. S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L. V., Schijlen, E., et al. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv* [Preprint]. doi: 10.1101/767764

International Brachypodium Initiative. (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* 463, 763–768. doi: 10.1038/nature08747

Jurka, J. (1998). Repeats in genomic DNA: mining and meaning. *Curr. Opin. Struct. Biol.* 8, 333–337. doi: 10.1016/s0959-440x(98)80067-5

Kass, D. H., Batzer, M. A., and Deininger, P. L. (1995). Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. *Mol. Cell. Biol.* 15, 19–25. doi: 10.1128/mcb.15.1.19

Kejnovsky, E., Hawkins, J. S., and Feschotte, C. (2012). "Plant transposable elements: biology and evolution," in *Diversity of Genomes in Plants*, eds J. F. Wendel, J. Greilhuber I, J. Leitch, and J. Dolezel (Berlin: Springer), 17–34. doi: 10.5808/GI.2014.12.3.87

Kejnovsky, E., Hobza, R., Kubat, Z., Widmer, A., Marais, G. A. B., and Vyskot, B. (2007). High intrachromosomal similarity of retrotransposon long terminal repeats: evidence for homogenization by gene conversion on plant sex chromosomes? *Gene* 390, 92–97. doi: 10.1016/j.gene.2006.10.007

Kejnovsky, E., Leitch, I., and Leitch, A. (2009). Contrasting evolutionary dynamics between angiosperm and mammalian genomes. *Trends Ecol. Evol.* 24, 572–582. doi: 10.1016/j.tree.2009.04.010

Kijima, T. E., and Innan, H. (2009). On the estimation of the insertion time of LTR retrotransposable elements. *Mol. Biol. Evol.* 27, 896–904. doi: 10.1093/molbev/msp295

Kim, T.-M., Hong, S.-J., and Rhyu, M.-G. (2004). Periodic explosive expansion of human retroelements associated with the evolution of the hominoid primate. *J. Korean Med. Sci.* 19, 177–185. doi: 10.3346/jkms.2004.19.2.177

Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498. doi: 10.1093/oxfordjournals.molbev.a026248

Krzywinski, J., Sangare, D., and Besansky, N. J. (2005). Satellite DNA from the Y chromosome of the malarial vector Anopheles gambiae. *Genetics* 169, 185–196. doi: 10.1534/genetics.104.034264

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090

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

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lexa, M., Lapar, R., Jedlicka, P., Vanat, I., Cervenansky, M., and Kejnovsky, E. (2018). "TE-nester: a recursive software tool for structure-based discovery of nested transposable elements," in *Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine*, Madrid, 2776–2778. doi: 10.1109/BIBM.2018.8621071

Li, W. (1997). *Molecular Evolution*. Sunderland, MA: Sinauer.

Lim, K. Y., Kovarik, A., Matyasek, R., Bezdek, M., Lichtenstein, C. P., and Leitch, A. R. (2000). Gene conversion of ribosomal DNA in Nicotiana tabacum is associated with undermethylated, decondensed and probably active gene units. *Chromosoma* 109, 161–172. doi: 10.1007/s004120050424

Llorens, C., Futami, R., Covelli, L., Dominguez-Escriba, L., Viu, J. M., Tamarit, D., et al. (2011). The gypsy database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* 39(Suppl. 1), D70–D74. doi: 10.1093/nar/gkq1061

Ma, J., and Bennetzen, J. L. (2004). Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12404–12410. doi: 10.1073/pnas.0403715101

Ma, J., Devos, K. M., and Bennetzen, J. L. (2004). Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* 14, 860–869. doi: 10.1101/gr.1466204

Macas, J., Novak, P., Pellicer, J., Cizkova, J., Koblizkova, A., Neumann, P., et al. (2015). In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. *PLoS One* 10:e0143424. doi: 10.1371/journal.pone.0143424

Mansai, S. P., and Innan, H. (2010). The power of the methods for detecting interlocus gene conversion. *Genetics* 184, 517–527. doi: 10.1534/genetics.109.111161

Mansai, S. P., Kado, T., and Innan, H. (2011). The rate and tract length of gene conversion between duplicated genes. *Genes* 2, 313–331. doi: 10.3390/genes2020313

Martins, H., and Villesen, P. (2011). Improved integration time estimation of endogenous retroviruses with phylogenetic data. *PLoS One* 6:e14745. doi: 10.1371/journal.pone.0014745

Maumus, F., and Quesneville, H. (2014). Ancestral repeats have shaped epigenome and genome composition for millions of years in Arabidopsis thaliana. *Nat. Commun.* 5:4104. doi: 10.1038/ncomms5104

McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., et al. (2017). The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354. doi: 10.1111/tpj.13781

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., et al. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318, 245–250. doi: 10.1126/science.1143609

Neumann, P., Novak, P., Hostakova, N., and Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA* 10:1. doi: 10.1186/s13100-018-0144-1

Ou, S., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., et al. (2007). The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* 35, D883–D887. doi: 10.1093/nar/gkl976

Pereira, V. (2004). Insertion bias and purifying selection of retrotransposons in the Arabidopsis thaliana genome. *Genome Biol.* 5:R79. doi: 10.1186/gb-2004-5-10-r79

Pereira, V. (2008). Automated paleontology of repetitive DNA with REANNOTATE. *BMC Genomics* 9:614. doi: 10.1186/1471-2164-9-614

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

Rawat, V., Abdelsamad, A., Pietzenuk, B., Seymour, D. K., Koenig, D., Weigel, D., et al. (2015). Improving the annotation of *Arabidopsis lyrata* using RNA-seq data. *PLoS One* 10:e0137391. doi: 10.1371/journal.pone.0137391

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276–277. doi: 10.1016/s0168-9525(00)02024-2

Roeder, G. S., and Fink, G. R. (1982). Movement of yeast transposable elements by gene conversion. *Proc. Natl. Acad. Sci. U.S.A.* 79, 5621–5625. doi: 10.1073/pnas.79.18.5621

Roy, A. M., Carrol, M. L., Nguyen, S. V., Salem, A.-H., Oldridge, M., Wilkie, A. O. M., et al. (2000). Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res.* 10, 1485–1495. doi: 10.1101/gr.152300

SanMiguel, P., Gaut, B., Tikhonov, A., Nakajima, Y., and Bennetzen, J. L. (1998). The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20, 43–45. doi: 10.1038/1695

SanMiguel, P. J., Ramakrishna, W., Bennetzen, J. L., Busso, C., and Dubcovsky, J. (2002). Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5Am. *Funct. Integr. Genomics* 2, 70–80. doi: 10.1007/s10142-002-0056-4

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008). Genome structure of the legume, *Lotus japonicus*. *DNA Res.* 15, 227–239. doi: 10.1093/dnares/dsn008

Sawyer, S. A. (1999). *GENECONV: A Computer Package for the Statistical Detection of Gene Conversion*. Washington, DC: University in St. Louis.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534

Sharma, S. K., Bolser, D., de Boer, J., Sønderkær, M., Amoros, W., Carboni, M. F., et al. (2013). Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3*, 2031–2047. doi: 10.1534/g3.113.007153

Shirazu, K., Schulman, A. H., Lahaye, T., and Schulze-Lefert, P. (2000). A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res.* 10, 908–915. doi: 10.1101/gr.10.7.908

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., et al. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* 15:312. doi: 10.1186/1471-2164-15-312

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119

Trombetta, B., Fantini, G., D'Atanasio, E., Sellitto, D., and Cruciani, F. (2016). Evidence of extensive non-allelic gene conversion among LTR elements in the human genome. *Sci. Rep.* 6:28710. doi: 10.1038/srep28710

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691

Vitte, C., and Panaud, O. (2003). Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol. Biol. Evol.* 20, 528–540. doi: 10.1093/molbev/msg055

Vitte, C., and Panaud, O. (2005). LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet. Genome Res.* 110, 91–107. doi: 10.1159/000084941

Wicker, T., Grundlach, H., Spannagl, M., Uauy, C., Borrill, P., Ramirez-Gonzales, R. H., et al. (2018). Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biol.* 19:103. doi: 10.1186/s13059-018-1479-0

Wicker, T., and Keller, B. (2007). Genome-wide comparative analysis of *copia* retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. *Genome Res.* 17, 1072–1081. doi: 10.1101/gr.6214107

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.-D., Dubcovsky, J., et al. (2003). Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. *Plant Cell* 15, 1186–1197. doi: 10.1105/tpc.011023

Xu, Y., and Du, J. (2014). Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (*Solanum lycopersicum*) plants. *Plant J.* 80, 582–591. doi: 10.1111/tpj.12656

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286

Zhang, Q. J., and Gao, L. Z. (2017). Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* species. *G3* 7, 1875–1885. doi: 10.1534/g3.116.037572

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 May 2020
doi: 10.3389/fpls.2020.00514

Chromosomal Evolution and Apomixis in the Cruciferous Tribe Boechereae

*Terezie Mandáková[1], Petra Hloušková[1], Michael D. Windham[2], Thomas Mitchell-Olds[2], Kaylynn Ashby[3], Bo Price[3], John Carman[3] and Martin A. Lysak[1]**

[1] CEITEC, Masaryk University, Brno, Czechia, [2] Department of Biology, Duke University, Durham, NC, United States, [3] Plants, Soils, and Climate Department, Utah State University, Logan, UT, United States

The mustard family (Brassicaceae) comprises several dozen monophyletic clades usually ranked as tribes. The tribe Boechereae plays a prominent role in plant research due to the incidence of apomixis and its close relationship to *Arabidopsis*. This tribe, largely confined to western North America, harbors nine genera and c. 130 species, with >90% of species belonging to the genus *Boechera*. Hundreds of apomictic diploid and triploid *Boechera* hybrids have spurred interest in this genus, but the remaining Boechereae genomes remain virtually unstudied. Here we report on comparative genome structure of six genera (*Borodinia*, *Cusickiella*, *Phoenicaulis*, *Polyctenium*, *Nevada*, and *Sandbergia*) and three *Boechera* species as revealed by comparative chromosome painting (CCP). All analyzed taxa shared the same seven-chromosome genome structure. Comparisons with the sister Halimolobeae tribe (n = 8) showed that the ancestral Boechereae genome (n = 7) was derived from an older n = 8 genome by descending dysploidy followed by the divergence of extant Boechereae taxa. As tribal divergence post-dated the origin of four tribe-specific chromosomes, it is proposed that these chromosomal rearrangements were a key evolutionary innovation underlaying the origin and diversification of the Boechereae in North America. Although most Boechereae genera exhibit genomic conservatism, intra-tribal cladogenesis has occasionally been accompanied by chromosomal rearrangements (particularly inversions). Recently, apomixis was reported in the Boechereae genera *Borodinia* and *Phoenicaulis*. Here, we report sexual reproduction in diploid *Nevada*, diploid *Sandbergia*, and tetraploid *Cusickiella* and aposporous apomixis in tetraploids of *Polyctenium* and *Sandbergia*. In sum, apomixis is now known to occur in five of the nine Boechereae genera.

Keywords: apomixis, apospory, autopolyploidy, Cruciferae, descending dysploidy, karyotype evolution, North America, speciation

OPEN ACCESS

Edited by:
Steven Dodsworth, University of Bedfordshire, United Kingdom

Reviewed by:
Ana Paula Moraes, Universidade Federal do ABC, Brazil
Aretuza Sousa Dos Santos, Ludwig Maximilian University of Munich, Germany

****Correspondence:***
Martin A. Lysak
martin.lysak@ceitec.muni.cz;
lysak@sci.muni.cz

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

Received: *13 February 2020*
Accepted: *06 April 2020*
Published: *28 May 2020*

Citation:
Mandáková T, Hloušková P, Windham MD, Mitchell-Olds T, Ashby K, Price B, Carman J and Lysak MA (2020) Chromosomal Evolution and Apomixis in the Cruciferous Tribe Boechereae. Front. Plant Sci. 11:514. doi: 10.3389/fpls.2020.00514

INTRODUCTION

Geographically well-defined clades provide ideal study systems for understanding the role of whole-genome duplications (WGDs, polyploidy) and chromosomal rearrangements in speciation and diversification. Frequently, a group of species confined to an island, mountain range, or (sub)continent is assumed to have originated in this region, perhaps following

an earlier dispersal event from another part of the world (e.g., Linder and Barker, 2014; Givnish et al., 2016). With the advent of molecular phylogenetics, in many cases, the inferred monophyly of a group has confirmed its geographical determinant and helped to elucidate its origin as well as directionality of later migrations and dispersals (e.g., Cowie and Holland, 2008; Ogutcen and Vamosi, 2016; Huang et al., 2018; Carter et al., 2019; Kim et al., 2019). A geographically restricted clade might be supported by different synapomorphies, such as morphological traits, specific metabolic pathways, pollination syndromes, or a shared WGD, some falling in the category of rare genomic changes (RGCs, Rokas and Holland, 2000). Structural chromosomal changes may underlie incipient reproductive isolation inducing species splits and the evolution of separate gene pools, i.e., cladogenesis (Faria and Navarro, 2010). Dysploidal (i.e., chromosome number changes caused by fusions and fissions) as well as non-dysploidal (i.e., deletions, duplications, inversions, and translocations) chromosomal rearrangements can modify recombination frequency, gene expression, the duration of cellular processes (replication, mitosis, and meiosis), and the degree of infertility of heterozygous hybrids. Thus, some chromosomal rearrangements may precipitate lineage splitting yet occur within a monophyletic clade (Freyman and Höhna, 2017).

The economically important mustard family (3977 species in 351 genera, BrassiBase[1], accessed on February 1, 2020) radiated into four (Franzke et al., 2011) to six (Huang et al., 2016) lineages or super-tribes ∼23 million years ago (Hohmann et al., 2015). These lineages have been divided into 52 monophyletic tribes (BrassiBase) ranging in size from the monospecific Shehbazieae (German and Friesen, 2014) to the Arabideae, which harbors more than 390 species (Jordon-Thaden et al., 2013; Karl and Koch, 2013). Many crucifer tribes do not differ in their basal or ancestral chromosome numbers. For example, tribes of lineage II/B and lineage III/E have the same number of ancestrally shared linkage groups and the same dysploid chromosomal rearrangements (Mandáková and Lysak, 2008; Mandáková et al., 2017a). By contrast, tribes of lineage I/A, such as Boechereae (x = 7), Descurainieae (x = 7), Erysimeae (mostly x = 7), Turritideae (x = 6), and Yinshanieae (x = 6 and 7) (Warwick and Al-Shehbaz, 2006; BrassiBase), appear to represent tribes that originated after independent reductions of the ancestral chromosome number (n = 8) to n = 7 and n = 6. None of the diploid Brassicaceae tribes with a clade-specific descending dysploidy have been investigated genomically, so it remains unclear whether intra-tribal diversification (i.e., speciation and origin of new genera) has involved non-dysploidal chromosomal rearrangements.

Here we focus on the tribe Boechereae which harbors c. 130 species. The vast majority of Boechereae taxa occurs only in North America, with one of these extending to Greenland and three species being endemic to the Russian Far East (Alexander et al., 2013; Doudkin and Volkova, 2013). Molecular studies (Beilstein et al., 2010; Nikolov et al., 2019) using various chloroplast and nuclear gene markers support the Boechereae (with a shared chromosome base number of x = 7) as a monophyletic clade sister to the New World tribe Halimolobeae (x = 8; 39 species in five genera, Al-Shehbaz, 2012). Alexander et al. (2013) recognized nine genera of Boechereae, seven of which (*Anelsonia* J. F. Macbride and Payson, *Cusickiella* Rollins, *Nevada* N. H. Holmgren, *Phoenicaulis* Nuttall, *Polyctenium* Greene, *Sandbergia* Greene, and *Yosemitea* P. J. Alexander and Windham) are mono- or bispecific, and, except for *Sandbergia whitedii* (Piper) Greene, restricted to the western United States. *Boechera* Á. Löve and D. Löve is by far the most diverse genus of the tribe, largely confined to the western part of the North American continent (Alexander et al., 2013). One group of eight species often assigned to *Boechera* was transferred to the genus *Borodinia* N. Busch by Alexander et al. (2013). This species group has the most discrete geographic range, apparently restricted to eastern North America and the Russian Far East [*Borodinia macrophylla* (Turcz.) O. E. Schulz]. Despite their largely allopatric distributions, *Boechera* and *Borodinia* species have hybridized in nature to produce one widespread sexual tetraploid and a series of apomictic triploids and tetraploids that erase any morphological distinctions between the two genera (Windham et al., 2014). When subsumed within *Boechera*, these lineages are informally designated the "western" and "eastern" clades, respectively.

The species now assigned to *Boechera* (x = 7) were originally included in *Arabis* L. (tribe Arabideae; x = 8), but a series of molecular analyses (Koch et al., 2000; Beilstein et al., 2010; Nikolov et al., 2019) has shown that these genera belong to different major lineages of Brassicaceae. *Boechera* is phylogenetically closely related to the model genus *Arabidopsis* Heynh. (**Figure 1A**) and is best known for its classic agamic complex consisting of numerous, morphologically diverse, facultative, and obligate gametophytic apomicts. These are generally of hybrid origin, arising from a diverse array of sexual diploids with more restricted habitats. The genus is named for Danish botanist Tyge Böcher, who first documented apomixis in *Boechera holboellii* (Hornem.) Á. Löve and D. Löve (Böcher, 1951). The relatively close relationship of these species to *Arabidopsis*, combined with its diversity of ploidies (Alexander et al., 2015) and apomixis types, at both the diploid and polyploid levels (Carman et al., 2019), have made the genus a major focus for apomixis research (e.g., Naumova et al., 2001; Schranz et al., 2005; Lee et al., 2017; Kliver et al., 2018; Rojek et al., 2018; Brukhin et al., 2019). Until recently, apomixis within Boechereae was thought to be confined to the large genus *Boechera*. However, Mandáková et al. (2020) documented the occurrence of apomixis at the diploid, triploid, and tetraploid levels in one of the smaller genera of Boechereae (*Phoenicaulis*), raising the possibility that apomixis might also occur in other Boechereae genera.

In flowering plants, apomixis can be defined as asexual seed formation where clonal embryos originate either from unreduced eggs produced in unreduced female gametophytes (gametophytic apomixis) or from somatic cells of the ovule wall without an intervening unreduced gametophyte generation (sporophytic apomixis). In sporophytic apomixis, a reduced gametophyte forms, which supports the clonal embryo while it develops. The reduced gametophyte may or may not contain a sexually derived embryo (Asker and Jerling, 1992; Hand and Koltunow, 2014). Gametophytic apomixis is a prominent mode of reproduction

[1]https://brassibase.cos.uni-heidelberg.de/

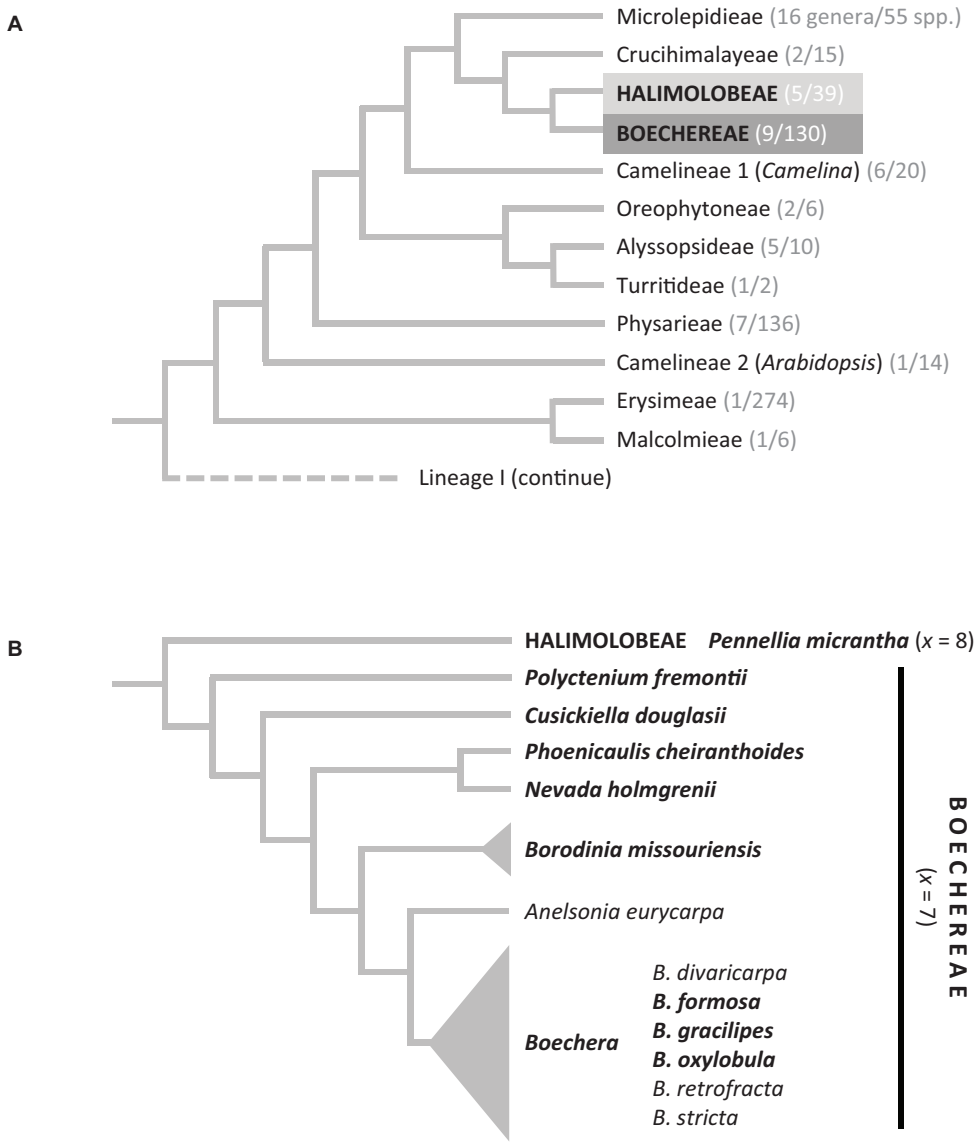

FIGURE 1 | (A) Phylogenetic position of the Boechereae in the Brassicaceae based on Nikolov et al. (2019). In parentheses, the number of genera and species is given based on data from BrassiBase (https://brassibase.cos.uni-heidelberg.de/). **(B)** Generic relationships within the Boechereae based on Beilstein et al. (2010) (*Sandbergia* and *Yosemitea* not shown).

in *Boechera* (Böcher, 1951; Roy, 1995; Naumova et al., 2001; Schranz et al., 2005), where it has greatly increased the diversity of genotypes and phenotypes by stabilizing the products of reticulate evolution (Beck et al., 2012; Carman et al., 2019). Three types of gametophytic apomixis occur in *Boechera*, and these are differentiated based on where the unreduced gametophyte forms: (*i*) if from a megasporocyte [megaspore mother cell (MMC)], it is referred to as Antennaria type diplospory, (*ii*) if from an apomeiotic dyad member of a first division meiotic restitution event, it is referred to as Taraxacum type diplospory, and (*iii*) if from a nucellar or parietal cell, it is referred to as Hieracium type apospory (Carman et al., 2019). The Antennaria type appears to be an oddity in *Boechera* that has been observed only rarely in plants that otherwise reproduce by Taraxacum type diplospory. In contrast, Taraxacum type diplospory and apospory are more commonly encountered in natural populations of *Boechera* than is sexual reproduction (Carman et al., 2019).

In the light of the fragmentary knowledge of genome evolution and reproductive modes in the Boechereae, we embarked on comparative cytogenetic and embryological analysis of several taxa representing the tribal diversity. We followed several aims: (1) To expand cytogenomic sampling of the species and genera of Boechereae to provide a more complete understanding of chromosomal evolution in the tribe, (2) to determine whether diversification within the tribe was accompanied by clade-specific chromosomal rearrangement, (3) to test whether an ancestral

$n = 7$ genome inferred for *Boechera* and *Phoenicaulis* (Mandáková et al., 2015b, 2020) provides an accurate reconstruction of the ancestral genome of the tribe as a whole, and (4) to gain insights into the reproductive modes of previously unsampled Boechereae genera by conducting embryological analyses on most of the species analyzed cytogenomically.

MATERIALS AND METHODS

Species Analyzed

Ten Boechereae species were selected to span the phylogenetic diversity of the tribe (**Figure 1**). These included *Boechera formosa* (Greene) Windham and Al-Shehbaz ($2n = 14$), *Boechera gracilipes* (Greene) Dorn ($2n = 14$), *Boechera oxylobula* (Greene) W. A. Weber ($2n = 14$), *Borodinia missouriensis* (Greene) P. J. Alexander and Windham ($2n = 14$), *Cusickiella douglasii* (A. Gray) Rollins ($2n = 28$), *Nevada holmgrenii* (Rollins) N. H. Holmgren ($2n = 14$), *Phoenicaulis cheiranthoides* ($2n = 14, 21, 28$), *Polyctenium fremontii* (S. Watson) Greene ($2n = 28$), *Sandbergia perplexa* (L. F. Henderson) Al-Shehbaz ($2n = 14$), and *S. whitedii* ($2n = 21$). As the tribe Halimolobeae was repeatedly retrieved as a sister clade to Boechereae (e.g., Beilstein et al., 2010; Couvreur et al., 2010; Alexander et al., 2013), *Pennellia micrantha* (A. Gray) Nieuwland ($2n = 16$) was selected to represent the Halimolobeae clade outgroup. The origins of the analyzed populations are listed in **Supplementary Table S1**; multiple individuals were analyzed from each population.

The analyzed plants were either collected in the wild or grown in a growth chamber from seeds collected in the wild. Young inflorescences of the analyzed plants were collected and fixed in freshly prepared fixative (ethanol: acetic acid, 3: 1) overnight, transferred to 70% ethanol and stored at −20°C.

Chromosome Preparation

Chromosome spreads from fixed young flower buds containing immature anthers were prepared according to published protocols (Lysak and Mandáková, 2013; Mandáková and Lysak, 2016a). Chromosome preparations were treated with 100 μg/ml RNase in 2 × sodium saline citrate (SSC; 20 × SSC: 3 M sodium chloride, 300 mM trisodium citrate, pH 7.0) for 60 min and with 0.1 mg/ml pepsin in 0.01 M HCl at 37°C for 5 min, then post-fixed in 4% formaldehyde in distilled water, and dehydrated in an ethanol series (70, 90, and 100%, 2 min each).

DNA Probes

The BAC clone T15P10 (AF167571) of *Arabidopsis thaliana* (L.) Heynh. bearing 35S rRNA gene repeats was used for *in situ* localization of nucleolar organizer regions (NORs), and the *A. thaliana* clone pCT4.2 (M65137), corresponding to a 500 bp 5S rDNA repeat, was used for localization of 5S rDNA loci. For Comparative Chromosome Painting (CCP), 674 chromosome-specific BAC clones of *A. thaliana* (The Arabidopsis Information Resource, TAIR[2]) were used to establish contigs corresponding to the 22 genomic blocks (GBs) and eight chromosomes (AK1–AK8) of the Ancestral Crucifer Karyotype (ACK; Lysak et al., 2016). See **Supplementary Tables S2–S8** for the list of BAC clones used to identify the 22 GBs on chromosomes of the Boechereae species. To determine and characterize inversions and split GBs, some BAC contigs were split into smaller subcontigs and differentially labeled (e.g., Aa, Ab, Ca, Cb, see **Supplementary Tables S3–S8**). All DNA probes were labeled with home-made biotin-dUTP, digoxigenin-dUTP, or Cy3-dUTP by nick translation as described by Mandáková and Lysak (2016b).

Comparative Chromosome Painting (CCP)

DNA probes were pooled to follow the design of a given experiment, ethanol precipitated, dried, and dissolved in 20 μl of 50% formamide and 10% dextran sulfate in 2 × SSC. The 20 μl of the dissolved probe were pipetted on a chromosome-containing slide and immediately denatured on a hot plate at 80°C for 2 min. Hybridization was carried out in a moist chamber at 37°C overnight. Post-hybridization washing was performed in 20% formamide in 2 × SSC at 42°C three times (5 min each time). Hybridized probes were visualized either as the direct fluorescence of Cy3-dUTP or through fluorescently labeled antibodies against biotin-dUTP and digoxigenin-dUTP following Mandáková and Lysak (2016b). Chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI, 2 μg/ml) in Vectashield antifade. Fluorescence signals were analyzed and photographed using a Zeiss Axioimager epifluorescence microscope with a CoolCube camera (MetaSystems). Images were acquired separately for all four fluorochromes using appropriate excitation and emission filters (AHF Analysentechnik). The four monochromatic images were pseudocoloured, merged, and cropped using Photoshop CS (Adobe Systems) and ImageJ (National Institutes of Health).

In silico Sequence Analysis

Boechera stricta (Graham) Al-Shehbaz (v1.2; Lee et al., 2017), *Arabidopsis lyrata* (L.) O'Kane and Al-Shehbaz (v2.1; Hu et al., 2011), and *A. thaliana* (TAIR 10) genome assemblies and annotations were downloaded from the Phytozome webpage[3]. Inter-genome collinearity was analyzed by SynOrths, identifying whether two homeologous genes are a conserved syntenic pair based on their sequence similarity and the support of homeologous flanking genes (Cheng et al., 2012).

Embryological Analyses

Clusters of pre-anthesis staged floral buds were fixed in 3:1 fixative for 48 h and stored in 70% EtOH. Ovaries were excised, cleared, measured, and mounted following Mandáková et al. (2020). An Olympus (Center Valley, PA, United States) BX53 microscope with differential interference contrast (DIC) optics and equipped with a DP74 digital camera with cellSens Dimension 1 software (Olympus) was used to investigate parietal cell, MMC, and female gametophyte origins.

[2]http://www.arabidopsis.org

[3]https://phytozome.jgi.doe.gov/pz/portal.html

RESULTS

Based on the ACK, and on previously analyzed Boechereae species (Mandáková et al., 2015a, 2020), detailed comparative cytogenetic maps were constructed by CCP for each of the 10 Boechereae species and for *P. micrantha* (**Figure 2** and **Supplementary Tables S2–S8**).

The Outgroup *Pennellia micrantha* Genome Structurally Mirrors the ACK

Comparative chromosome painting in *P. micrantha* ($2n$ = 16, Halimolobeae) was successful in identifying all 22 conserved GBs making up the eight chromosomes (Hal1–Hal8, **Figures 2**, **3** and **Supplementary Table S2**). The ACK-like *Pennellia* genome further corroborrated the earlier assumption (Mandáková et al., 2015b, 2020) that the Most Recent Common Ancestor (MRCA) of Boechereae and Halimolobeae had eight chromosomes and structurally resembled the ancestral genome of crucifer lineage I.

Overall Structural Stasis of Boechereae Genomes

Comparative chromosome painting with painting probes designed according to the structure of the seven linkage groups in *Boechera* (Mandáková et al., 2015a) and *Phoenicaulis* (Mandáková et al., 2020), were effective in identifying all seven or 14 chromosome pairs among the 10 Boechereae species analyzed. All 10 genomes had a very similar organization (**Figure 2**), except for a few species-specific chromosomal rearrangements (see below). The overall structural genome similarity among different Boechereae genera allowed us to reconstruct the genome of the MRCA of Boechereae.

Ancestral Boechereae Genome

By comparing the 10 Boechereae genomes studied herein with those of three diploid *Boechera* taxa (Mandáková et al., 2015b), *Phoenicaulis* (Mandáková et al., 2020), and *P. micrantha*, we inferred the ancestral Boechereae genome with seven pairs of chromosomes (Boe1–Boe7, **Figures 2**, **3**). Three of these pairs (Boe4, Boe6, and Boe7) retained their ancestral structure as in the ACK, or Halimolobeae, whereas four pairs (Boe1–Boe3 and Boe5) are specific to the Boechereae genomes (**Figures 2–4** and **Supplementary Table S3**).

CCP chromosome painting analyses allowed us to reconstruct the origin of the four Boechereae-specific chromosomes. The origin of chromosomes Boe1 and Boe2 (**Figure 4A**) most likely included an initial 0.52-Mb pericentric inversion on ancestral chromosome AK1 with breakpoints within GBs A [between BAC clone F13B4 (At1g13620) and T16N11 (At1g15410)] and C [between F8L10 (At1g53170) and F12M16 (At1g53160)]. The size of this inversion and other documented rearrangements were inferred from the physical length (Mb) of *A. thaliana* BAC contigs spanning these chromosome regions. The inversion-bearing AK1 chromosome underwent a whole-arm translocation with paleochromosome AK2 resulting in chromosomes Boe1 (GBs Aa, Ca, and D) and Boe2 (Cb, Ab, B, and E). Chromosome Boe3 (F, G, W, and X) originated by a whole-arm translocation between paleochromosomes AK3 and AK8 (**Figure 4B**). The second translocation chromosome (GBs V and H) was involved in an end-to-end translocation with chromosome AK5, mediating the chromosome number reduction (8 → 7) in Boechereae. The collinearity of GBs K-L and M-N and the absence of the original centromere suggest that the "chromosome fusion" was accompanied or followed by a removal of the AK5 paleocentromere (**Figure 4B**). Remnants of the AK5 paleocentromere, apparent as heterochromatic knobs and/or unpainted chromosome segments, were not observed in any of the analyzed species (**Figures 3**, **4B**).

Clade-Specific Chromosomal Rearrangements

The ancestral Boechereae genome remained conserved in the diploid *B. gracilipes* (**Figures 2**, **3** and **Supplementary Table S3**) and the tetraploid *P. fremontii* (**Figure 2** and **Supplementary Table S3**). Given that both *Polyctenium* subgenomes had identical chromosome structure and the pachytene chromosomes formed quadrivalents (**Figure 5**), the analyzed accession of *P. fremontii* was most likely of autotetraploid origin.

In the three cytotypes of *P. cheiranthoides* ($2x$, $3x$, and $4x$; Mandáková et al., 2020), diploid *N. holmgrenii*, diploid *S. perplexa*, and triploid *S. whitedii*, chromosome Boe4 was altered by an 8.24-Mb paracentric inversion spanning the entire block J (whole long arm). The breakpoints were located in the pericentromeric and subtelomeric regions (**Figures 2**, **6** and **Supplementary Tables S4, S5**). The absence of subgenome differentiation in the analyzed triploid and tetraploid populations of *Phoenicaulis* and *S. whitedii* suggests intra-specific, autopolyploid origins for these polyploids. Additionally, *N. holmgrenii* exhibited a 2.55-Mb whole-arm pericentric inversion on Boe5. In this population, the short arm (GB V) was inverted, rendering the acrocentric chromosome telocentric (**Figures 2**, **6** and **Supplementary Table S5**).

Chromosome Boe4 of tetraploid *C. douglasii* also displayed inversions (**Figure 2** and **Supplementary Table S6**). A 2.06-Mb pericentric inversion, spanning the whole short arm (GB I), converted the chromosome from acrocentric to telocentric. Boe4 was also modified by a 5.35-Mb paracentric inversion [breakpoint between blocks I and J, and within block J—between T28M21 (At2g40090) and T3G21 (At2g40240)]. The *Cusickiella* population analyzed for this study was most likely of autotetraploid origin given that both subgenomes were structurally similar, including a reshuffling of Boe4 (data not shown). Finally, in *B. formosa* and *B. oxylobula*, Boe5 was altered by a 9.83-Mb pericentric inversion with breakpoints within block V [between K14A3 (At5g47175) and MQL5 (At5g47150)] and K-L [between MQP15 (At3g30655) and MED5 (At3g30663)], converting the chromosome from acrocentric to metacentric (**Figures 2**, **6** and **Supplementary Table S7**).

Borodinia missouriensis had the most reshuffled genome encountered among the taxa analyzed (**Figures 2**, **6** and **Supplementary Table S8**). A whole-arm translocation between Boe1 (GBs Aa, Ca, and D) and Boe3 (F, G, W, and X) produced two *B. missouriensis*-specific translocation chromosomes. The

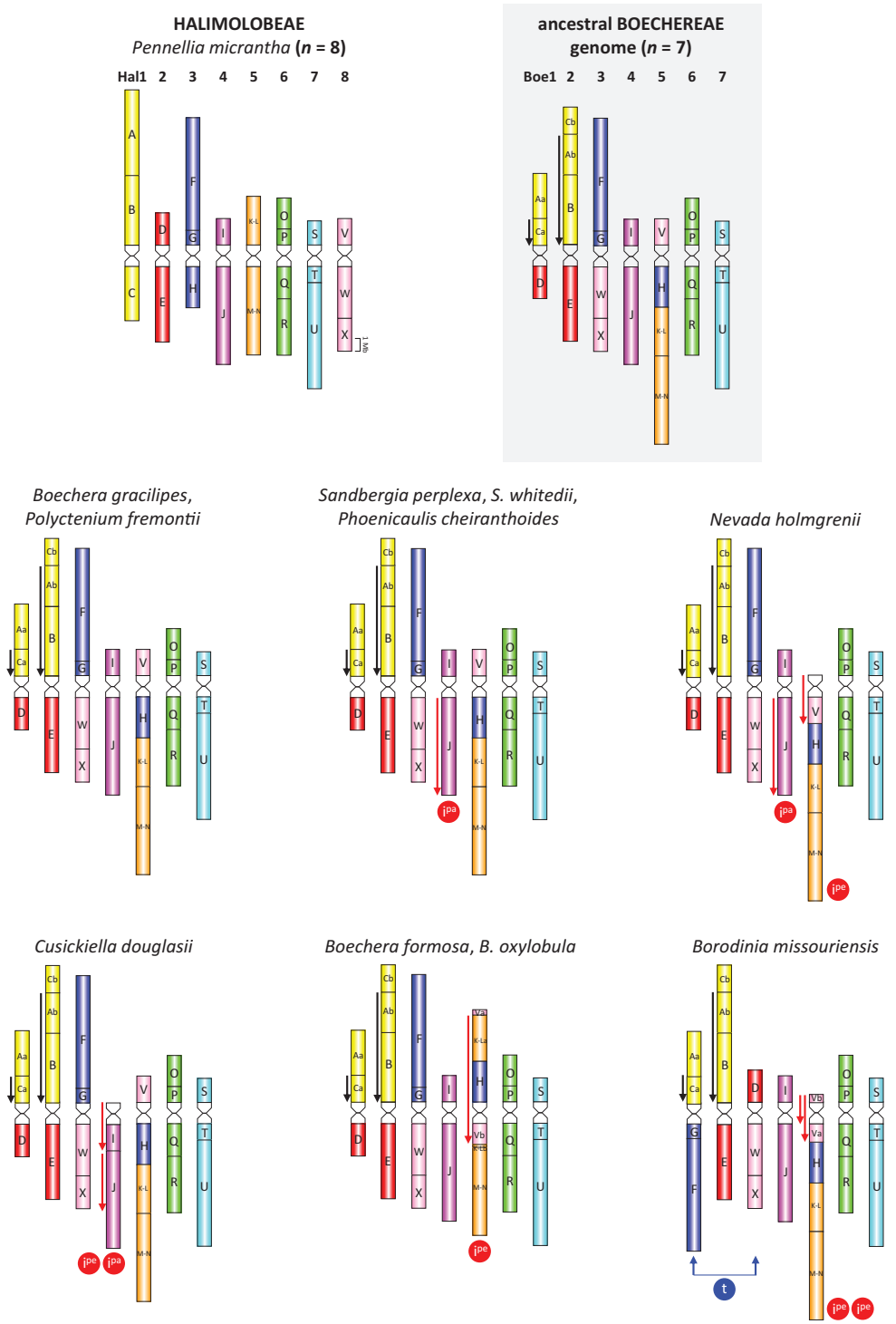

FIGURE 2 | Comparative cytogenomic maps of *Pennellia micrantha* (Halimolobeae), the ancestral Boechereae genome, and 10 analyzed Boechereae species. As subgenomes of the polyploid species/cytotypes have the same structure, only a single (sub)genome is shown for triploids (*P. cheiranthoides*, *S. whitedii*) and tetraploids (*C. douglasii*, *P. cheiranthoides*, *P. fremontii*). Color coding of 22 genomic blocks (A to X) reflects their position on the eight ancestral chromosomes (AK1–AK8) in the Ancestral Crucifer Karyotype (Lysak et al., 2016). Blocks split into two segments are labeled as "a" and "b." Downward-pointing arrows denote the inverse orientation of GBs compared to their position in the ACK represented here by the *P. micrantha* genome. Black arrows mark the Boechereae-specific inversions which occurred prior to the divergence of the tribe, whereas red arrows denote genus- and species-specific inversions that occurred after the divergence of Boechereae. I^{pe}: pericentric inversion; I^{pa}: paracentric inversion; t: whole-arm translocation. All ideograms are drawn to scale, whereby the size of GBs corresponds to the size of homeologous blocks in the *Arabidopsis thaliana* genome (The Arabidopsis Information Resource, TAIR; http://www.arabidopsis.org). Genome structure of *P. cheiranthoides* was adopted from Mandáková et al. (2020).

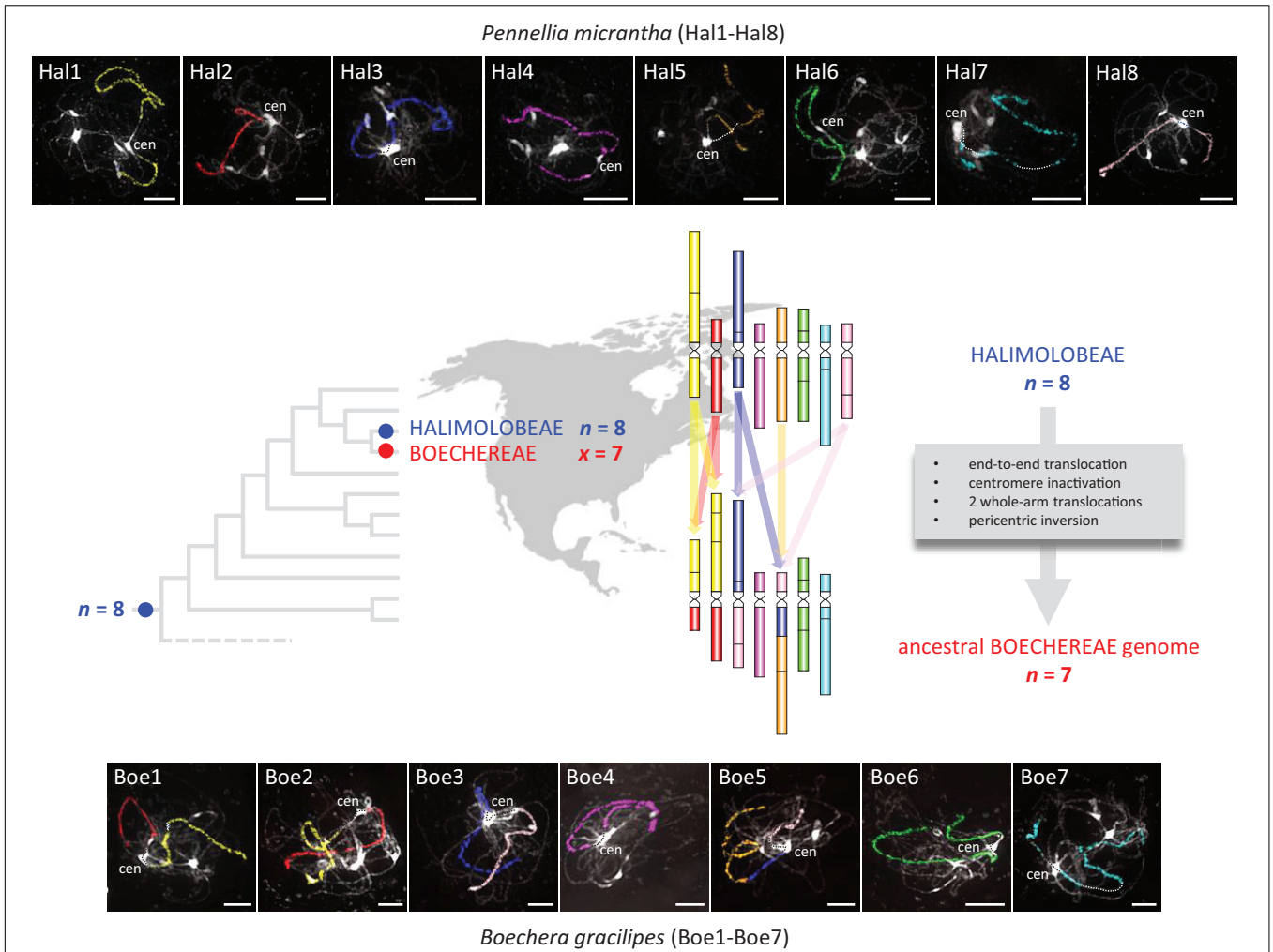

FIGURE 3 | The inferred origin of the ancestral Boechereae genome and painted chromosomes of *Pennellia micrantha* (Hal1–Hal8) and *Boechera gracilipes* (Boe1–Boe7). The ancestral Boechereae genome (n = 7) originated from a Halimolobeae-like genome (n = 8) through descending dysploidy mediated by an end-to-end translocation accompanied by a centromere inactivaton, two additional reciprocal translocations and a pericentric inversion (see **Figure 4** for more details). In the CCP images, different colors correspond to the eight ancestral chromosomes (AK1–AK8) in the Ancestral Crucifer Karyotype (Lysak et al., 2016), whereas capital letters refer to 22 genomic blocks (A to X). *A. thaliana* BAC clones defining each BAC contig/painting probe in *Pennellia* and *Boechera* are listed in **Supplementary Tables S2, S3**, respectively. Chromosomes were counterstained by DAPI. The fluorescence signals of the painting probes were captured as black and white photographs, and the signals were then pseudocolored to match the eight chromosomes of the ACK. Scale bars, 10 μm.

species also shared a 2.55-Mb whole-arm pericentric inversion spanning block V of Boe5 with *N. holmgrenii*. In *B. missouriensis*, this was followed by a small 0.57-Mb pericentric inversion splitting block V into Va and Vb and placing the Boe5 centromere close to the chromosome terminus between BACs K24F5 (At5g43211) and MNL12 (At5g43190).

Localization of rDNA Loci

In *P. micrantha*, NORs (35S rDNA loci) were localized on the termini of five chromosomes (Hal1, Hal3, Hal4, Hal6, and Hal7) and two 5S rDNA loci were adjacent to the pericentromeric heterochromatin of chromosomes Hal4 and Hal6 (**Supplementary Figure S1**). A single NOR and 5S rDNA locus were identified in eight Boechereae species analyzed; two NORs were found in *Cusickiella* and two 5S loci in *Nevada* (**Supplementary Figure S1**). NORs were located terminally on the short chromosome arm of Boe4 (*Boechera* spp. and *Cusickiella*) or Boe5 (*Borodinia*, *Cusickiella*, *Polyctenium*, and *Nevada*). In *Phoenicaulis* and both *Sandbergia* species, NORs were located interstitially, close to the pericentromere of Boe6. This supports the close relationship between *Phoenicaulis* and *Sandbergia*. Interestingly, all telocentric chromosomes were NOR-bearing (Boe4 in *Cusickiella* and Boe5 in *Borodinia* and *Nevada*). 5S rDNA loci were found positioned interstitially, close to the pericentromere of Boe5 (*Cusickiella*, *Nevada*, and *Phoenicaulis*), Boe6 (*Boechera* spp. and *Polyctenium*), and Boe7 (*Borodinia*, *Nevada*, and both *Sandbergia* spp.).

In triploid (*Phoenicaulis*, *S. whitedii*) and tetraploid (*Cusickiella*, *Phoenicaulis*, *Polyctenium*) taxa/cytotypes, the position of rDNA gene loci at the same chromosomal positions

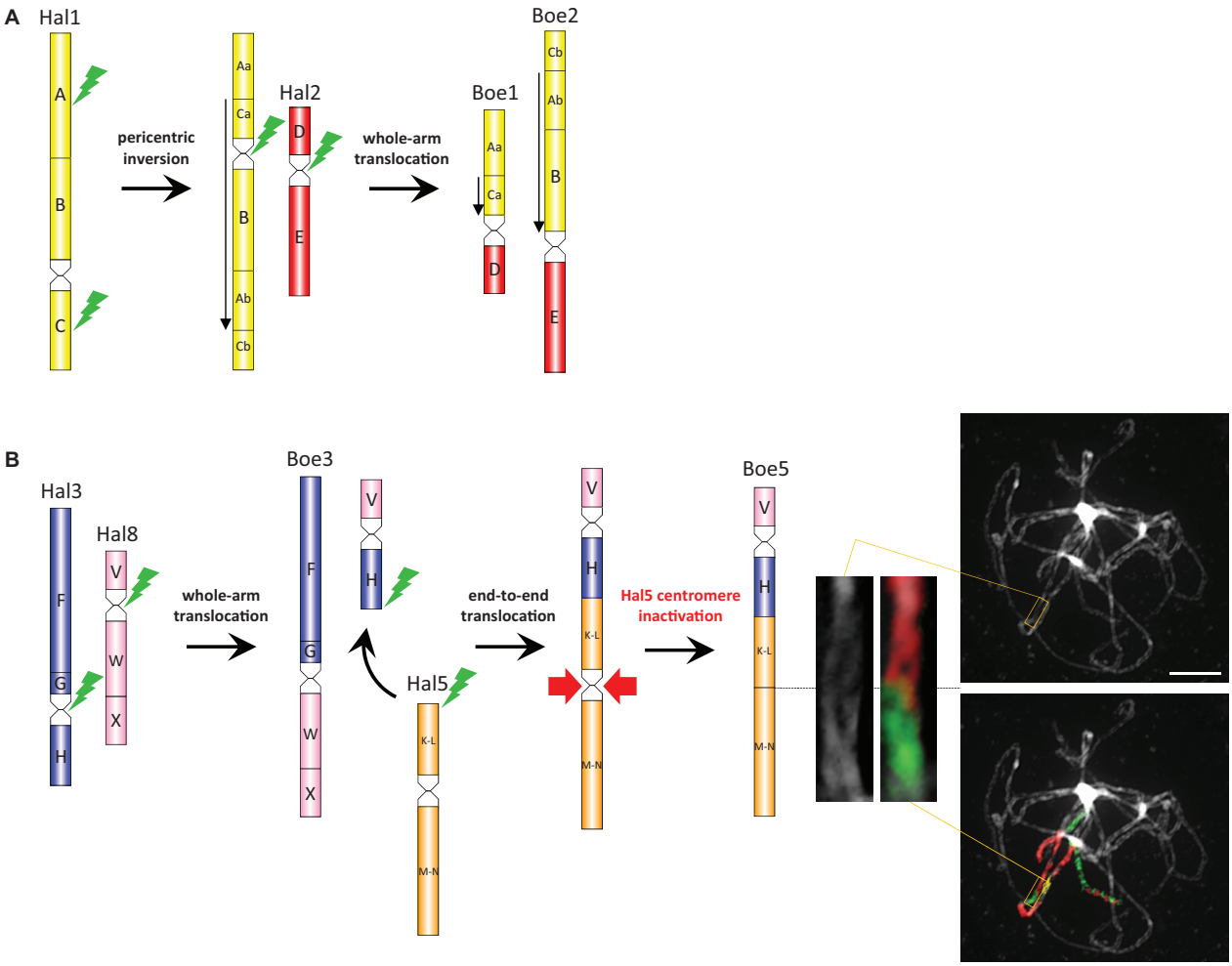

FIGURE 4 | The origin of Boechereae-specific chromosomes. Parsimoniously reconstructed origin of the ancestral Boechereae chromosomes Boe1, Boe2 **(A)**, Boe3, and Boe5 **(B)** from the ACK/Halimolobeae-like chromosomes Hal1–Hal3, Hal5, and Hal8. Chromosome images show painting of the Boe5 homeolog at pachytene in *Boechera gracilipes* including the K-L/M-N region with the inactive AK5 paleocentromere (magnified inset). Downward-pointing arrows refer to GBs inverted compared to their orientation in the ACK/Halimolobeae genome. Capital letters mark the ancestral GBs of the ACK (Lysak et al., 2016). Green flash-like symbols indicate chromosome breakpoints; red arrows indicate the lost of Hal5 (AK5) centromere on Boe5. Scale bar, 10 μm.

within three (triploid) or four (tetraploid) chromosome sets, further supported the purported autopolyploid origins of these genomes.

The Inactive Centromere Between Genomic Blocks K-L and M-N in *Boechera stricta*

The *B. stricta* genome contains an inactive centromere between GBs K-L and M-N on chromosome Boe5. To characterize this region at the sequence level, we compared scaffold 556 in the *B. stricta* assembly with orthologous genes on homeologous chromosome 3 and 5 in *A. thaliana* and *A. lyrata*, respectively (**Figure 7A**). The centromeric region is delimited by genes of the Peroxidase superfamily (loci At3g32980) and genes of the Transducin family (loci At3g33530). In *B. stricta*, the site of the eliminated (AK5) paleocentromere corresponds to a 13-kb region between orthologs Bostr.0556s0638 and Bostr.0556s0640. This region contains a single gene (Bostr.0556s0639), which is presumably paralogous to gene Bostr.13158s0074 (homology of 84.9%), located in the distant part of block M-N on the same chromosome. In the *Arabidopsis* genomes, orthologs of Bostr.13158s0074 are located within M-N on chromosomes At3 (At3g60740; At3: 22,447,245-22,453,364) and Al5 (scaffold_503309.1; Al5: 20,028,781−20,034,751), respectively. Additionally, the *A. lyrata* genome possesses a paralog, Al_scaffold_0002_1021, located on Al2 in GB E (position 9,412,104-9,422,712). No remnants of tandem repeats were detected within the 13-kb region (**Figure 7B**). Comparable distribution of transposable elements and their remnants was observed within the former centromeric region, in upstream and downstream 20 kb regions (**Figure 7C**) and along the whole Boe5 pseudo-chromosome. The absence of repeat segment enrichment within the short 13-kb region supports an almost complete

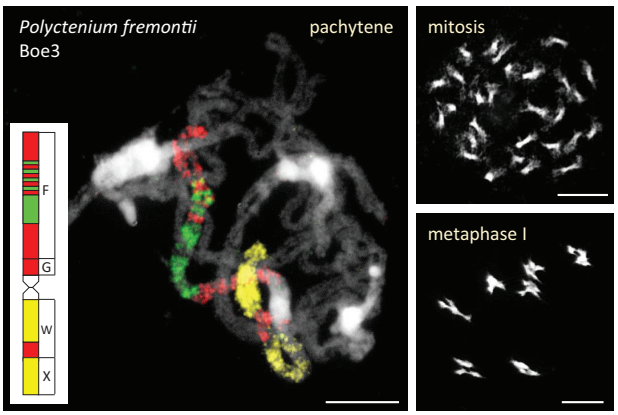

FIGURE 5 | Quadrivalents in *Polyctenium fremontii* ($2n = 4x = 28$). The 28 chromosomes form seven quadrivalents during the first meiotic division. The quadrivalent of Boe3 homeologs was identified by CCP of pachytene chromosomes with BAC contigs specific for genomic blocks F, G, W, and X (**Supplementary Table S3**). Chromosomes were counterstained by DAPI. Scale bars, 10 μm.

removal of the AK5 paleocentromere after the end-to-end "chromosome fusion" (**Figure 4B**).

The Heterochromatic *Het* Chromosome Was Absent in the Analyzed Boechereae Species

In apomictic *Boechera* species ($2n = 14$) and *Phoenicaulis* cytotypes ($2n = 21$ and 28, but not apomictic $2n = 14$), one of the Boe1 homologs (a *Het* chromosome) displayed expanded regions of pericentromeric heterochromatin (Mandáková et al., 2015a, 2020). In all analyzed species of the present study, pericentromeric heterochromatin regions of the Boe1 homologs were comparable in size.

High Frequency Apospory Occurs in *Polyctenium* and *Sandbergia*

Six accessions (**Supplementary Table S1**), representing five species distributed across four genera, were embryologically analyzed (**Figure 8**). Tetraploid *C. douglasii*, diploid *N. holmgrenii*, and diploid *S. perplexa*, were sexual. Tetraploid *P. fremontii* (MDW 2055, ES 1078) and triploid *S. whitedii* were highly aposporous. In the *P. fremontii* population ES 1078, the dyad to tetrad ratio was high (**Figure 8**), which in *Boechera* would generally indicate diplospory (Carman et al., 2019). However, three of the four observed gametophytes were forming from nucellar cells (apospory), and the fourth was forming from the surviving megaspore of a sexual tetrad. While this sample size is too small to rule out diplospory, our observations indicated that apospory initiates early during ovule development with meiosis regularly terminating as early as the sexual dyad to early tetrad stages. Such termination may have inflated the dyad to tetrad ratio (**Figure 8**).

Parietal cell frequencies were ≥ 70% for five of the six taxa studied (**Figure 8**). These frequencies are similar to those observed in *Phoenicaulis* (Mandáková et al., 2020), but they are much higher than those generally observed in *Boechera* (<50%) (Naumova et al., 2001; Carman et al., 2019). *The ES 1078 P. fremontii* sample was too small to determine this frequency. Parietal cells form from the distal daughter cell of the mitotic division of the archesporial cell. In these cases, the MMC forms from the proximal daughter cell (**Figures 9A,B**). In ovules of tenuinucellate species, the opposite normally occurs, i.e., the MMC differentiates distally, and the proximal cell is considered nucellar (Johri et al., 1992). In this respect, the Boechereae show tendencies toward crassinucellate development, with parietal cells sometimes undergoing further division to produce a parietal tissue that positions the meiocyte deeper within the ovule (**Figures 9C–G**).

Sexual and aposporous gametophyte formation were of the eight-nucleate Polygonum type (**Figures 9H–J**). In the aposporous taxa, one or more nucellar cells, and sometimes parietal cells, initiated vacuolate gametophyte formation as early as MMC differentiation (**Figures 9K,L**). As described above, this may have terminated meiosis prior to M_{II} and caused the abnormally high sexual dyad to tetrad ratio observed for the *P. fremontii* population ES 1078 (**Figure 8**). Nucellar, parietal, and nucellar epidermis cells degenerated quickly in front of the rapidly growing sexual or aposporous gametophytes (**Figures 9H–N**).

DISCUSSION

The Ancestor of the Boechereae Had Seven Chromosomes and Descended From an ACK-Like Genome With Eight Chromosomes

Using BAC-based CCP, we analyzed 14 samples of Boechereae representing seven of the nine genera recognized by Alexander et al. (2013). These analyses revealed a high level of genomic stasis across the tribe, with the earliest diverging genus *Polyctenium* (**Figure 1**; Beilstein et al., 2010; Couvreur et al., 2010) and *B. gracilipes* (a member of the derived "western *Boechera* clade," **Figure 1**) showing identical chromosome structures (**Figure 2**). Our analysis corroborrated our earlier hypothesis (Mandáková et al., 2015b, 2020) of an ancestral Boechereae genome with seven chromosomes ($n = 7$) derived from an older $n = 8$ genome by descending dysploidy. This precursor $n = 8$ genome structurally resembled the ACK, an ancestral genomic arrangement present in many tribes of crucifer lineage I (Lysak et al., 2016). Indeed, analysis of the *P. micrantha* genome from the Halimolobeae, sister group to the Boechereae, confirmed that the eight chromosomes of Halimolobeae genomes (Bailey et al., 2007) are homeologous to the eight chromosomes of the ACK. While the eight ancestral chromosomes remained conserved in Halimolobeae, the divergence of extant Boechereae appears to coincide with or follow a descending dysploidal change from $n = 8$ to $n = 7$.

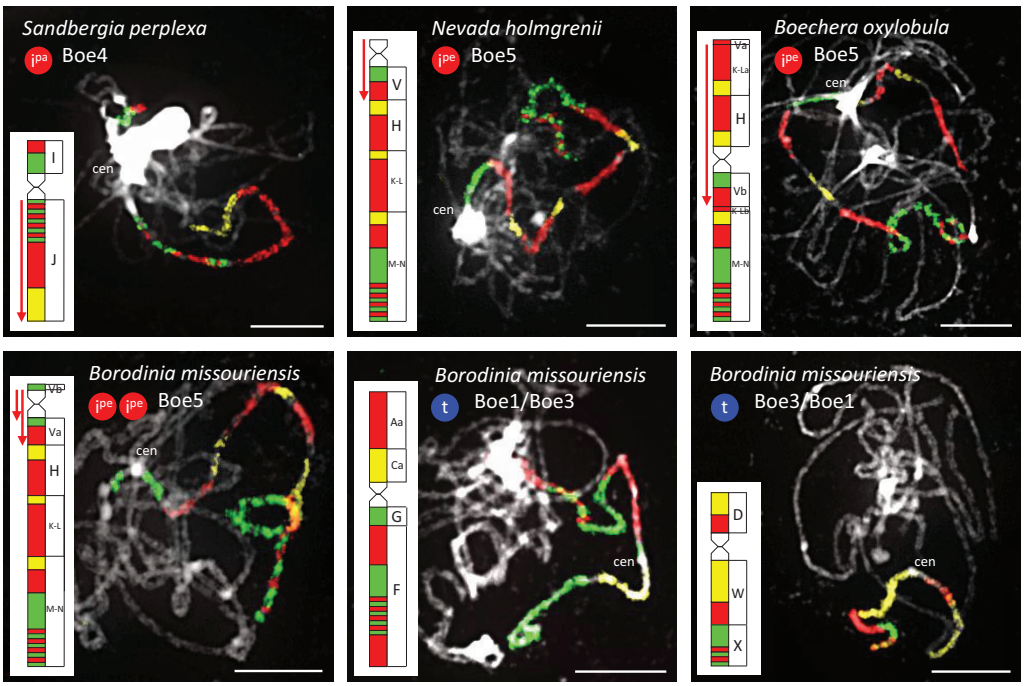

FIGURE 6 | Species- and genus-specific chromosomal rearrangements in diploid Boechereae revealed by CCP. Red downward-pointing arrows denote inverse orientation of GBs compared to their position in the ancestral Boechereae genome (**Figure 3**). I^{pe}: pericentric inversion; I^{pa}: paracentric inversion; t: whole-arm translocation. Chromosomes were identified by CCP with *A. thaliana* BAC contigs labeled by biotin-dUTP (red flourescence), digoxigenin-dUTP (green) and Cy3-dUTP (yellow). *A. thaliana* BAC clones defining each BAC contig/painting probe are listed in **Supplementary Tables S4, S5, S7, S8**. Chromosomes were counterstained by DAPI. Scale bars, 10 μm.

Tribe-Specific Descending Dysploidy

Given that all sampled Boechereae exhibit a chromosome base number of $x = 7$, it is likely that the descending dysploidy event documented above occurred prior to the initial divergence of the extant members of the tribe approximately 8 million years ago (Couvreur et al., 2010). This reduction of chromosome number did not result from a "simple fusion" of two chromosomes. The first step involved the origin of a transient translocation (AK3/AK8) chromosome by a whole-arm translocation, followed by an end-to-end translocation between AK3/AK8 and AK5 (→ chromosomes Boe3, Boe5). The chromosome-arm collinearity of the resulting Boe5 points to inactivation or loss of the AK5 centromere. Interestingly, this paleocentromere has disappeared from many other crucifer genomes independently, and it is the most frequently inactivated centromere detected so far among the tribes of lineage I (*Camelineae*: Lysak et al., 2016; *Cardamineae*: Mandáková et al., 2016; *Microlepidieae*, Mandáková et al., 2010, 2017b). As tribal divergence post-dates the origin of four tribe-specific chromosomes, we propose that the three translocations involving five out of eight ancestral chromosomes were a key evolutionary innovation underlying the origin and diversification of the Boechereae in North America.

Genus- and Species-Specific Inversions

Although reciprocal translocations clearly played a major role in the origin of the tribe, the only additional translocation noted within the tribe was a unique whole-arm transfer between Boe1 and Boe3 that produced two structurally unique chromosomes in *B. missouriensis* (**Figure 2**). However, chromosomal inversions proved to be common within the group, as they are in other mustards (e.g., Mandáková et al., 2015a; Fransz et al., 2016) and land plants in general (Hoffmann and Rieseberg, 2008). Autapomorphic pericentric inversions on chromosome Boe5 were observed in both *Borodinia* and *Nevada*, while *Cusickiella* exhibited unique pericentric and paracentric inversions on Boe4 (**Figure 2**).

Our analyses identified three potentially synapomorphic chromosomal rearrangements within the tribe: (1) a 9.83-Mb pericentric inversion on Boe5 shared by *B. formosa* and *B. oxylobula*, but not by *B. gracilipes*, (2) a 2.55-Mb pericentric inversion on Boe5 shared by *N. holmgrenii* and *Borodinia missouriensis*, and (3) a 8.24-Mb paracentric inversion on Boe4 shared by *N. holmgrenii*, all three ploidies of *P. cheiranthoides*, and both species of *Sandbergia* (**Figure 2**). Support for the synapomorphic status of these three chromosomal rearrangements is equivocal in the few phylogenetic analyses of Boechereae published to date.

With respect to the 9.83-Mb pericentric inversion apparently shared by *B. formosa* and *B. oxylobula*, the only phylogenetic analysis with appropriate taxon sampling is the concatenated nuclear gene tree of Alexander et al. (2013). This tree shows very weak support for a clade encompassing *B. oxylobula* and *B. gracilipes* but excluding *B. formosa*. A close relationship

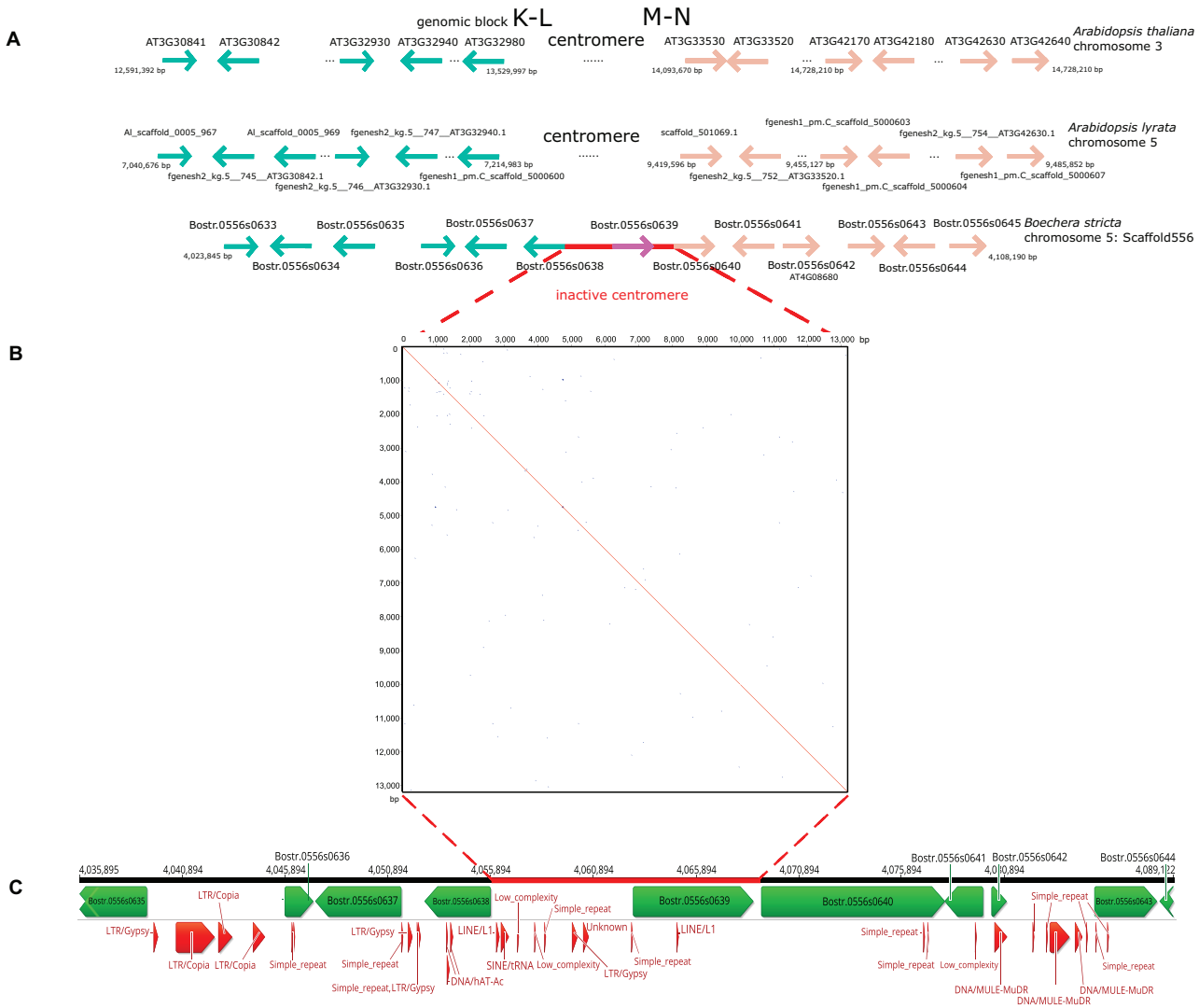

FIGURE 7 | Sequence comparison of the AK5 centromeric region on homeologous chromosomes in *Arabidopsis thaliana*, *A. lyrata* and *Boechera stricta*. **(A)** The functional centromere on chromosome 3 and 5 in *A. thaliana* and *A. lyrata*, respectively, is located between genomic blocks K-L and M-N, between At3g32980 and At3g33530. The 13-kb region corresponding to the inactive AK5 centromere is located between genes Bostr.0556s0638 and Bostr.0556s0640 on chromosome Boe5. **(B)** A dot-plot self-comparison of the 13-kb region on Boe5 showing the absence of tandem repeat arrays. **(C)** Annotation of the *B. stricta* scaffold 556 including the 13-kb region and regions 20 kb upstream and downstream. Transposable elements and their remnants do not exhibit a higher abundance within the 13-kb region.

between *B. oxylobula* (which has the inversion) and *B. gracilipes* (which does not) is congruent with morphology in this case, and it argues against the 9.83-Mb pericentric inversion being synapomorphic.

Concerning the 2.55-Mb pericentric inversion shared by *N. holmgrenii* and *B. missouriensis*, the two published analyses with appropriate taxon sampling yielded conflicting topologies. The concatenated nuclear gene tree of Alexander et al. (2013) provides very weak support for a sister relationship between *Nevada* and *Borodinia*, which would favor the interpretation of the 2.55-Mb inversion as a synapomorphic character. However, the Boechereae phylogeny presented by Beilstein et al. (2010) places *Nevada* as sister to *Phoenicaulis* not *Borodinia*, the latter being sister to a clade comprising *Boechera s.s.* and *Anelsonia* (**Figure 1B**). If this phylogeny is correct, the 2.55-Mb inversion apparently shared by *Nevada* and *Borodinia* would have originated independently.

The final potential chromosomal synapomorphy to be considered is the 8.24-Mb paracentric inversion shared by *N. holmgrenii*, all three ploidies of *P. cheiranthoides*, and both species of *Sandbergia* (**Figure 2**). Among previously published phylogenies, the Beilstein et al. (2010) topology hypothesizes a sister relationship between *Nevada* and *Phoenicaulis* (*Sandbergia* was not sampled). This would be congruent with the 8.24-Mb

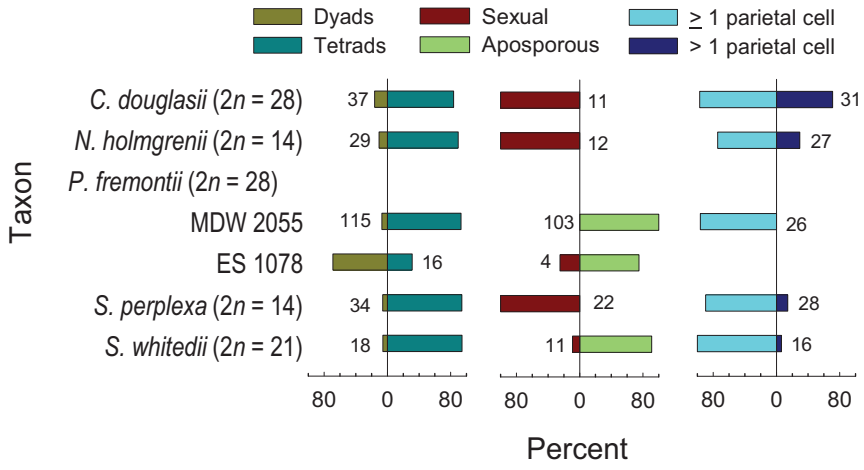

FIGURE 8 | Frequencies of dyads vs. tetrads, sexual vs. aposporous gametophytes, and parietal cells or tissues (≥1 parietal cell) vs. parietal tissues (>1 parietal cell). Numbers next to bars represent observations contributing to each variable pair.

inversion on Boe4 being interpreted as synapomorphic. However, the concatenated nuclear gene tree of Alexander et al. (2013) hypothesizes a sister relationship (very weakly supported) between *Nevada* and *Borodinia* and provides no resolution of inter-generic relationships for *Phoenicaulis* or *Sandbergia*. This creates a conflict between the only two inversions that could be a synapomorphy. If the Beilstein et al. (2010) phylogeny is correct, then the 8.24-Mb inversion discussed here could be synapomorphic but the 2.55-Mb inversion mentioned previously would not. On the other hand, if the Alexander et al. (2013) topology is correct, then the 2.55-Mb inversion could be synapomorphic but the 8.24-Mb inversion would not. A better resolved and supported phylogeny will be needed to assess the homology of the chromosomal rearrangements documented herein.

Chromosomal inversions appear to be relatively common in Boechereae, but the existence of breakage "hotspots" on several chromosomes can make it difficult to infer homology. In our dataset, most of the inversions detected are pericentric and occur in just one or two samples or taxa. The only inversion that appears to have any time depth is the 8.24-Mb paracentric inversion on Boe4 (shared by six samples representing four species and three genera). The others appear to be more recent, like the 8.4-Mb paracentric inversion on chromosome Bs1 that distinguishes the West genotype of *B. stricta* from other populations of the species (Lee et al., 2017). Some of these young inversions likely originated since the last glacial maximum, suggesting that this type of chromosomal rearrangement may be an ongoing and important contributor to reproductive isolation and speciation within the Boechereae.

The Boechereae *n* = 8 Ancestor in a Phylogeographic Context

Eurasia, specifically the Irano-Turanian floristic region, is believed to be the cradle of crucifer origin (Franzke et al., 2009). Although long-distance dispersal events contributed to plant migrations from Eurasia to North America and in the opposite direction (e.g., Wang et al., 2015; Huang et al., 2018), the Bering land bridge played the key role for the spreading of seed plants, including crucifers, from Eurasia to the North American subcontinent (e.g., Carlsen et al., 2010; Wen et al., 2010; Karl and Koch, 2013; Jiang et al., 2019). Among the clades belonging to crucifer lineage I, some are endemic to the New World (Halimolobeae and Physarieae), others have a bi-continental distribution in Asia and America (Camelineae, Cardamineae, Crucihimalayeae, Descurainieae, Erysimeae, and Smelowskieae), but the Boechereae are confined to North America with only three species occurring in the Russian Far East. The bi-continental distribution of several tribes in the lineage I, and tribes of other crucifer lineages (e.g., Arabideae: Karl and Koch, 2013), makes the Bering land bridge the most plausible colonization route for several (and perhaps all) crucifer clades to North America.

Since the Halimolobeae are consistently retrieved as the sister clade to Boechereae (e.g., Beilstein et al., 2010; Nikolov et al., 2019) and Halimolobeae genomes resemble the ACK genome with $n = 8$, it is likely that the ancestral genomes of Boechereae and Halimolobeae arose from this common ancestor. The present-day geographic ranges of the two tribes differ, with Boechereae concentrated in the United States and Canada (barely extending into Mexico) and Halimolobeae generally more southern in distribution. In fact, Halimolobeae has two distinct centers of diversity, one extending from central Mexico to the southwestern United States, and the other in the Andes, from Ecuador to central Argentina (Bailey et al., 2007). The geographic ranges of the two tribes overlap in a narrow band stretching east to west along the United States/Mexico border. As both tribes have the closest phylogenetic affinity to the primarily Eurasian clades of lineage I, we propose that the MRCA of Boechereae/Halimolobeae reached North America via the Bering land bridge. Couvreur et al. (2010) estimated that the Boechereae and Halimolobeae diverged c. 8 mya (late Miocene) and their

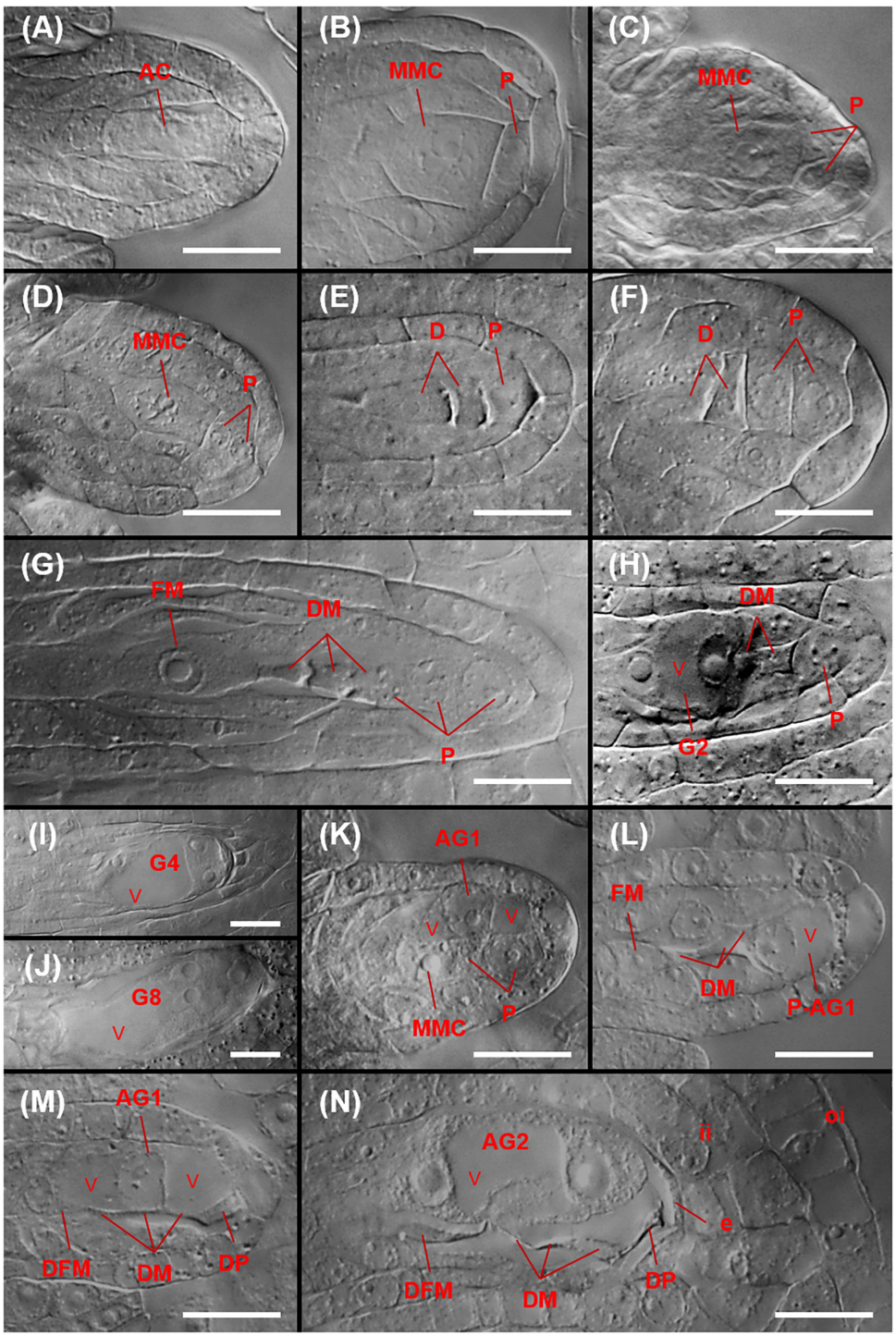

FIGURE 9 | Megasporogenesis and sexual and aposporous gametophyte formation in *P. fremontii* **(A,K–N)** and sexual *C. douglasii* **(B,D,F,G)**, *N. holmgrenii* **(C)**, and *S. perplexa* **(E,H–J)**. **(A)** Archesporial cell (AC) at the budding integument stage. **(B)** Mitotic division of an AC yielding a proximal MMC and a distal parietal cell (P). **(C,D)** Anticlinal and paraclinal divisions, respectively, of a P to yield a two-celled parietal tissue. **(E,F)** Dyads (D) with one and two Ps, respectively. **(G)** Tetrad showing the functional megaspore (FM) and three degenerating megaspores (DM). Also shown is a parietal tissue consisting of three Ps that formed from two paraclinal divisions of the original P. **(H)** Two-nucleate sexual gametophyte (G2) showing a central vacuole (v), two of three DM, and a P. **(I)** Four-nucleate sexual gametophyte (G4) with three nuclei visible. **(J)** Eight-nucleate sexual gametophyte (G8) showing egg apparatus formation at the micropylar end of the gametophyte. **(K)** 1-nucleate aposporous gametophyte (AG1) from a nucellar cell at the MMC stage. **(L)** AG1 from a parietal cell (P-AG1) at the tetrad stage. **(M)** AG1 from a nucellar cell at the late tetrad stage showing functional megaspore degeneration (DFM), DMs, degenerating nucellar cells, and a degenerating parietal cell (DP). **(N)** Two-nucleate aposporous gametophyte (AG2) at the late tetrad stage showing a DFM, DMs, degenerating nucellar cells, a degenerating parietal cell (DP), and degenerating epidermal cells. Scale bars, 20 μm.

MRCA diverged from its predominantly Eurasian sister lineage during mid to early Miocene. The Bering land bridge connected northeastern Asia and northwestern North America from the Cretaceous until the Pliocene (Gladenkov et al., 2002; Jiang et al., 2019). In the latter study, rates of dispersal from Eurasia to North America were significantly elevated throughout the Oligocene and early Miocene (c. 34 to 16 mya), particularly around 26 to 24 mya. These time estimates broadly coincide with the origin and diversification of Brassicaceae lineage I in early Miocene (∼23 to 18 mya; Hohmann et al., 2015), which likely used the Bering land bridge to disperse from northeastern Asia to North America.

Apomixis-Related Chromosomes in the Boechereae

Two heterochromatic chromosomes (*Het* and *Del*) have been previously described in *Boechera* apomicts (Kantama et al., 2007; Mandáková et al., 2015a). In eudiploid apomicts ($2n = 14$), a *Het* chromosome was identified as one of the Boe1 homologs (GBs A, C, and D). In aneuploid apomicts ($2n = 15, 22$), a centric fission partitioned *Het* chromosome to a larger *Het'* (GBs A and C) and to a smaller *Del* (block D) chromosome (Mandáková et al., 2015b). In *P. cheiranthoides*, apomictic triploids and tetraploids contained a heterochromatic *Het* (GBs A, C, and D), but it was absent in diploid apomicts (Mandáková et al., 2020). This observation indicates that aposporic reproduction in *Phoenicaulis* is not associated with the presence of a *Het* chromosome. This is further corroborated herein by the apparent absence of a *Het* chromosome in two other aposporous apomicts, tetraploid *P. fremontii* and triploid *S. whitedii*.

Autotetraploids in the Boechereae

While diploid or nearly diploid ($2n = 15$) species and hybrids seem to prevail in Boechereae (cf. BrassiBase[4], accessed on 1 February 2020), apomictic triploid ($2n = 21$) or nearly triploid ($2n = 22$) hybrids are very common in *Boechera* (Schranz et al., 2005; Li et al., 2017) and recently discovered in *Phoenicaulis* (Mandáková et al., 2020) and *S. whitedii* (this study). Tetraploids ($2n = 28$) are rarely reported and in *Boechera* all cases of tetraploidy studied to date involve interspecific hybridization (i.e., allopolyploidy; see Windham et al., 2014; Li et al., 2017). Autotetraploidy was previously documented only in *P. cheiranthoides* (Mandáková et al., 2020), where the tetraploid cytotype is more common and widespread than either the diploid or triploid. Here we report two new cases of autotetraploidy in Boechereae, involving *C. douglasii* and *P. fremontii*. Hence, *bona fide* autotetraploids occur in three out of nine Boechereae genera. The autotetraploid species/cytotypes reproduce either sexually (*Cusickiella*) or, more frequently, by apomixis (*Phoenicaulis* and *Polyctenium*). Autopolyploids often experience irregular chromosome segregation due to the multivalent formation, but apomixis can potentially bypass such problematic meioses (Darlington, 1939; Stebbins, 1971; Comai, 2005; Cosendai et al., 2011). Indeed, while the analyzed tetraploid plants of *P. fremontii* show exclusive quadrivalent pairing (**Figure 5**), they appear to be fully fertile due to apospory (**Figure 8**). Hence, apomixis appears to be stabilizing reproduction in triploids and autotetraploids, which otherwise would suffer from semi-sterility due to chromosome pairing irregularities.

[4]https://brassibase.cos.uni-heidelberg.de/

Apomixis Originated Several Times Independently During Brassicaceae and Boechereae Diversification

Following the pioneering work on apomixis in *B. holboellii* [as *Arabis holboellii* Hornem. in Böcher (1951)], Mulligan (1966) reported its occurrence in *Erysimum* L. (Erysimeae) and Mosquin and Hayley (1966) documented possible asexual seed production in *Parrya* R. Brown (Chorisporeae). Mulligan and Findlay (1970) identified several species of *Draba* L. (Arabideae) that they inferred to be apomictic, and subsequent embryological and single seed flow cytometry analyses of one of these species (*Draba oligosperma* Hook.) were suggestive of apospory (Jordon-Thaden and Koch, 2012). Detailed cytological studies of *Draba*, *Erysimum*, and *Parrya* are needed to verify the regular occurrence of asexual seed production and the specific pathway involved. However, the available anecdotal evidence for this reproductive pathway occurring in four distantly related crucifer tribes (and its induction by hybridization in a fifth tribe, Brassiceae; Ellerstrom, 1983) suggests that apomixis has evolved independently multiple times within the family.

The Boechereae appear to be a "hotspot" for the origin and diversification of apomictic taxa. Our earlier publications have embryologically confirmed apomixis (either diplospory or apospory) in more than 20 diploid and triploid taxa of *Boechera* (Windham et al., 2015; Carman et al., 2019), and apospory in diploid, triploid, and tetraploid *Phoenicaulis* (Mandáková et al., 2020), and in diploid *Borodinia laevigata* (Muhl. ex Willd.) P. J. Alexander and Windham [as *Boechera laevigata* (Muhl. ex Willd.) Al-Shehbaz in Carman et al., 2019]. Here, two new genera are added to the list of aposporous apomicts, *Polyctenium* (tetraploid) and *Sandbergia* (triploid). Two genera of Boechereae (*Anelsonia* and *Yosemitea*) remain unsampled, and the exclusive occurrence of sexual reproduction in *Nevada* and *Cusickiella* must be confirmed by more extensive sampling. In summary, apomixis is now known to occur in five of the nine genera of Boechereae, and in two of these (*Phoenicaulis* and *Polyctenium*) it is the only reproductive pathway documented to date. Based on current sampling, *Sandbergia* exhibits equal proportions of apomictic (*S. whitedii*) and sexual reproduction (*S. perplexa*). Among *Boechera* species, apomictic taxa of hybrid origin greatly outnumber their sexual diploid progenitors (Li et al., 2017), while in *Borodinia*, sexual populations appear to predominate.

High frequency apospory (70% of ovules) was observed in man-made *Raphanus* L. × *Brassica* L. hybrids (Brassiceae) (Ellerstrom and Zagorcheva, 1977; Ellerstrom, 1983), which is consistent with wide hybridization occasionally inducing apomixis in otherwise sexual species (Carman, 1997). The long-term reproductive stability conferred by apomixis to sterile or semisterile inter-specific hybrids could provide novel genotypes with sufficient time (possibly hundreds of years) to fortuitously produce, by facultative sexual reproduction, recombinants that

are more or less genomically stable (autopolyploidized, sensu Sybenga, 1996) and sexually fertile (Carman, 1997, 2007; Carman et al., 2019). If chromosome aberrations have occurred, the newly formed recombinant genomes might warrant specific or generic status. Hence, apomixis in genomically unstable taxa may facilitate (Carman, 1997; Horandl and Hojsgaard, 2012; Hojsgaard et al., 2014; Carman et al., 2019) rather than terminate (Darlington, 1939; Stebbins, 1971) speciation. That apomixis is prevalent in many large genera, e.g., among the rose, aster, and grass families (Carman, 1997; Hojsgaard et al., 2014), as well as in *Boechera*, supports this hypothesis.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

TM, JC, and ML conceived the experiments. TM-O, PH, KA, and BP conducted the study and processed the data. ML, TM, JC, and MW wrote the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work was supported by Czech Ministry of Education, Youth and Sports within the program INTER−EXCELLENCE (Project No. LTAUSA17002), by the CEITEC 2020 project (Grant No. LQ1601), the Utah Agricultural Experiment Station project awards, Utah State University, Logan to JC, and the U.S. National Science Foundation (award DEB−0816560 to MW) and the U.S. National Institutes of Health (Grant R01 GM086496 to TM-O).

ACKNOWLEDGMENTS

We thank Donovan Bailey, Eric Schranz, and Millenium Seed Bank (Royal Botanic Gardens, Kew) for providing seeds, and Dmitry German for his comments on the manuscript. Core Facility Plants Sciences of CEITEC MU is acknowledged for the cultivation of plants used in this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00514/full#supplementary-material

REFERENCES

Alexander, P. J., Windham, M. D., Beck, J. B., Al-Shehbaz, I. A., Allphin, L., and Bailey, C. D. (2013). Molecular phylogenetics and taxonomy of the genus *Boechera* and related genera (Brassicaceae: Boechereae). *Syst. Bot.* 38, 192–209. doi: 10.1600/036364413x661917

Alexander, P. J., Windham, M. D., Beck, J. B., Al-Shehbaz, I. A., Allphin, L., and Bailey, C. D. (2015). Weaving a tangled web: divergent and reticulate speciation in *Boecher fendleri* sensu lato (Brassicaceae; Boechereae). *Syst. Bot.* 40, 572–596. doi: 10.1600/036364415x688745

Al-Shehbaz, I. A. (2012). A generic and tribal synopsis of the *Brassicaceae* (Cruciferae). *Taxon* 61, 931–954. doi: 10.1002/tax.615002

Asker, S. E., and Jerling, L. (1992). *Apomixis in Plants*. Boca Raton, FL: CRC Press.

Bailey, C. D., Al-Shehbaz, I. A., and Rajanikanth, G. (2007). Generic limits in tribe Halimolobeae and description of the new genus *Exhalimolobos* (Brassicaceae). *Syst. Bot.* 32, 140–156. doi: 10.1600/036364407780360166

Beck, J. B., Alexander, P. J., Allphin, L., Al-Shehbaz, I. A., Rushworth, C., Bailey, C. D., et al. (2012). Does hybridization drive the transition to asexuality in diploid *Boechera* (Brassicaceae)? *Evolution* 66, 985–995. doi: 10.1111/j.1558-5646.2011.01507.x

Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18724–18728. doi: 10.1073/pnas.0909766107

Böcher, T. W. (1951). Cytological and embryological studies in the amphiapomictic *Arabis holboellii* complex. *Det Kongelige Danske Videnskabernes Selskab. Biol. Skrif.* 6, 1–59.

Brukhin, V., Osadtchiy, J. V., Florez-Rueda, A. M., Smetanin, D., Bakin, E., Nobre, M. S., et al. (2019). The *Boechera* genus as a resource for apomixis research. *Front. Plant Sci.* 10:392. doi: 10.3389/fpls.2019.00392

Carlsen, T., Elven, R., and Brochmann, C. (2010). The evolutionary history of Beringian *Smelowskia* (Brassicaceae) inferred from combined microsatellite and DNA sequence data. *Taxon* 59, 427–438. doi: 10.1002/tax.592008

Carman, J. G. (1997). Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. *Biol. J. Linn. Soc.* 61, 51–94. doi: 10.1111/j.1095-8312.1997.tb01778.x

Carman, J. G. (2007). "Do duplicate genes cause apomixis?," in *Apomixis: Evolution, Mechanisms and Perspectives*, eds E. Horandl, U. Grossniklaus, P. J. van Dijk, and T. F. Sharbel (Liechtenstein: A. R. G. Gantner Verlag K. G), 63–91.

Carman, J. G., Mateo de Arias, M., Gao, L., Zhao, X., Kowallis, B., Sherwood, D. A., et al. (2019). Apospory in addition to diplospory is common in *Boechera* where it may facilitate speciation by recombination-driven apomixis-to-sex reversals. *Front. Plant Sci.* 10:724. doi: 10.3389/fpls.2019.00724

Carter, K. A., Liston, A., Bassil, N. V., Alice, L. A., Bushakra, J. M., Sutherland, B. L., et al. (2019). Target capture sequencing unravels *Rubus* evolution. *Front. Plant Sci.* 10:1615. doi: 10.3389/fpls.2019.01615

Cheng, F., Wu, J., Fang, L., and Wang, X. (2012). Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. *Front. Plant Sci.* 3:198. doi: 10.3389/fpls.2019.00198

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Cosendai, A. C., Rodewald, J., and Hörandl, E. (2011). Origin and distribution of autopolyploids via apomixis in the alpine species *Ranunculus kuepferi* (Ranunculaceae). *Taxon* 60, 355–364. doi: 10.1002/tax.602006

Couvreur, T. L., Franzke, A., Al-Shehbaz, I. A., Bakker, F. T., Koch, M. A., and Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol. Biol. Evol.* 27, 55–71. doi: 10.1093/molbev/msp202

Cowie, R. H., and Holland, B. S. (2008). Molecular biogeography and diversification of the endemic terrestrial fauna of the Hawaiian Islands. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3363–3376. doi: 10.1098/rstb.2008.0061

Darlington, C. D. (1939). *The Evolution Of Genetic Systems*. Cambridge: Cambridge University Press.

Doudkin, R. V., and Volkova, S. A. (2013). A new species of *Boechera* (Brassicaceae) from the Primorsky Territory, Russia. *Novon* 22, 411–414. doi: 10.3417/2010077

Ellerstrom, S. (1983). Apomictic progeny from *Raphanobrassica*. *Hereditas* 99:315. doi: 10.1111/j.1601-5223.1983.tb00906.x

Ellerstrom, S., and Zagorcheva, L. (1977). Sterility and apomictic embryo-sac formation in *Raphanobrassica*. *Hereditas* 87, 107–119. doi: 10.1111/j.1601-5223.1977.tb01251.x

Faria, R., and Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* 25, 660–669. doi: 10.1016/j.tree.2010.07.008

Fransz, P., Linc, G., Lee, C. R., Aflitos, S. A., Lasky, J. R., Toomajian, C., et al. (2016). Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. *Plant J.* 88, 159–178. doi: 10.1111/tpj.13262

Franzke, A., German, D., Al-Shehbaz, I. A., and Mummenhoff, K. (2009). Arabidopsis family ties: molecular phylogeny and age estimates in Brassicaceae. *Taxon* 58, 425–437. doi: 10.1002/tax.582009

Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* 16, 108–116. doi: 10.1016/j.tplants.2010.11.005

Freyman, W. A., and Höhna, S. (2017). Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. *Syst. Biol.* 67, 195–215. doi: 10.1093/sysbio/syx065

German, D. A., and Friesen, N. W. (2014). *Shehbazia* (*Shehbazieae*, Cruciferae), a new monotypic genus and tribe of hybrid origin from Tibet. *Turczaninowia* 17, 17–23. doi: 10.14258/turczaninowia.17.4.3

Givnish, T. J., Spalink, D., Ames, M., and Lyon, S. P. (2016). Orchid historical biogeography, diversification, Antarctica and the paradox of orchid dispersal. *J. Biogeogr.* 43, 1905–1916. doi: 10.1111/jbi.12854

Gladenkov, A. Y., Oleinik, A. E., and Marincovich, L. (2002). A refined age for the earliest opening of Bering strait. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 183, 321–328. doi: 10.1016/s0031-0182(02)00249-3

Hand, M. L., and Koltunow, A. M. (2014). The genetic control of apomixis: asexual seed formation. *Genetics* 197, 441–450. doi: 10.1534/genetics.114.163105

Hoffmann, A. A., and Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu. Rev. Ecol. Evol. Syst.* 39, 21–42. doi: 10.1146/annurev.ecolsys.39.110707.173532

Hohmann, N., Wolf, E. M., Lysak, M. A., and Koch, M. A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* 27, 2770–2784. doi: 10.1105/tpc.15.00482

Hojsgaard, D., Klatt, S., Baier, R., Carman, J. G., and Horandl, E. (2014). Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics. *Crit. Rev. Plant Sci.* 33, 414–427. doi: 10.1080/07352689.2014.898488

Horandl, E., and Hojsgaard, D. (2012). The evolution of apomixis in angiosperms: A reappraisal. *Plant Biosyst.* 146, 681–693.

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

Huang, C. H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., et al. (2016). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol. Biol. Evol.* 33, 394–412. doi: 10.1093/molbev/msv226

Huang, J., Yang, L. Q., Yu, Y., Liu, Y. M., Xie, D. F., Li, J., et al. (2018). Molecular phylogenetics and historical biogeography of the tribe Lilieae (Liliaceae): bi-directional dispersal between biodiversity hotspots in Eurasia. *Ann. Bot.* 122, 1245–1262. doi: 10.1093/aob/mcy138

Jiang, D., Klaus, S., Zhang, Y.-P., Hillis, D. M., and Li, J.-T. (2019). Asymmetric biotic interchange across the Bering land bridge between Eurasia and North America. *Nat. Sci. Rev.* 6, 739–745. doi: 10.1093/nsr/nwz035

Johri, B. M., Ambegaokar, K. B., and Srivastava, P. S. (1992). *Comparative Embryology of Angiosperms*, Vol. 1. New York, NY: Springer-Verlag.

Jordon-Thaden, I., and Koch, M. A. (2012). "Detection of apomixis in an octoploid, alpine species, *Draba oligosperma* (Brassicaceae)," in *Proceedings of the International Plant and Animal Genome Conference XX*, San Diego, CA.

Jordon-Thaden, I. E., Al-Shehbaz, I. A., and Koch, M. A. (2013). Species richness of the globally distributed, arctic–alpine genus *Draba* L. *(Brassicaceae)*. *Alpine Bot.* 123, 97–106. doi: 10.1007/s00035-013-0120-9

Kantama, L., Sharbel, T. F., Schranz, M. E., Mitchell-Olds, T., de Vries, S., and de Jong, H. (2007). Diploid apomicts of the *Boechera holboellii* complex display large-scale chromosome substitutions and aberrant chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14026–14031. doi: 10.1073/pnas.0706647104

Karl, R., and Koch, M. A. (2013). A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. *Ann. Bot.* 112, 983–1001. doi: 10.1093/aob/mct165

Kim, C., Kim, S.-C., and Kim, J.-H. (2019). Historical biogeography of melanthiaceae: a case of Out-of-North America through the bering land Bridge. *Front. Plant Sci.* 10:396. doi: 10.3389/fpls.2019.00396

Kliver, S., Rayko, M., Komissarov, A., Bakin, E., Zhernakova, D., Prasad, K., et al. (2018). Assembly of the *Boechera retrofracta* genome and evolutionary analysis of apomixis-associated genes. *Genes* 9:185. doi: 10.3390/genes9040185

Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498. doi: 10.1093/oxfordjournals.molbev.a026248

Lee, C. R., Wang, B., Mojica, J. P., Mandáková, T., Prasad, K. V. S. K., Goicoechea, J. L., et al. (2017). Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat. Ecol. Evol.* 1:119.

Li, F. W., Rushworth, C. A., Beck, J. B., and Windham, M. D. (2017). Boechera microsatellite website: an online portal for species identification and determination of hybrid parentage. *Database* 2017:baw169. doi: 10.1093/database/baw169

Linder, H. P., and Barker, N. P. (2014). Does polyploidy facilitate long-distance dispersal? *Ann. Bot.* 113, 1175–1183. doi: 10.1093/aob/mcu047

Lysak, M. A., and Mandáková, T. (2013). "Analysis of plant meiotic chromosomes by chromosome painting," in *Methods in molecular biology*, ed. N. J. Clifton (New York, NY: Humana Press), 13–24. doi: 10.1007/978-1-62703-333-6_2

Lysak, M. A., Mandáková, T., and Schranz, M. E. (2016). Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Curr. Opin. Plant Biol.* 30, 108–115. doi: 10.1016/j.pbi.2016.02.001

Mandáková, T., and Lysak, M. A. (2008). Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (Brassicaceae). *Plant Cell* 20, 2559–2570. doi: 10.1105/tpc.108.062166

Mandáková, T., and Lysak, M. A. (2016a). Chromosome preparation for cytogenetic analyses in *Arabidopsis*. *Curr. Protoc. Plant Biol.* 1, 43–51. doi: 10.1002/cppb.20009

Mandáková, T., and Lysak, M. A. (2016b). Painting of *Arabidopsis* chromosomes with chromosome-specific BAC clones. *Curr. Protoc. Plant Biol.* 1, 359–371. doi: 10.1002/cppb.20022

Mandáková, T., Ashby, K., Price, B. J., Windham, M. D., Carman, J. G., and Lysak, M. A. (2020). Genome structure and apomixis in *Phoenicaulis* (Brassicaceae; Boechereae). *J. Syst. Evol.* doi: 10.1111/jse.12555

Mandáková, T., Gloss, A. D., Whiteman, N. K., and Lysak, M. A. (2016). How diploidization turned a tetraploid into a pseudotriploid. *Am. J. Bot.* 103, 1–10.

Mandáková, T., Heenan, P. B., and Lysak, M. A. (2010). Island species radiation and karyotypic stasis in *Pachycladon* allopolyploids. *BMC Evol. Biol.* 10:367. doi: 10.1186/1471-2148-10-367

Mandáková, T., Hloušková, P., German, D. A., and Lysak, M. A. (2017a). Monophyletic origin and evolution of the largest crucifer genomes. *Plant Physiol.* 174, 2062–2071. doi: 10.1104/pp.17.00457

Mandáková, T., Pouch, M., Harmanová, K., Zhan, S. H., Mayrose, I., and Lysak, M. A. (2017b). Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Mol. Ecol.* 26, 6445–6462. doi: 10.1111/mec.14379

Mandáková, T., Schranz, M. E., Sharbel, T. F., de Jong, H., and Lysak, M. A. (2015a). Karyotype evolution in apomictic *Boechera* and the origin of the aberrant chromosomes. *Plant J.* 82, 785–793. doi: 10.1111/tpj.12849

Mandáková, T., Singh, V., Krämer, U., and Lysak, M. A. (2015b). Genome structure of the heavy metal hyperaccumulator *Noccaea caerulescens* and its stability on metalliferous and nonmetalliferous soils. *Plant Physiol.* 169, 674–689. doi: 10.1104/pp.15.00619

Mosquin, T., and Hayley, D. E. (1966). Chromosome numbers and taxonomy of some Canadian Arctic plants. *Can. J. Bot.* 44, 1209–1218. doi: 10.1139/b66-132

Mulligan, G. A. (1966). Chromosome numbers of the family Crucifereae III. *Can. J. Bot.* 44, 309–319. doi: 10.1139/b66-037

Mulligan, G. A., and Findlay, J. N. (1970). Sexual reproduction and agamospermy in the genus *Draba*. *Can. J. Bot.* 48, 269–271.

Naumova, T. N., van der Laak, J., Osadtchiy, J., Matzk, F., Kravtchenko, A., Bergervoet, J., et al. (2001). Reproductive development in apomictic populations of *Arabis holboellii* (Brassicaceae). *Sex. Plant Reprod.* 14, 195–200. doi: 10.1007/s00497-001-0118-0

Nikolov, L. A., Shushkov, P., Nevado, B., Ga, X., Al-Shehbaz, I. A., Filatov, D., et al. (2019). Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytol.* 222, 1638–1651. doi: 10.1111/nph.15732

Ogutcen, E., and Vamosi, J. C. (2016). phylogenetic study of the tribe Antirrhineae: genome duplications and long-distance dispersals from the old world to the new world. *Am. J. Bot.* 103, 1071–1081. doi: 10.3732/ajb.1500464

Rojek, J., Kapusta, M., Kozieradzka-Kiszkurno, M., Majcher, D., Gorniak, M., Sliwinska, E., et al. (2018). Establishing the cell biology of apomictic reproduction in diploid *Boechera stricta* (Brassicaceae). *Ann. Bot.* 122, 513–539. doi: 10.1093/aob/mcy114

Rokas, A., and Holland, P. W. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15, 454–459. doi: 10.1016/s0169-5347(00)01967-4

Roy, B. A. (1995). The breeding systems of six species of *Arabis* (Brassicaceae). *Am. J. Bot.* 82, 869–877. doi: 10.1002/j.1537-2197.1995.tb15703.x

Schranz, M. E., Dobes, C., Koch, M. A., and Mitchell-Olds, T. (2005). Sexual reproduction, hybridization, apomixis, and polyploidization in the genus *Boechera* (Brassicaceae). *Am. J. Bot.* 92, 1797–1810. doi: 10.3732/ajb.92.11.1797

Stebbins, G. L. (1971). *Chromosomal Evolution in Higher Plants*. London: E. Arnold.

Sybenga, J. (1996). Chromosome pairing affinity and quadrivalent formation in polyploids: do segmental allopolyploids exist? *Genome* 39, 1176–1184. doi: 10.1139/g96-148

Wang, X.-J., Shi, D.-C., Wang, X.-Y., Wang, J., Sun, Y.-S., and Liu, J.-Q. (2015). Evolutionary migration of the disjunct salt cress eutrema salsugineum (= *Thellungiella salsuginea*, Brassicaceae) between Asia and North America. *PLoS One* 10:e0124010. doi: 10.1371/journal.pone.0124010

Warwick, S. I., and Al-Shehbaz, I. A. (2006). Brassicaceae: chromosome number index and database on CD-Rom. *Pl. Syst. Evol.* 259, 237–248. doi: 10.1007/s00606-006-0421-1

Wen, J., Ickert-Bond, S., Nie, Z. L., and Li, R. (2010). "Timing and modes of evolution of eastern Asian - North American biogeographic disjunctions in seed plants," in *Darwin's Heritage Today: Proceedings of the Darwin 2010 Beijing International Conference*, eds M. Long, H. Gu, and Z. Zhou (Beijing: Higher Education Press), 252–269.

Windham, M. D., Beck, J., Alexander, P., Li, F.-W., Rushworth, C., Bailey, C. D., et al. (2014). *Newly Documented Hybrids In The Tribe Boechereae (Brassicaceae) Challenge Current Generic Circumscriptions In The Group*. Available online at: http://www.2014.botanyconference.org/engine/search/index.php?func=detail&aid=312

Windham, M. D., Beck, J. B., Li, F.-W., Allphin, L., Carman, J. G., Sherwood, D. A., et al. (2015). Searching for diamonds in the apomictic rough: a case study involving *Boechera lignifera* (Brassicaceae). *Syst. Bot.* 40, 1031–1044. doi: 10.1600/036364415x690076

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: 03 June 2020
doi: 10.3389/fpls.2020.00719

Genomic Blocks in *Aethionema arabicum* Support Arabideae as Next Diverging Clade in Brassicaceae

*Nora Walden[1], Thu-Phuong Nguyen[1], Terezie Mandáková[2], Martin A. Lysak[2] and Michael Eric Schranz[1]**

[1] Biosystematics Group, Wageningen University, Wageningen, Netherlands, [2] Central European Institute of Technology, Faculty of Science, Masaryk University, Brno, Czechia

The tribe Aethionemeae is sister to all other crucifers, making it a crucial group for unraveling genome evolution and phylogenetic relationships within the crown group Brassicaceae. In this study, we extend the analysis of Brassicaceae genomic blocks (GBs) to *Aethionema* whereby we identified unique block boundaries shared only with the tribe Arabideae. This was achieved using bioinformatic methods to analyze synteny between the recently updated genome sequence of *Aethionema arabicum* and other high-quality Brassicaceae genome sequences. We show that compared to the largely conserved genomic structure of most non-polyploid Brassicaceae lineages, GBs are highly rearranged in *Aethionema*. Furthermore, we detected similarities between the genomes of *Aethionema* and *Arabis alpina*, in which also a high number of genomic rearrangements compared to those of other Brassicaceae was found. These similarities suggest that tribe Arabideae, a clade showing conflicting phylogenetic position between studies, may have diverged before diversification of the other major lineages, and highlight the potential of synteny information for phylogenetic inference.

Keywords: *Aethionema*, Brassicaceae, comparative genomics, genomic blocks, synteny, Arabideae

OPEN ACCESS

Edited by:
Steven Dodsworth, University of Bedfordshire, United Kingdom

Reviewed by:
Patrick P. Edger, Michigan State University, United States
Zhen Li, Ghent University, Belgium

***Correspondence:**
Michael Eric Schranz
eric.schranz@wur.nl

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

Received: *03 February 2020*
Accepted: *06 May 2020*
Published: *03 June 2020*

Citation:
Walden N, Nguyen T-P, Mandáková T, Lysak MA and Schranz ME (2020) Genomic Blocks in Aethionema arabicum Support Arabideae as Next Diverging Clade in Brassicaceae. Front. Plant Sci. 11:719. doi: 10.3389/fpls.2020.00719

INTRODUCTION

The Brassicaceae is an economically important plant family, containing the *Brassica* crops, rapeseed and several ornamental taxa (e.g., *Aubrieta*, *Iberis*). Due to the availability of abundant genomic resources, such as the high-quality reference genome for model plant *Arabidopsis thaliana*, the family has become a model system for studying plant trait, genome and chromosomal evolution. The Brassicaceae family diverged from its sister-family, the Cleomaceae, ∼43 mya (million years ago) (Schranz and Mitchell-Olds, 2006; Edger et al., 2018). Divergence of tribe Aethionemeae, sister lineage to all other Brassicaceae, with its single genus *Aethionema* occurred ∼32 mya (Hohmann et al., 2015). The subsequent diversification of the rest of the family, or "crown-group," started ∼23 mya (Hohmann et al., 2015). The crown-group includes ∼3,900 species in 350 genera, grouped into 51 monophyletic tribes[1] (BrassiBase; Koch et al., 2018). These tribes are further grouped into either three or five major lineages, termed I–III or A–E (Koch and Al-Shehbaz, 2009; Franzke et al., 2011; Huang et al., 2016; Nikolov et al., 2019).

[1]https://brassibase.cos.uni-heidelberg.de/

Despite the wealth of sequence information used for recent phylogenetic reconstructions, the deeper nodes of the crown group Brassicaceae, including between lineages, are still not fully resolved. All data show Aethionemeae as the first diverging lineage. However, differing branching orders of the crown-group lineages have been reported. This is largely due to conflicting signals between plastome and nuclear data. Recent phylogenies based on extensive nuclear genome data support lineage III/E, including for example *Euclidium syriacum*, as sister to lineages I/A, including model species *A. thaliana*, and II/B, including the *Brassica* crops (Huang et al., 2016; Kiefer et al., 2019; Nikolov et al., 2019; **Figure 1A**). Plastome sequence based phylogenies on the other hand consistently place lineage II/B and III/E as sister to lineage I/A (Guo et al., 2017; Mabry et al., 2019; Nikolov et al., 2019; **Figure 1B**). Additionally, tribe Arabideae, including the important model plant *Arabis alpina*, is placed either as sister to lineages I/A and II/B (Kiefer et al., 2019; Nikolov et al., 2019; **Figure 1A**) or within lineage II/B (e.g., Hohmann et al., 2015; Guo et al., 2017; **Figure 1B**). Given the importance of Brassicaceae as a model system, a resolved and reliable backbone phylogeny is a crucial prerequisite for understanding genome and trait evolution on a family-wide scale.

To facilitate comparative genomics and studies of genome evolution, a reference system of genomic blocks (GBs) was established for Brassicaceae genomes (Schranz et al., 2006). Originally inferred from genetic maps and the first available whole genome sequences (*A. thaliana*, *Arabidopsis lyrata*, *Capsella rubella* and *Brassica rapa*), this resulted in the description of the ACK (ancestral crucifer karyotype) with eight chromosomes termed AK1–AK8 and 24 GBs (A–X). Despite its name, the ACK was not the ancestral genome of the Brassicaceae, but rather it can be seen as the hypothetical genome of the MRCA (most recent common ancestor) of lineages I/A and II/B. Since its description, the release of additional whole genome sequences as well as comparative cytogenetic analyses have led to the family-wide expansion of the genomic-block concept and reduction to 22 conserved GBs (Lysak et al., 2016). The PCK (Proto-Calepineae Karyotype, $n = 7$) was described as the ancestral karyotype for clades of lineage II/B (Mandáková and Lysak, 2008), and the CEK (Clade E Karyotype, $n = 7$) as the ancestral karyotype of lineage III/E (Mandáková et al., 2017a). The CEK genome bears some resemblance to the organization of GBs in the *A. alpina* genome (Willing et al., 2015). A high number of within-GB breakpoints compared to the ACK has been observed in both the CEK and *A. alpina* genomes, thus raising the question whether these clades may indeed be more closely related than supported by plastome data. Analysis of GBs can be conducted either with cytogenetic methods, or using bioinformatics. Cytogenetic methods usually rely on Bacterial Artifical Chromosome (BAC) clones, often from *A. thaliana* when studying Brassicaceae species, for chromosome painting. This method has the advantage that even species with little genomic information available can be studied; however, it is limited by the need for BAC clones from more or less closely related taxa. In contrast, bioinformatic methods can be applied even to distantly related species, but high-quality genomic information is required to identify genomic collinearity (syntenic blocks). The evolutionary distance between *Aethionema* and *Arabidopsis* limits the success of chromosome painting, and a high-quality reference genome was so far unavailable. Thus, no species from the more distantly related tribe Aethionemeae has been analyzed in the context of the GBs using either method.

The genus *Aethionema* comprises 57 species (BrassiBase, see footnote 1; Koch et al., 2018). It most likely originated in the Anatolian Diagonal, and dispersed throughout the Irano-Turanian region and large parts of the Mediterranean (Mohammadin et al., 2017). Over the past few years, the genus has been used to study fruit dimorphism (Lenser et al., 2016, 2018; Wilhelmsson et al., 2019) and seed germination (Mérai et al., 2019), in particular using the species *Aethionema arabicum* (e.g., the ERA-CAPS SeedAdapt project). Its divergence occurred sometime after the Brassicaceae-specific At-α WGD (whole genome duplication) (**Figure 1**). Ancient WGDs are thought to be associated with diversification (Tank et al., 2015), and duplicated genes originating from these events may play an important role for evolving new traits (Hofberger et al., 2013). Remnants of repeated ancient WGDs are found in all land plants (One Thousand Plant Transcriptomes Initiative, 2019), with some of them shared between orders and families, and some family-specific. In addition to At-α, the At-β event specific to the core Brassicales (Edger et al., 2015) is of particular importance to the Brassicaceae. The evolution of glucosinolates, secondary compounds involved in herbivore defense, and the family's coevolution with Pieridae butterflies is most likely associated with gene family expansion due to WGD following At-β and At-α (Hofberger et al., 2013; Edger et al., 2015). Following polyploidization, genomes often undergo genome size reduction eventually leading to diploidization, a process also referred to as genome fractionation. Genome downsizing is generally accompanied by chromosomal rearrangements and gene loss. The phylogenetic position of *Aethionema* as sister to all other Brassicaceae makes this lineage a crucial link that is needed to understand genome evolution after WGD in Brassicaceae.

The observation that diversifications after WGDs often occur after a considerable time lag and exclude a species-poor sister lineage that shares the WGD has been formalized in the "WGD radiation lag-time model" (Schranz et al., 2012). Diversification of the species-rich and successfully diversifying Brassicaceae crown group contrasted by species-poor tribe Aethionemeae follows this pattern (Schranz et al., 2012). Fractionation and unequal gene loss may be responsible, but this hypothesis still needs to be tested (Schranz et al., 2012). An updated and substantially improved reference genome for *Ae. arabicum* was recently released (*Ae. arabicum* genome v3.0; Nguyen et al., 2019). Its chromosome-level assembly of the eleven linkage groups corresponding to the chromosomes ($n = 11$) allows us to apply the concept of GBs to *Aethionema*, and analyze the genomic structure of the sister clade to all other Brassicaceae in more detail.

Phylogenetics have so far failed to resolve the deeper nodes within Brassicaceae, even using ever larger transcriptome data sets. Instead of relying on nucleotide sequences, using genomic features such as synteny and/or chromosomal rearrangements could therefore prove to be a useful tool to resolve such

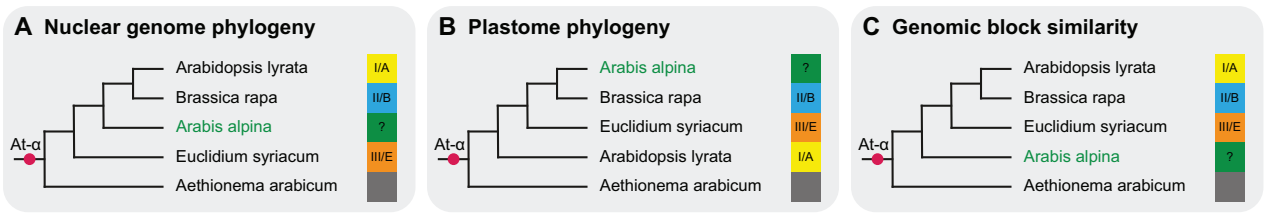

FIGURE 1 | Three potential evolutionary scenarios for the backbone phylogeny of Brassicaceae. **(A)** Nuclear data places lineage III/E as sister to the rest of the family, with Arabideae branching off next and outside of lineage II/B (Kiefer et al., 2019; Nikolov et al., 2019). **(B)** Phylogenetic reconstruction based on plastome data consistently places lineage I/A as sister to lineages II/B and III/E, and Arabideae in lineage II/B or extended lineage II (e.g., Hohmann et al., 2015; Guo et al., 2017). **(C)** Genomic blocks of *Arabis alpina* and *Aethionema arabicum* show some similarities indicating that Arabideae may be the first branching clade in crown group Brassicaceae (this study). The At-α WGD occurred sometime before divergence of Brassicaceae. Note that these three scenarios are not a comprehensive summary of all Brassicaceae phylogenies – other tree topologies have been published (e.g., Beilstein et al., 2010), however, here we only show topologies from recent genomic studies based on a high number of genes. Species names are representatives of their evolutionary lineages and were not necessarily included in the cited studies.

problematic nodes and disentangle phylogenetic placement of Brassicaceae lineages. Here, we present the syntenic blocks in the genome of *Ae. arabicum* and explore open questions concerning genome evolution and phylogenetics in Brassicaceae: Given the early divergence of *Aethionema* and its position as the species-poor sister group, is its genomic structure similar to the ACK and the largely conserved genomic structure of crown group Brassicaceae? Are the same breakpoints observed between *Aethionema*, *Arabis* and CEK genomes, and can synteny be used to obtain evidence for the phylogenetic position of early diverging lineages? We show that compared to the ACK, the syntenic blocks in the genome of *Aethionema* are broken into a high number of sub-blocks across its linkage groups. Among the high number of breakpoints, we observed, some are shared with *A. alpina* and *E. syriacum* representing the ancestral CEK genome. Our results suggest that Arabideae may have diverged before diversification of lineages I–III.

RESULTS

Genomic Blocks in the *Aethionema arabicum* Reference Genome

Our analysis revealed 13,719 syntenic genes between *Aethionema* and *A. thaliana* that are in syntenic blocks. These blocks are defined as regions sharing at least 20 collinear genes (our chosen threshold for the detection of GBs) when disregarding syntenic regions originating from the At-α, At-β, and segmental duplication events. The duplicated regions could easily be identified using K_s values; orthologous blocks generally had median K_s values of around 0.77 (purple colored in **Figure 2A**), while mean K_s values for blocks derived from At-α and older duplication events (WGD-derived paralogs) were higher, around 1.37 (blue and turquoise colored in **Figure 2A**). The average length of the syntenic blocks was 199 ± 183 (mean ± SD) syntenic genes, ranging from 23 to 833 genes. Using the same analysis and parameters on the *Arabis* genome resulted in the detection of 16,588 syntenic genes with an average block length of 313 ± 373 genes, ranging from 25 to 1,507 genes in a block. Here, K_s was lower for orthologous blocks (0.41) and At-α derived blocks (1.01). This difference is most likely the result of the additive effect of lineage-specific substitution, leading to a higher number of substitutions in the pairwise comparison of the more divergent species. The difference in syntenic block length is also reflected by the number of syntenic blocks: 69 were detected in the eleven linkage groups of the *Aethionema* genome (**Figure 2B**), compared to 53 in *Arabis*. All 22 GBs from the ACK (following the updated definition by Lysak et al. (2016) were present in *Aethionema*. However, genomic block G was not detected when setting the minimum number of genes in a syntenic block to 20. When not restricting block size, G was detected (with six syntenic genes) on LG9 between F4 and H, the same position this block has in the ACK.

Placement of At-α Relative to the Evolution of Brassicaceae

Analysis of the syntenic regions in the genome of *Ae. arabicum* revealed the duplicated regions originating from gene and ancient whole genome duplications. In reciprocal analyses (self–self comparisons), K_s values of around 0.8 are generally indicative of duplicates retained from At-α (Schranz et al., 2012). Higher K_s values are characteristic for duplicated genes originating from the At-β and older WGDs (Tang and Lyons, 2012). In pairwise analyses, like we conducted here, these characteristic K_s values are increased relative to the divergence of the selected species, as each lineage accumulates their own mutations after divergence, thus median K_s for orthologs was between 0.58 and 0.94 (**Figures 2A,C**).

Notably, in K_s histograms of *Aethionema* vs. other Brassicaceae the peak from orthologs is almost indistinguishable from that resulting from At-α duplicates. However, the origin of syntenic regions can clearly be distinguished in the syntenic dotplots (**Figure 2A** and **Supplementary Figures 1–3**). Median K_s in all histograms of pairwise comparisons involving *Aethionema* is around 0.8 (**Supplementary Figure 4**). This is not the case when comparing other Brassicaceae species, where K_s values are between 0.44 and 0.52. The similar K_s values between orthologs and paralogs (At-α derived) in *Aethionema* are consistent with only a relatively short time passing between At-α and divergence of *Aethionema* from the rest of Brassicaceae, but

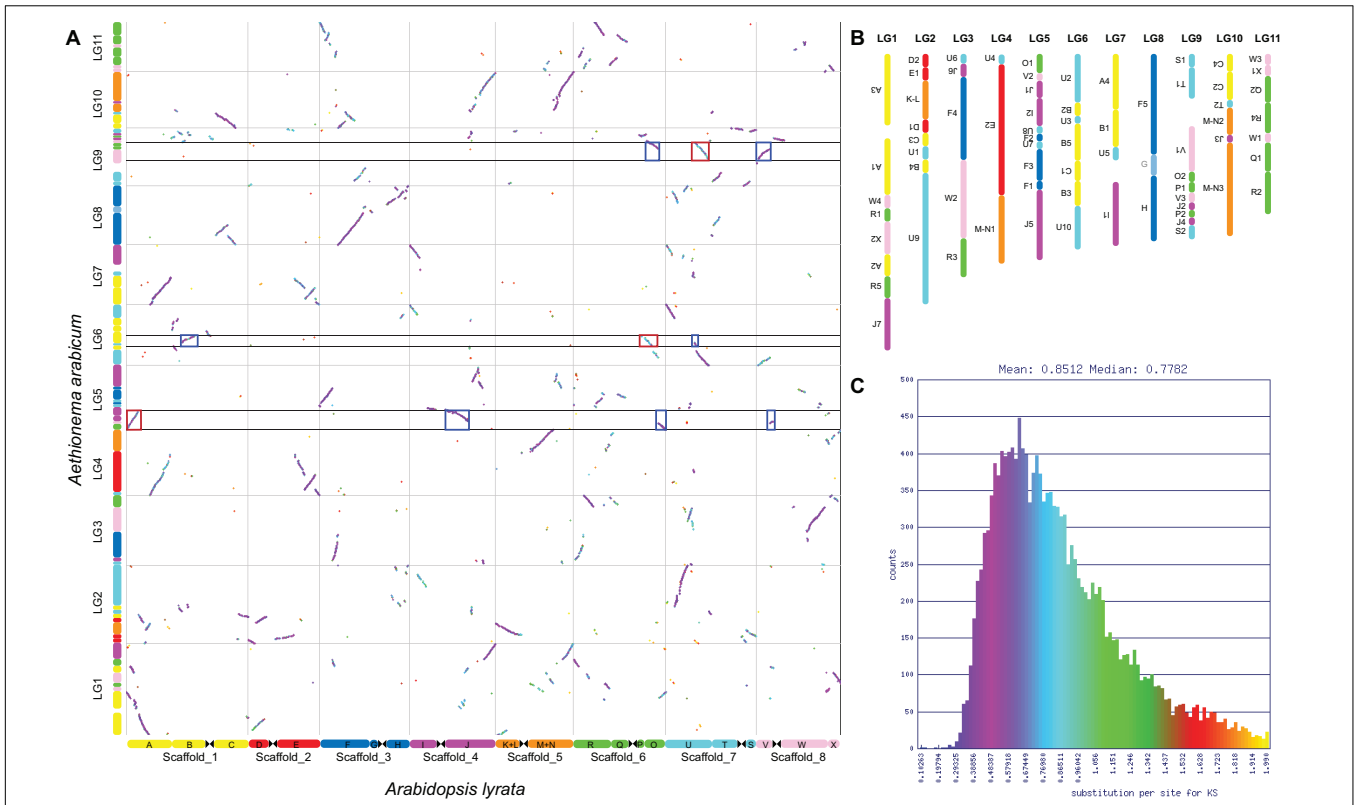

FIGURE 2 | *Aethionema arabicum* genomic blocks. **(A)** Syntenic dotplot of *Ae. arabicum* and *Arabidopsis lyrata*, which closely resembles the ACK. The dotplot was generated using SynMap implemented in CoGe. Syntenic genes are colored by K_S values to help differentiate between orthologs and At-α, At-β or segmental duplication derived paralogs. Assignment to genomic blocks is given on the left for *Aethionema* and below for *Arabidopsis*. Red boxes highlight At-α derived blocks syntenic to contiguous blocks detected in the *Aethionema*/*Arabis* comparison, blue boxes their orthologous counterparts. **(B)** Genomic blocks on the eleven linkage groups of *Ae. arabicum*. Up to eight sub-blocks were detected and blocks are ordered relative to the ACK. Inversions are indicated by upside-down block names. For reasons of consistency, genomic block G is shown despite its size below our chosen 20 genes threshold. **(C)** Histogram of synonymous substitution rate K_S. Color scheme corresponds to that used in panel **(A)**.

diversification of the crown group having occurred with some delay. Nevertheless, the small differences between orthologs and paralogs in *Aethionema* are sufficient to differentiate between the two in the syntenic dotplots (**Figure 2A**), as well as using median K_s of syntenic blocks. In addition, gene content between orthologs is more similar compared to paralogs, due to unequal fractionation, with paralogs generally containing more syntenic genes.

Conserved Blocks and Boundaries Within Brassicaceae

Most GBs in crown group diploid Brassicaceae species are conserved, i.e., they are not broken into sub-blocks. Interestingly, when within-block breaks and rearrangements are observed, this most often involved AK6 and AK8 [PCK; (Mandáková and Lysak, 2008)] and additionally AK4 [Arabideae; (Mandáková et al., 2020)], with a maximum number of three sub-blocks. This is, however, not the case in *Aethionema*. GBs from all eight chromosomes of the ACK show multiple within-block breakpoints, with the exception of three GBs (G, H, and K–L). The three GBs that did not break into sub-blocks in *Aethionema* are also conserved as a unit throughout most Brassicaceae lineages. Both G and H are present as a single block each on AK3 of the ACK, and they are also conserved in the PCK, CEK, *Arabis* and in the more rearranged *A. thaliana* genome. K-L is conserved as a unit in most lineages of Brassicaceae, but split into K and L in *A. thaliana*, and bioinformatic analysis also detected a small segment (35 genes) of K-L on *A. alpina* chromosome 3, while the largest part of this GB was detected on chromosome 5.

Apart from conserved blocks, also conserved shared GB associations can be observed across the family. While in crown group species only few new GB associations were created through rearrangements (e.g., translocations, inversions), this is the case for almost all blocks in *Aethionema*. Only four GB associations are shared between ACK, PCK, CEK, Arabideae, and *Aethionema*. The A-B association on AK1 can be found on LG-7 (A4-B1), the F-G association on AK3 on LG-8 (F4-G), the G–H association on AK3 is located on LG-8 as well, and the I–J association on AK4 is located on LG-5 (I2–J1). Notably, the I–J association is not found in *Arabis*, but a recent analysis of Arabideae revealed that this association is conserved in *Pseudoturritis turrita*, the sister to all other Arabideae, while it is not retained in later diverging

Arabideae genera (Mandáková et al., 2020) and therefore the lack of this association can be considered the derived state in *Arabis*.

Shared Sub-Block Associations Between Aethionema and Arabis

To explore whether genomic features could help resolve the deeper nodes of the Brassicaceae phylogeny, we searched for shared breakpoints and boundaries between blocks and sub-blocks in the genome sequences of *Aethionema* and *Arabis* that are not present in the ACK. As no chromosome-level assembly from any species of lineage III/E are yet available, we could not extend our analysis to this lineage. Instead, we searched the *E. syriacum* genome for the syntenic regions of interest identified in the *Aethionema–Arabis* analysis. Three regions of interest were identified that represent shared block and sub-block boundaries or similar breakpoints.

We identified three shared unique boundaries between *Aethionema* and *Arabis*: J1-V2-O; U3-B5; and V1-O2-P. First, the association J1-V2-O1 from LG-5 (**Figure 3A**) corresponds to the Ja-V-O association on chromosome Ar6 across Arabideae (Willing et al., 2015; Mandáková et al., 2020). In *Euclidium*, the genome segment containing the J-V-O segment is assembled in one fragment, but contains only partial and inverted V and O orthologs. Comparative chromosome painting did not detect V on chromosome Es3 in *E. syriacum* (Mandáková et al., 2017a). Interestingly, on chromosome Ct6 in *Chorispora tenella* (also lineage III/E), V and a small fragment of neighboring block W have inserted between Ja and Oa (Mandáková et al., 2017a). However, the genome sequence of the species would be required to determine if the breakpoints are identical with those identified in *Aethionema* and Arabideae genomes. Second, the boundary U3-B5 on LG-6 (**Figure 3B**) is found on chromosome 7 of the *Arabis* genome. The B segment was not detected using chromosome painting on this chromosome in *Arabis* (Willing et al., 2015; Mandáková et al., 2020), but could be detected using SynMap (**Supplementary Figure 1A**). In *Euclidium*, this fragment of the genome was not assembled contiguously, therefore no conclusions on lineage III/E can be drawn until a chromosome-level assembly becomes available. Finally, the association V1-O2-P1 on LG-9 (**Figure 3C**) is also detected on chromosome 6 in the *Arabis* genome using bioinformatic methods. In contrast to the other three regions, this association is also shared with *Euclidium* according to our analysis of its genome sequence. However, cytogenetic analyses did not show evidence of block V in vicinity of O in the *Euclidium* genome (Mandáková et al., 2017a), potentially due to the small size of the respective blocks (in *Arabis*, the V sub-block was only 73 syntenic genes long). Detailed figures of the three genomic regions in all pairwise comparisons are shown in **Supplementary Figures 5–7**.

In support of the potential "ancestral state" of the three aforementioned shared breakpoints between *Aethionema* and *Arabis*, the older At-α derived paralogous blocks are highly syntenic to all three regions (highlighted in red boxes in **Figure 2A** and **Supplementary Figures 1–3**; the respective orthologs are highlighted in blue). The J1-V2-O1 block detected in *Aethionema* and *Arabis* is syntenic to a part of A on the AK1, the U3-B5 block is syntenic to a part of O on AK6, and the V1-O2-P1 block from *Aethionema*, *Arabis* and *Euclidium* is syntenic to a part of U on AK7. This similarity of At-α blocks and continuous blocks in *Aethionema* and *Arabis* can be seen as strong evidence for the ancestral status of these genomic regions in the two species, with subsequent rearrangements leading to the blocks building up the ACK.

In genomes of lineage III/E and *Arabis*, blocks from AK4, AK6, AK7, and AK8 are subject to extensive rearrangements and within-block breaks. The association of GBs P and V in particular is observed in *Arabis* on chromosome Ar6 and conserved across Arabideae (Mandáková et al., 2020). Interestingly, the association V3-P2 on LG-9 is also detected in *Aethionema*. However, different block borders are associated, suggesting multiple independent origins. In *Euclidium*, this genomic region was again not assembled contiguously, and no conclusion on lineage III/E can be drawn from genome sequence data. However, cytogenetic analyses have not shown evidence for an association of blocks V and P in any species from this lineage (Mandáková et al., 2017a).

DISCUSSION

Here, we analyzed GBs in the genome of *Ae arabicum*; comparison with *A. alpina* and *E. syriacum* provide evidence for a new placement of *Arabis* within the Brassicaceae. The phylogenetic position of *Aethionema* as sister of all other Brassicaceae lineages makes this genus particularly interesting in the context of crucifer genome evolution. Due to the earlier availability of genomes and genetic maps of species from lineages I/A and II/B, comparative genomics in Brassicaceae was traditionally conducted relative to the ACK (n = 8). The recent update and improvement of the *Ae arabicum* genome sequence (Nguyen et al., 2019) to chromosome-level assembly of the eleven linkage groups has allowed us to apply the system of GBs previously used for other Brassicaceae lineages to tribe Aethionemeae. Compared to other Brassicaceae species, the GBs of the ACK are broken into multiple sub-blocks in *Aethionema*. Interestingly, the genomes of tribe Arabideae (Mandáková et al., 2020) and lineage III/E (Mandáková et al., 2017a) also contain a higher number of sub-blocks than "diploid" genomes from lineage I/A and II/B. A high number of within-block breaks is also observed in diploidized mesopolyploid genomes (e.g., Lysak et al., 2016; Mandáková et al., 2017b) or meso-neopolyploid ones, such as allohexaploid genomes of *Camelina sativa* (Mandáková et al., 2019) and *B. rapa* (Cheng et al., 2013) from lineage I/A and II/B, respectively. However, mesopolyploidization has not been observed in any tribe belonging to lineage III/E (Mandáková et al., 2017a,b), and the Arabideae have also not undergone a WGD post-dating the At-alpha (Willing et al., 2015; Mandáková et al., 2020). Thus, there seem to be two different reasons for elevated fractionation of GBs. While homeologous and ectopic recombination between the duplicated GBs, accompanying post-polyploid diploidization of the mesopolyploid genomes, explains the high number of within-block breakpoints in these genomes, the very definition of ACK and 22 building blocks (Schranz et al.,

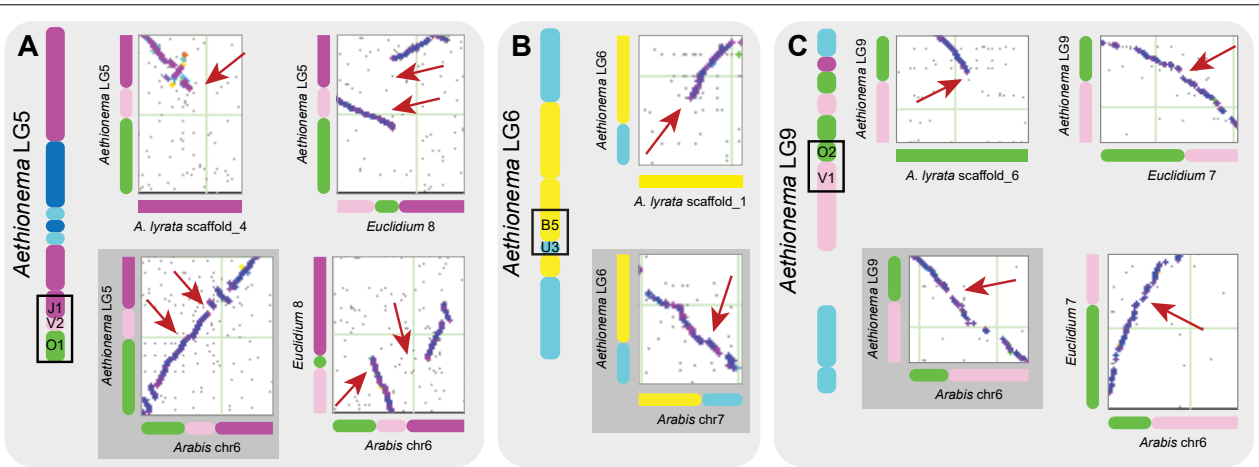

FIGURE 3 | Three syntenic regions in *Aethionema* and *Arabis*. **(A)** Sub-blocks O1, V2 and J1 are present as a unit on LG5 of *Aethionema* and chromosome 6 of *Arabis*, but not contiguous in *A. lyrata*. In *Euclidium*, large parts of O and parts of V and J are missing, and the O-V fragment is inverted relative to *Aethionema* and *Arabis*. **(B)** U3 and B5 are contiguous in *Aethionema* on LG6 and *Arabis* on chromosome 7, but not in *A. lyrata*. **(C)** V1 and O2 on LG9 of *Aethionema* and chromosome 6 of *Arabis* are associated, which is also detected in *Euclidium*, but not in *Arabidopsis*. More detailed figures of the three genomic regions in all pairwise comparisons are shown in **Supplementary Figures 5–7**.

2006; Lysak et al., 2016) does not reflect the structure of more ancestral genomes of lineage III/E, Arabideae and *Aethionema*. The ACK represents a diploidized genome derived from a more ancestral paleotetraploid genome most likely resembling that of *Ae. arabicum* and to lesser extent those of Arabideae and lineage III/E tribes. With the caveat that phylogenetic position of some crucifer genera and clades remains unresolved as evident by low support values at deeper nodes (Nikolov et al., 2019) and conflicting topologies between studies (e.g., Beilstein et al., 2010; Huang et al., 2016; Nikolov et al., 2019), the phylogenetic placement of ACK is revisited as an ancestral genome of lineage I/A and lineage II/B (Mandáková et al., 2017a).

The backbone phylogeny of Brassicaceae has been a subject of debate in recent years, with conflicting signals from plastome and nuclear genome data, and low resolution at deeper nodes despite large data sets. The use of non-nucleotide genomic data, such as patterns of synteny, may thus help in resolving these nodes. Tools to reconstruct phylogenetic trees based on genome rearrangement patterns have been developed recently (Drillon et al., 2020; Zhao et al., 2020), and first angiosperm-wide phylogenetic reconstructions based on synteny data have shown that these methods may provide alternative phylogenetic evidence for controversial nodes (Zhao et al., 2020). Our analysis of the genome structure of *Aethionema* and *Arabis* in the context of the entire Brassicaceae family was aimed at identifying evidence for the phylogenetic position of Arabideae relative to lineage II/B, where the tribe is consistently placed using plastome (Franzke et al., 2011; Hohmann et al., 2015; Guo et al., 2017; **Figure 3A**), but not nuclear genome data (Nikolov et al., 2019; **Figure 3B**), and lineage III/E. Comparative cytogenomic analyses of lineage III/E and Arabideae genera have revealed extensive chromosome reshuffling in these potentially earlier diverging branches of Brassicaceae (Willing et al., 2015; Mandáková et al., 2017a). Our analysis of GBs in *Ae. arabicum* shows that four block associations are shared between *Aethionema* and *Arabis*, indicating that these associations may have been the ancestral state. The breakpoints of these blocks and sub-blocks coincided with assembly borders in the genome sequence of *Euclidium* in one case, and thus we had to rely on evidence from cytogenetic data to infer the status of synteny at this region in lineage III/E. One shared association was observed in *Aethionema*, *Arabis* and *Euclidium*. This boundary notably does not involve any within-block borders, but the outer borders of blocks O and V (from AK 6 and AK7, respectively). Two further associations were shared between *Aethionema* and *Arabis* on LG5 of the *Aethionema* genome. The three sub-blocks (O1-V2-J1) were detected in the same order in *Aethionema* and *Arabis*, but the genome assembly of *Euclidium* indicated two gaps and an inversion at this location. Interestingly, here the V2 sub-block was not located at the edge of GB V from ACK, thus only internal breakpoints were involved.

The presence of identical characters in different lineages can, in short, be explained by two different processes: Either they are derived and originated independently in the respective lineages, or they are ancestral and were lost sometime in the past in the lineages that do not contain them. Two possible explanations and evolutionary scenarios may thus be invoked for the interpretation of our results. The first scenario follows previous interpretations of the ACK as the ancestral genome of Brassicaceae. In this case, the rearranged genomes of lineage III/E, *A. alpina* and *Ae arabicum* are derived from an ancestral Brassicaceae genome similar to the ACK. Their apparent similarity could be the result of frequent reuse of breakpoints. In the second scenario, the blocks from the ACK are the derived state and originated from an ancestral Brassicaceae genome somewhat resembling the genomes of *Aethionema*, *Arabis* and lineage III/E. Having a lower number of required changes, this seems to be the slightly more parsimonious scenario given our current data, and synteny with continuous At-α derived blocks additionally supports our claim.

Altogether, our results suggest that the *Arabis* clade diverged first within the Brassicaceae crown group, followed by lineage III/E and finally the most species-rich groups of lineages I/A and II/B (see **Figure 1C**).

To further advance our understanding of genome evolution in Brassicaceae, genome reconstruction of the family's most recent common ancestor, the post At-α genome, before divergence of *Aethionema*, is needed. This would allow for a redefinition of GBs relative to this presumed ancestral genome and for analysis of genome evolution in all lineages of the family. Whereas multiple high-quality genomes from lineages I/A and II/B are available, comparable genome sequences are not available yet for other crucifer clades. Chromosome-level assemblies from lineage III/E would allow us to test hypotheses regarding the backbone phylogeny and placement of lineage III/E as well as tribe Arabideae in more detail. In particular, the similarity of lineage III/E genomes with the ACK should be studied further. Additionally, the genome sequences of other *Aethionema* species, preferably some that diverged from *Ae. arabicum* early in the evolution of the tribe, would allow us to determine an ancestral karyotype of tribe Aethionemeae and to conclude whether the eleven *Aethionema* linkage groups represent the relic At-α genome frozen in time or a reshuffled paleotetraploid genome. This would also give us the opportunity to study the genome evolution of this species-poor sister clade, and could shed some light on why Aethionemeae did not diversify like the rest of Brassicaceae.

MATERIALS AND METHODS

Genomic Block Identification

We identified syntenic blocks in the updated reference genome of *Ae. arabicum* v.3 (Nguyen et al., 2019) relative to the 22 GBs in ACK (Schranz et al., 2006; Lysak et al., 2016). Note that throughout the manuscript, we refer to bioinformatically or cytogenomically detected syntenic blocks in extant species as "syntenic blocks" or simply "blocks" and to those of the ACK as "genomic blocks" or "GBs." The CoGe platform[2] (Lyons et al., 2008) was used to detect syntenic regions between *Ae. arabicum* and *A. thaliana*, as GBs in the ACK are defined using the *A. thaliana* gene IDs as start and end coordinates. Orthologous genes were identified using the BlastZ algorithm, and synteny was identified using DAGchainer (Haas et al., 2004). To obtain larger syntenic regions, we set the maximum distance between two matches in DAGchainer to 25, and only retained blocks with a minimum number of 20 retained pairs. For reasons of consistency, an exception was made for block G, which could only be identified with default settings, as it only contained six genes. We merged syntenic regions using QuotaAlign (Tang et al., 2011) with a maximum distance of 50 genes. In order to differentiate between syntenic blocks representing the GBs and those retained from At-α, we calculated the synonymous substitution rate (K_s) using CodeML (Yang, 2007) as implemented in CoGe. For values <2, K_s is relatively linear with time (Vanneste et al., 2013) and can be used to distinguish between orthologs between species, At-α derived genes or blocks with a K_s value around 0.8 (Schranz et al., 2012), recently duplicated genes, and duplicated genes with an even older origin, for example from the At-β WGD. Blocks with median K_s values corresponding to a WGD were discarded. Note that due to the additive effect of lineage-specific substitutions, K_s values for pairwise comparisons are higher with longer divergence time. In the *Aethionema-Arabidopsis* comparison, orthologs had a mean K_s of 0.77, while it was 0.42 in the Arabis-Arabidopsis comparison; mean Ks values of At-α derived paralogs were 1.37 and 1.01, respectively. Additionally, we checked each syntenic block for redundancy, i.e., if a block spanning the same part of the ACK was present more than once; the block with lower K_s was retained. The *Aethionema* blocks were generated from the remaining syntenic blocks; in the few cases where neighboring blocks from the same GB were not merged by QuotaAlign because of their distance, we merged them manually. Note that while gene names from *A. thaliana* define the borders, direction of blocks is given relative to the ACK. For visualization, we generated a syntenic dotplot and K_s plot using *Ae. arabicum* and *Arabidopsis lyrata*, a species with high similarity to the ACK. In this case, default parameters were set for DAGchainer and syntenic regions were not merged. Minimum length of chromosomes was set to 5,000,000 bp to retain only chromosomes from the genomes.

Comparison With Other Species

We compared the GBs from Aethionema with those from other species by running similar CoGe analyses with the following three species pairs: *Ae. arabicum* – *A. lyrata* (representative of lineage I/A and close to ACK), *Ae. arabicum* – *A. alpina* (unclear phylogenetic position), *Ae. arabicum* – *E. syriacum* (representative of lineage III/E) and *A. alpina* – *A. thaliana*. Blocks were only reconstructed for *A. alpina* (using the same parameters as above) and used to identify boundaries between (sub-)blocks shared between *Aethionema* and *Arabis*. Syntenic dotplots of these regions were finally compared between all species. Minimum length of chromosomes was again set to 5,000,000 bp to retain only chromosomes from the genomes, except for *Euclidium,* where shorter chromosomal length of 500,000 bp was allowed. The genome sequence of *Euclidium* is not quite assembled on a chromosomal-level, and block boundaries sometimes coincided with assembly borders. As we could not determine whether this was an artefact of assembly or the syntenic block boundary was located at the chromosome (arm) edge, we also used cytogenetic evidence from the literature for interpretation of our results.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study and download links are given in **Supplementary Table 1**.

[2]https://genomevolution.org/coge/

AUTHOR CONTRIBUTIONS

MS conceived the study. NW and T-PN analyzed the data. NW wrote the manuscript with input from TM, ML, and MS.

FUNDING

NW was supported by the German Research Foundation (DFG) with grant no. HO 6443/1. ML and TM were funded by a research grant from the Czech Science Foundation (Grant No. 15-18545S) to ML and by the CEITEC 2020 project (Grant No. LQ1601). T-PN and MS were funded by a grant from the Netherlands Organization for Scientific Research (NWO 0849.13.004) as part of the ERA-CAPS "SeedAdapt" consortium project (www.seedadapt.eu).

ACKNOWLEDGMENTS

We thank the members of the SeedAdapt Consortium and Dr. Laurie Grandont for fruitful discussions about the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00719/full#supplementary-material

REFERENCES

Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R., and Mathews, S. (2010). Dated molecular phylogenies indicate a miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18724–18728. doi: 10.1073/pnas.0909766107

Cheng, F., Mandáková, T., Wu, J., Xie, Q., Lysak, M. A., and Wang, X. (2013). Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *Plant Cell* 25, 1541–1554. doi: 10.1105/tpc.113.110486

Drillon, G., Champeimont, R., Oteri, F., Fischer, G., and Carbone, A. (2020). Phylogenetic reconstruction based on synteny block and gene adjacencies. *Molecular Biol. Evol.* msaa114. doi: 10.1093/molbev/msaa114

Edger, P. P., Hall, J. C., Harkess, A., Tang, M., Coombs, J., Mohammadin, S., et al. (2018). *Brassicales* phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Am. J. Bot.* 105, 463–469. doi: 10.1002/ajb2.1040

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366. doi: 10.1073/pnas.1503926112

Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A., and Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of *Brassicaceae*. *Trends Plant Sci.* 16, 108–116. doi: 10.1016/j.tplants.2010.11.005

Guo, X., Liu, J., Hao, G., Zhang, L., Mao, K., Wang, X., et al. (2017). Plastome phylogeny and early diversification of *Brassicaceae*. *BMC Genomics* 18:e176. doi: 10.1186/s12864-017-3555-3553

Haas, B. J., Delcher, A. L., Wortman, J. R., and Salzberg, S. L. (2004). DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20, 3643–3646. doi: 10.1093/bioinformatics/bth397

Hofberger, J. A., Lyons, E., Edger, P. P., Chris Pires, J., and Eric Schranz, M. (2013). Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family. *Genome Biol. Evol.* 5, 2155–2173. doi: 10.1093/gbe/evt162

Hohmann, N., Wolf, E. M., Lysak, M. A., and Koch, M. A. (2015). A time-calibrated road map of *Brassicaceae* species radiation and evolutionary history. *Plant Cell* 27, 2770–2784. doi: 10.1105/tpc.15.00482

Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., et al. (2016). Resolution of *Brassicaceae* phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol. Biol. Evol.* 33, 394–412. doi: 10.1093/molbev/msv226

Kiefer, C., Willing, E.-M., Jiao, W.-B., Sun, H., Piednoël, M., Hümann, U., et al. (2019). Interspecies association mapping links reduced CG to TG substitution rates to the loss of gene-body methylation. *Nat. Plants* 5, 846–855. doi: 10.1038/s41477-019-0486-489

Koch, M. A., and Al-Shehbaz, I. A. (2009). "Molecular systematics and evolution of 'wild' crucifers (*Brassicaceae* or *Cruciferae*)," in *Biology and Breeding of Crucifers*, ed. S. K. Gupta (Boca Raton, FL: CRC Press), 1–19.

Koch, M. A., German, D. A., Kiefer, M., and Franzke, A. (2018). Database taxonomics as key to modern plant biology. *Trends Plant Sci.* 23, 4–6. doi: 10.1016/j.tplants.2017.10.005

Lenser, T., Graeber, K., Cevik, O. S., Adiguzel, N., Donmez, A. A., Grosche, C., et al. (2016). Developmental control and plasticity of fruit and seed dimorphism in *Aethionema arabicum*. *Plant Physiol.* 172, 1691–1707. doi: 10.1104/pp.16.00838

Lenser, T., Tarkowská, D., Novák, O., Wilhelmsson, P. K. I., Bennett, T., Rensing, S. A., et al. (2018). When the BRANCHED network bears fruit: how carpic dominance causes fruit dimorphism in *Aethionema*. *Plant J.* 94, 352–371. doi: 10.1111/tpj.13861

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., et al. (2008). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with Rosids. *Plant Physiol.* 148, 1772–1781. doi: 10.1104/pp.108.124867

Lysak, M. A., Mandáková, T., and Schranz, M. E. (2016). Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Curr. Opin. Plant Biol.* 30, 108–115. doi: 10.1016/j.pbi.2016.02.001

Mabry, M. E., Brose, J. M., Blischak, P. D., Sutherland, B., Dismukes, W. T., Bottoms, C. A., et al. (2019). Phylogeny and multiple independent whole-genome duplication events in the *Brassicales*. *BioRxiv* [Preprint]. doi: 10.1101/789040

Mandáková, T., Hloušková, P., German, D. A., and Lysak, M. A. (2017a). Monophyletic origin and evolution of the largest crucifer genomes. *Plant Physiol.* 174, 2062–2071. doi: 10.1104/pp.17.00457

Mandáková, T., Li, Z., Barker, M. S., and Lysak, M. A. (2017b). Diverse genome organization following 13 independent mesopolyploid events in *Brassicaceae* contrasts with convergent patterns of gene retention. *Plant J.* 91, 3–21. doi: 10.1111/tpj.13553

Mandáková, T., Hloušková, P., Koch, M. A., and Lysak, M. A. (2020). Genome evolution in Arabideae was marked by frequent centromere repositioning. *Plant Cell* 32, 650–665. doi: 10.1105/tpc.19.00557

Mandáková, T., and Lysak, M. A. (2008). Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (*Brassicaceae*). *Plant Cell* 20, 2559–2570. doi: 10.1105/tpc.108.062166

Mandáková, T., Pouch, M., Brock, J. R., Al-Shehbaz, I. A., and Lysak, M. A. (2019). Origin and evolution of diploid and allopolyploid *Camelina* genomes were accompanied by chromosome shattering. *Plant Cell* 31:2596. doi: 10.1105/tpc.19.00366

Mérai, Z., Graeber, K., Wilhelmsson, P., Ullrich, K. K., Arshad, W., Grosche, C., et al. (2019). *Aethionema arabicum*: a novel model plant to study the light control of seed germination. *J. Exp. Bot.* 70, 3313–3328. doi: 10.1093/jxb/erz146

Mohammadin, S., Peterse, K., Kerke, S. J., van de Chatrou, L. W., Dönmez, A. A., Mummenhoff, K., et al. (2017). Anatolian origins and diversification of *Aethionema*, the sister lineage of the core Brassicaceae. *Am. J. Bot.* 104, 1042–1054. doi: 10.3732/ajb.1700091

Nguyen, T.-P., Mühlich, 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* 9:3521. doi: 10.1534/g3.119.400657

Nikolov, L. A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I. A., Filatov, D., et al. (2019). Resolving the backbone of the *Brassicaceae* phylogeny for investigating trait diversity. *New Phytol.* 222, 1638–1651. doi: 10.1111/nph.15732

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

Schranz, M., Lysak, M., and Mitchellolds, T. (2006). The ABC's of comparative genomics in the *Brassicaceae*: building blocks of crucifer genomes. *Trends Plant Sci.* 11, 535–542. doi: 10.1016/j.tplants.2006.09.002

Schranz, M. E., and Mitchell-Olds, T. (2006). Independent ancient polyploidy events in the sister families *Brassicaceae* and *Cleomaceae*. *Plant Cell* 18, 1152–1165. doi: 10.1105/tpc.106.041111

Schranz, M. E., Mohammadin, S., and Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr. Opin. Plant Biol.* 15, 147–153. doi: 10.1016/j.pbi.2012.03.011

Tang, H., and Lyons, E. (2012). Unleashing the genome of brassica rapa. *Front. Plant Sci.* 3:172. doi: 10.3389/fpls.2012.00172

Tang, H., Lyons, E., Pedersen, B., Schnable, J. C., Paterson, A. H., and Freeling, M. (2011). Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinform.* 12:102. doi: 10.1186/1471-2105-12-102

Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., Hinchliff, C. E., et al. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* 207, 454–467. doi: 10.1111/nph.13491

Vanneste, K., Van de Peer, Y., and Maere, S. (2013). Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* 30, 177–190. doi: 10.1093/molbev/mss214

Wilhelmsson, P. K. I., Chandler, J. O., Fernandez-Pozo, N., Graeber, K., Ullrich, K. K., Arshad, W., et al. (2019). Usability of reference-free transcriptome assemblies for detection of differential expression: a case study on *Aethionema arabicum* dimorphic seeds. *BMC Genom.* 20:95. doi: 10.1186/s12864-019-5452-5454

Willing, E.-M., Rawat, V., Mandáková, T., Maumus, F., James, G. V., Nordström, K. J. V., et al. (2015). Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nat. Plants* 1:14023. doi: 10.1038/nplants.2014.23

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Zhao, T., Xue, J., Kao, S., Li, Z., Zwaenepoel, A., Schranz, M. E., et al. (2020). Novel phylogeny of angiosperms inferred from whole-genome microsynteny analysis. *bioRxiv* [Preprint]. doi: 10.1101/2020.01.15.908376

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.

The reviewer ZL declared past co-authorship with one of the authors, MS, to the handling editor.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 September 2020
doi: 10.3389/fpls.2020.536507

Holocentric Karyotype Evolution in *Rhynchospora* Is Marked by Intense Numerical, Structural, and Genome Size Changes

Paula Burchardt[1], *Christopher E. Buddenhagen*[2†], *Marcos L. Gaeta*[1†], *Murilo D. Souza*[1], *André Marques*[3*] and *André L. L. Vanzela*[1*]

[1] Laboratório de Citogenética e Diversidade Vegetal, Departamento de Biologia Geral, CCB, Universidade Estadual de Londrina, Londrina, Brazil, [2] Forage Science, AgResearch Limited, Hamilton, New Zealand, [3] Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany

Cyperaceae is a family of Monocotyledons comprised of species with holocentric chromosomes that are associated with intense dysploidy and polyploidy events. Within this family the genus *Rhynchospora* has recently become the focus of several studies that characterize the organization of the holocentric karyotype and genome structures. To broaden our understanding of genome evolution in this genus, representatives of *Rhynchospora* were studied to contrast chromosome features, C-CMA/DAPI band distribution and genome sizes. Here, we carried out a comparative analysis for 35 taxa of *Rhynchospora*, and generated new genome size estimates for 20 taxa. The DNA 2C-values varied up to 22-fold, from 2C = 0.51 pg to 11.32 pg, and chromosome numbers ranged from 2*n* = 4 to 61. At least 37% of our sampling exhibited 2*n* different from the basic number *x* = 5, and chromosome rearrangements were also observed. A large variation in C-CMA/DAPI band accumulation and distribution was observed as well. We show that genome variation in *Rhynchospora* is much larger than previously reported. Phylogenetic analysis showed that most taxa were grouped in clades corresponding to previously described taxonomic sections. Basic chromosome numbers are the same within every section, however, changes appeared in all the clades. Ancestral chromosome number reconstruction revealed *n* = 5 as the most likely ancestral complements, but *n* = 10 appears as a new possibility. Chromosome evolution models point to polyploidy as the major driver of chromosome evolution in *Rhynchospora*, followed by dysploidy. A negative correlation between chromosome size and diploid number open the discussion for holokinetic drive-based genome evolution. This study explores relationships between karyotype differentiation and genome size variation in *Rhynchospora*, and contrasts it against the phylogeny of this holocentric group.

Keywords: C-CMA/DAPI banding, chromosome numbers, DNA C-value, flow cytometry, karyotype diversity

OPEN ACCESS

Edited by:
Hanna Weiss-Schneeweiss, University of Vienna, Austria

Reviewed by:
Ekaterina D. Badaeva, Russian Academy of Sciences, Russia
Eva Maria Temsch, University of Vienna, Austria
Tae-Soo Jang, Chungnam National University, South Korea

****Correspondence:***
André L. L. Vanzela
andrevanzela@uel.br
André Marques
amarques@mpipz.mpg.de

†ORCID:
Christopher E. Buddenhagen
orcid.org/0000-0002-3016-1054
Marcos L. Gaeta
orcid.org/0000-0003-1260-9641

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

Received: *20 February 2020*
Accepted: *21 August 2020*
Published: *10 September 2020*

Citation:
Burchardt P, Buddenhagen CE, Gaeta ML, Souza MD, Marques A and Vanzela ALL (2020) Holocentric Karyotype Evolution in Rhynchospora Is Marked by Intense Numerical, Structural, and Genome Size Changes. Front. Plant Sci. 11:536507. doi: 10.3389/fpls.2020.536507

INTRODUCTION

The size, morphology and composition of chromosomes have been useful parameters for comparing karyotypes of phylogenetically related species, and to resolve some taxonomic conflicts (Guerra, 2000). These features have been widely regarded as drivers of evolutionary processes (Comai, 2005; Doyle et al., 2008), as they are the result of duplication or deletion of entire chromosomes, polyploidy, fission, fusion, and/or chromosome translocation (Greilhuber, 1995; Luceño and Guerra, 1996; Schubert and Lysak, 2011). We focus on karyotype evolution of Rhynchosporeae, a monophyletic clade in the Cyperaceae that presents holocentric chromosomes, a trait that supposedly originated independently in four distinct clades of plants (Melters et al., 2012).

Greilhuber (1995) considered holocentric chromosomes to be a synapomorphy of the Cyperid clade (Thurniaceae, Juncaceae and Cyperaceae), and this characteristic has been accepted for *Rhynchospora* since then (Vanzela et al., 2000; Vanzela and Colaço, 2002; Arguelho et al., 2012). However, Guerra et al. (2019) reported that *Juncus* L., a genus regarded as exclusively holocentric, contains monocentric species. Also recently it was reported that Prionium serratum (Thurniaceae) is also monocentric (Baez et al., 2020). More detailed work in this area is needed to provide insights into the evolution of holocentric chromosomes in Cyperales, such as whether it arose independently in each of the major clades.

In holocentric chromosomes, kinetochore proteins are arranged along the chromosomes and their kinetic activity appears to be distributed along almost the entire chromatid surface (Melters et al., 2012; Heckmann and Houben, 2013). In this case, any fragments produced by chromosome fission/fusion may segregate regularly, making them more likely to be inherited during cell division, and this can lead to increases and decreases in chromosome numbers, giving rise to dysploidy (Mola and Papeschi, 2006; Arguelho et al., 2012). In contrast, fission events in monocentric chromosomes may generate acentric fragments that are unable to segregate normally and are lost during cell division (Carrano and Heddle, 1973; Escudero et al., 2015).

Cyperaceae members are well known for having large chromosome number variation associated with chromosome rearrangements, and *Carex* L. has the highest record of chromosome fission and fusion (Hipp, 2007; Roalson, 2008; Hipp et al., 2009). The increase in chromosome number by polyploidy has also been proposed in genera such as *Eleocharis* R.Br. (Da Silva et al., 2010; Zedek et al., 2010) and *Rhynchospora* Vahl (Vanzela et al., 2000). High chromosome numbers were also found in *Cyperus cyperoides* Kuntze (n = 112; Tejavathi, 1988), *C. esculentus* L. (n = 104; Sharma, 1970), *Carex hirta* L. (n = 56–57; Luceño, 1994), and *Rhynchospora faurieri* Franch. (n = 31; Hoshino, 1987). But, the reduction below the probable basic chromosome number (x = 5) also happens in Cyperaceae *via* dysploidy, such as n = 2 in *Rhynchospora* (Vanzela et al., 1996) and n = 3 in both *Eleocharis* (Da Silva et al., 2005) and *Fimbristylis* Vahl (Rath and Patnaik, 1977). Lower numbers have also been reported in other holocentric families, such as Juncaceae (Malheiros et al., 1947), Droseraceae (Kondo et al., 1994) and Convolvulaceae (Pazy and Plitmann, 1987; Pazy and Plitmann, 1994).

A number of cytogenetic studies compared *Rhynchospora* karyotypes (Luceño et al., 1998a; Vanzela and Guerra, 2000; Vanzela et al., 2000; Vanzela et al., 2003; Sousa et al., 2011; Michelan et al., 2012; Cabral et al., 2014; Marques et al., 2015; Ribeiro et al., 2017; Ribeiro et al., 2018), but none compared karyotype diversity with heterochromatin distribution and DNA C-value variation together, encompassing different clades. The genus *Rhynchospora* is the third largest clade in Cyperaceae (Araújo et al., 2012), with ca. 350 species distributed worldwide (Buddenhagen et al., 2017). Chromosome numbers vary from $2n$ = 4 in *R. tenuis* (Vanzela et al., 1996) to $2n$ = 62 in *R. faurieri* (Hoshino, 1987), although $2n$ = 10 is the most common number. Reports suggest a wide diversity with odd and even numbers, like $2n$ = 4, 5, 8, 10, 12, 18, 20, 24, 26, 30, 36, 37, 45, 48, 50, and 58 (Luceño et al., 1998a; Luceño et al., 1998b; Vanzela et al., 2000; Arguelho et al., 2012; Ribeiro et al., 2018), and this high variability can also be observed within a single species, e.g. *R. globosa* ($2n$ = 24, 36, 37, 45, 48, 50, and 58).

Rhynchospora has been used as a model for detailed studies aiming to characterize holocentric chromosome structure and adaptations taking place in these organisms (Cabral et al., 2014; Marques et al., 2015; Marques et al., 2016; Rocha et al., 2016). A prior study examining the interspecific relationships in a phylogenetic context using cytogenetic data and DNA content has suggested polyploidy as the main driver of karyotype and genome evolution in *Rhynchospora* (Ribeiro et al., 2018). Despite this and the high diversity of the genus, there are few studies approaching phylogenetic relationships with genomes and karyotype data. The most comprehensive analysis of the genus was based on a traditional herbarium based taxonomic study of 211 species (Kükenthal, 1949; Kükenthal, 1950a; Kükenthal, 1950b; Kükenthal, 1951).

In order to perform a comprehensive assessment of evolutionary forces that have a role in *Rhynchospora* karyotype differentiation, the number of Brazilian populations and species was expanded and phylogenetically compared, including never studied species. Efforts were intended to compare intra- and interspecific variations in chromosome number, as well as to estimate DNA C-values and C-CMA/DAPI bands distribution. Data were compared and analyzed in a phylogenetic context, including samples from 14 different taxonomic sections of the genus. Our data provide a window into the group's intraspecific variation, which helps to support polyploidy and dysploidy as the major drivers of genome and karyotype evolution in *Rhynchospora*, and indicate the importance of wide sampling to include possible inter and intraspecific variations in holocentric karyotypes.

MATERIALS AND METHODS

Plant Material

Living plants of 24 taxa were collected from different localities in Brazil. Plants were grown in pots in the greenhouse of the Center for Biological Sciences at the State University of Londrina and

vouchers were deposited in the Herbarium of Londrina State University (FUEL Herbarium). The **Supplementary data 1** for this study contains descriptions of the plant material, chromosome counts and nuclear DNA measurements, including previously published data gathered from the literature. Some chromosome records obtained from previously fixed materials, and that are part of the Laboratório de Citogenética e Diversidade Vegetal (LCDV, UEL, Brazil) collection, were used for comparison purposes and are indicated in tables and figures as "from LCDV" and includes citations.

Genome Size Estimates

Holoploid genome sizes (2C-values) were assessed for available living specimens of *Rhynchospora*. *Raphanus sativus* L. 'Saxa' (2C = 1.11 pg; Doležel et al., 1998), *Solanum lycopersicum* L. 'Stupicke polni tyckove rane' (2C = 1.96 pg; Doležel et al., 1992), and *Pisum sativum* L. 'Ctirad' (2C = 9.09 pg; Doležel et al., 1998), were each used as internal standards. Young leaves were processed immediately after collection. Fragments (1 cm^2) of young leaves of the sample and the internal standard were chopped together (Galbraith et al., 1983), for 30 s (Noirot et al., 2005), with a brand-new razor blade in a 60 mm × 10 mm Petri dish containing 125 µl of OTTO-I lysis buffer (Otto, 1990), supplemented with 2.0 mM polyethylene glycol and 50 µg/ml RNAse. To the nuclei suspensions, another 125 µl of OTTO-I lysis buffer was added, and the homogenates were sieved through 25 µm nylon filters into 2.0 microcentrifuge tubes, then centrifuged at 100× g for 5 min. The supernatant of each sample was poured out, and the pellet resuspended and incubated for 5 min in 25 µl of OTTO-I lysis buffer. The suspensions were stained with 375 µl OTTO-II solution (Otto, 1990; Doležel and Göhde, 1995), supplemented with 75 µM propidium iodide, 2.0 mM polyethylene glycol and 50 µg/ml RNAse (Doležel et al., 1992; Meister, 2005). The staining step was performed in the dark for 40 min, followed by filtration through a 25 µm nylon mesh. DNA content of at least 10,000 stained nuclei was determined for each sample using a BD Accuri C6 flow cytometer (BD Biosciences), using three independent DNA estimations of one to three individual plants. Chromosome counts were done for every population sampled, except for *R. pilosa*, for which we used a previously reported chromosome number. Total 2C-values were calculated as sample peak mean/standard peak mean × 2C DNA content of standard (pg). The hypothetical monoploid genome sizes (1C*x*) were calculated by dividing the 2C-values by the ploidy level. Pearson's correlation test and linear regression analysis were performed in the R statistical software environment with the ggplot2 package (Wickham, 2016). The average chromosome size was calculated for each sample, including genome size and chromosome count data available from our study and literature. Average chromosome size was calculated from genome size data as 2C (Mbp)/2*n* (1 pg = 978 Mbp; Doležel et al., 2003) and plotted together with 2*n* values for comparison.

Cytogenetic Analyses

Three to five plants of each of the 23 species, comprising 37 populations, provided meristems for cytogenetic analysis. Root tips from greenhouse cultivated plants were pretreated in a solution containing 2 mM 8-hydroxyquinoline at 10°C for 24 h, fixed in 3:1 ethanol:glacial acetic acid (v:v) for 24 h at room temperature, and stored at –20°C. Samples were digested in 2% cellulase plus 20% pectinase (w/v), both Sigma, at 37°C for 45 to 60 min. For conventional staining, digested root tips were washed in distilled water, hydrolyzed in HCl 1 M for 10 min at 60°C, and squashed in a drop of 60% acetic acid. Coverslips were removed after freezing in liquid nitrogen. Samples were stained in 3% Giemsa and mounted in Entellan (Merck).

Chromomycin A_3 (CMA) and 4-6 diamidino-2-phenyl indole (DAPI) serve to identify heterochromatic blocks by binding to GC- and AT-rich repeats in the genome, respectively. C-CMA/DAPI banding was carried out using fixed root tips digested for 3 to 4 h in a mixture of 2% cellulase and 20% pectinase (w/v), and squashed in a drop of 60% acetic acid. Coverslips were removed in liquid nitrogen and, after air drying for 3 days, samples were treated for C-banding procedure [45% acetic acid at 60°C for 10 min, 5% $Ba(OH)_2$ at room temperature for 10 min and 2× SSC, pH 7.0, at 60°C for 1 h and 30 min]. Subsequently, samples were stained with CMA_3 for 90 min and DAPI for 30 min, as described by Vanzela and Guerra (2000) and mounted in glycerin:McIlvaine buffer (1:1, v:v), pH 7.0, and 2.5 mM $MgCl_2$.

Slides were examined using a Leica DM 4500B epifluorescence microscope and images were acquired using a Leica DFC 300FX camera. All the images were optimized for contrast and brightness using Gimp-2.8 and Inkscape 0.92.3 programs on the Linux platform.

Phylogenetic Comparison

A phylogenetic analysis was undertaken for *Rhynchospora* taxa for which either chromosome number or genome size data is available. This was not intended as a new phylogenetic proposition for the genus, but as a means of complementing the cytogenetic analysis and supporting further comparisons. For this purpose, we extracted plastid sequences from an Illumina short-read target capture data for 35 of 115 available *Rhynchospora* species (Buddenhagen, 2016; Buddenhagen et al., 2016) using the Geneious mapper tool with default parameters available in Geneious version 7.1. Plastome-derived sequences from 35 species of *Rhynchospora* were obtained using the plastomes of *Hypolytrum nemorum* (Vahl) Spreng. (GenBank accession number NC_036036.1) and *Carex neurocarpa* Mack. (GenBank accession number NC_036037.1) as a reference sequence for mapping (see **Supplementary data 2**). A consensus sequence was made for each sampled species, and as criteria, gene regions with more than 15% gaps were stripped, and alignments for gene regions with more than one missing taxon were not used. Filtered regions were aligned with PASTA (Mirarab et al., 2015) and gene alignment files were concatenated in Geneious. Partitions used for tree estimation corresponded to the annotated regions. There were 54 distinct plastid regions, listed in **Supplementary data 3**, used in the final alignment (43,979 bases). The concatenated alignment was then used to estimate maximum likelihood phylogeny using the software IQ-TREE (Nguyen et al., 2015) with 1,000 bootstraps. The consensus tree was edited in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Ancestral Chromosome Number Reconstruction

To test the best models for chromosome evolution in *Rhynchospora* we applied the RevBayes (Freyman and Höhna, 2018) implementation of the ChromEvol models (Glick and Mayrose, 2014). The program utilizes a model of anagenetic transition rates including gain (fission, dysploidy) or loss (fusion) of chromosomes, polyploidization, and demi-polyploidization. The best fitting model was assessed using the AIC (Glick and Mayrose, 2014). Furthermore, the ancestral states of chromosome number along the branches were estimated using PastML (Ishikawa et al., 2019) applying two prediction methods, Maximum Likelihood (JOINT+F81) and Maximum Parsimony (Accelerated Transformation), respectively. Since these analyses can take only one state per sample, only the lowest chromosome number for each species was used in the case of samples with more than one cytotype, with different chromosome numbers and ploidy levels.

RESULTS

Genome Size Variations

Samples from 39 populations were studied, comprising 24 taxa of *Rhynchospora*. Genome size data and other details are presented in **Figures 1A, B** and **Supplementary data 1**. The 2C-values showed a 22-fold variation between the lowest, 0.51 pg in *R. rugosa* ($2n = 36$), and the highest, 11.32 pg in *R. globosa* ($2n = 61$). Most samples fall within a narrow range of 2C-values (from 0.51 pg to 1.28 pg). Outliers to this parameter are *R. pubera*, *R. tenuis* subsp. *austrobrasiliensis* (referred to as *R. austrobrasiliensis*) and *R. globosa*, all of which have larger holoploid genome sizes. When we evaluated the hypothetical monoploid 1C*x*-value distribution, a difference of about 15× appeared, from 0.06 pg in the polyploid *R. pilosa* ($2n = 10x = 50$) to 0.94 pg in the polyploid/dysploid *R. globosa* with $2n = 12x = 61$, and about 27× in relation to the diploid *R. pubera* with $2n = 10$ (**Figure 1A** and **Supplementary data 1**). Our estimates of nuclear DNA content did not differ much among populations of diploid *R. breviuscula*, *R. nervosa* subsp. *ciliata*

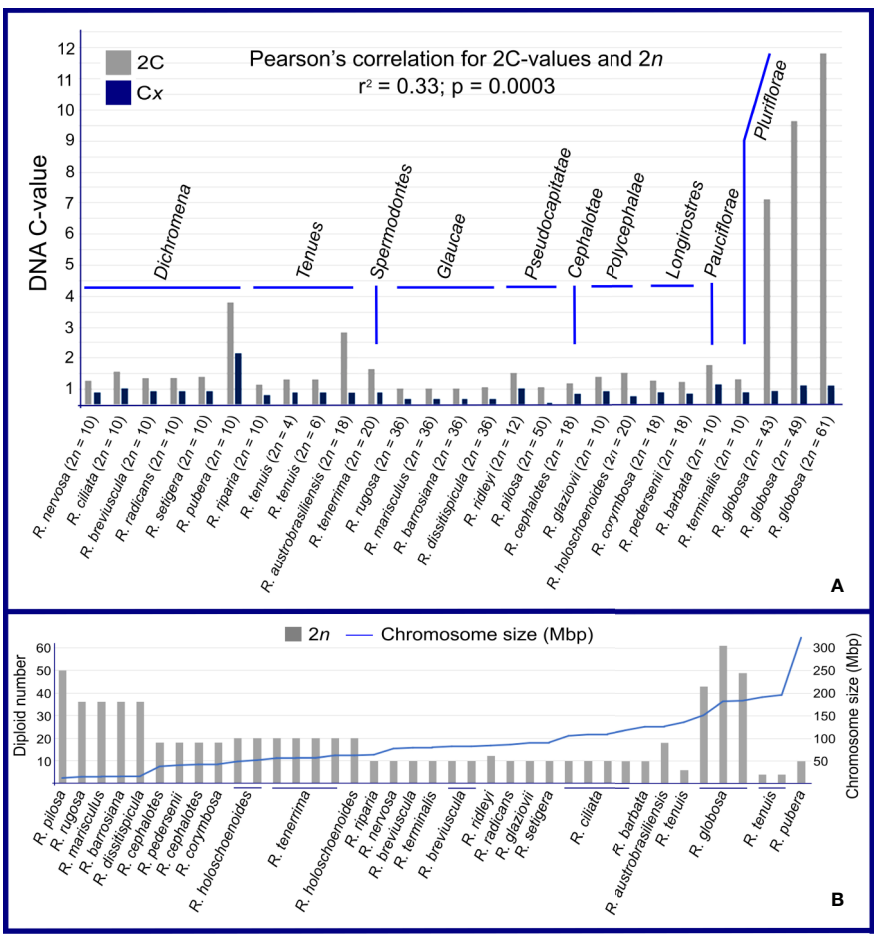

FIGURE 1 | Genome sizes in *Rhynchospora*. **(A)** Comparison between holoploid genome sizes (2C- and 1 Cx-values). DNA content varied throughout the genus, within sections and among taxa with the same chromosome number. Note that section *Glaucae* is the exception, with stable DNA contents and chromosome numbers. **(B)** This graph shows a negative association between 2*n* and average chromosome size in Mbp in almost every sample. The opposite is true for *Rhynchospora globosa*, which presented high chromosome numbers and large chromosomes.

(referred to as *R. ciliata*), and tetraploid *R. holoschoenoides* and *R. tenerrima* species. However, different polyploid populations of *R. globosa* had divergent genome sizes.

We compared genome size differences among sections and summarize here the most interesting observations. The sect. *Glaucae* and sect. *Longirostres* have the most constant 2C DNA contents, differing 1.05× and 1.09× in relation to *Dichromena*, *Tenues*, *Pseudocapitatae*, and *Polycephalae*. Four cases drew our attention: i) in sect. *Dichromena*, the diploid *R. pubera* with $2n = 10$ has 3 times more DNA content than the other diploid species in the section, ii) in sect. *Tenues*, the polyploid *R. austrobrasiliensis*, with $2n = 18$, exhibited three times more DNA content than *R. tenuis* with $2n = 6$, iii) in sect. *Pseudocapitatae*, *R. ridleyi* with $2n = 12$ presented twice the 2C-value as the polyploid *R. pilosa* with $2n = 50$, and iv) in the polyploid/dysploid *R. globosa* (sect. *Pluriflorae*), the 2C-value variation was superior to the numerical changes involving few chromosomes, i.e. from 43 to 49 (**Figure 1A**). In sect. *Pluriflorae*, a 13.9× difference was observed between *R. terminalis* var. *rosemariana* with $2n = 10$ and *R. globosa* with $2n = 61$. *Pearson's* correlation test between 2C-values and $2n$ chromosome numbers resulted in a low correlation ($r^2 = 0.33$; $p = 0.0003$). Average chromosome sizes varied from 11 Mbp in *R. pilosa* ($2n = 50$) to 320 Mbp in *R. pubera* ($2n = 10$). We found a clear negative correlation between chromosome numbers ($2n$) and average chromosome sizes (**Figure 1B**). Except for *R. globosa*, most species with high chromosome numbers tended to show smaller average chromosome sizes, while species with few chromosomes showed larger average chromosome sizes.

Karyotype Diversity

Conventional cytogenetic analysis showed a wide variety of karyotypes, with numbers ranging from $2n = 4$ to 61 and chromosomes differing in size from about 1.3 to 7 μm (**Figure 2**). In this data set, new chromosome counts were reported for *R. albobracteata*, *R. terminalis* var. *rosemariana*, *R. dissitispicula* and *R. pedersenii*, new populations for *R. barbata*, *R. barrosiana*, *R. cephalotes*, *R. corymbosa*, *R. globosa*, *R. holoschoenoides*, *R. marisculus*, *R. nervosa* subsp. *nervosa* (referred to as *R. nervosa*), *R. ciliata*, *R. riparia*, *R. rugosa*, *R. tenuis*, *R. austrobrasiliensis*, and a new chromosome race for *R. tenuis* ($2n = 6$) and cytotypes for *R. globosa* ($2n = 43$, 49, and 61). Chromosome numbers for all the taxa and populations, including some data from literature, are summarized in **Supplementary data 1**. To follow a logical cytotaxonomic order, results are also presented in **Supplementary data 4–7**, according to their phylogenetic relatedness and also considering Kükenthal's taxonomic classification (1949; 1950a; 1950b; 1951). Within this numerical chromosome variation, numbers derived from $x = 5$ ($2n = 10$, 20, and 30) were the most common. Multiples of $x = 5$ were found in over 63% of the samples, while multiples of $x = 6$ or 9 were registered in a smaller number of accessions (~23% and ~14%, respectively).

Among representatives of sect. *Dichromena*, karyotypes were more symmetrical in relation to the other sections (**Figure 2** and **Supplementary data 4**). Despite this relative chromosome homogeneity, *R. pubera* (**Supplementary data 4**) has chromosomes almost 3× larger than other species, and a twofold size difference among its chromosomes. Contrastingly, all cytotypes in *R. nervosa* ($2n = 10$, 20, and 30) had symmetrical karyotypes (**Supplementary data 4**). In sect. *Tenues* (**Supplementary data 5**), a gradual reduction in chromosome size was observed in all diploid, dysploid, and polyploid species. The chromosome race in *R. tenuis* with $2n = 5$ stands out for having one chromosome 3 times smaller than the largest chromosome. Section *Spermodontes*, represented here by *R. tenerrima* (**Figure 2F**), showed much smaller chromosomes than those from sect. *Tenues*. Sections *Glaucae*, *Cephalotae*, *Polycephalae*, and *Longirostres* (**Figures 2G, H** and **Supplementary data 6**) exhibited up to twofold differences in chromosome sizes within karyotypes, and also some of the smallest chromosomes in *R. rugosa*, *R. marisculus, R. dissitispicula* and *R. cephalotes* (**Figure 2G** and **Supplementary data 6**). Section *Pluriflorae* exhibited the most significant variation among samples. While the diploid *R. terminalis* and the tetraploid *R. albobracteata* presented symmetrical karyotypes (**Figures 2I, J**), *R. globosa* had strongly asymmetrical karyotypes among its polyploid and dysploid cytotypes ($2n = 36$ to 61) (**Figure 2K** and **Supplementary data 7**).

C-CMA/DAPI Band Variation

Chromosome banding was performed to check for possible differences and similarities among karyotypes, in addition to the features already observed by conventional cytogenetics. As criteria, the occurrence, number and position of CMA^+, $DAPI^+$, and $CMA^+/DAPI^+$ bands were evaluated. Fifteen taxa were analyzed, and a high diversity of band profiles was observed (**Figures 3–5**). Variation included number as well as location (terminal, subterminal, and/or interstitial), and this was evident when we compared different populations of *R. nervosa* and *R. globosa*. In sect. *Dichromena*, interstitial or terminal/subterminal $CMA^+/DAPI^+$ or $CMA^+/DAPI^0$ bands were common to all species, and the accumulation of interstitial bands in *R. ciliata* was visible. $CMA^+/DAPI^+$ was observed only in *R. setigera*, *R. ciliata* (both with $2n = 10$), and a population of *R. nervosa* with $2n = 10$ and another with $2n = 20$ (**Figures 3A–F** and **Supplementary data 8**). *Rhynchospora pubera*, the species with the largest chromosomes and genome size in sect. *Dichromena*, accumulated fewer bands than other species with smaller genomes (**Figure 4A**).

Rhynchospora tenerrima (sect. *Spermodontes*), which is the closest species to sect. *Tenues*, exhibited both CMA^+ and an evident interstitial $DAPI^+$ band (**Figures 3G, H** and **4C**). In comparison to species from sect. *Tenues*, terminal CMA^+ bands prevailed, except for *R. austrobrasiliensis* which has insterstitial bands (**Figures 3G–I**, **4B** and **Supplementary data 9A–F**). Samples from sect. *Glaucae*, *Polycephalae* and *Longirostres* (**Figures 3** and **5** and **Supplementary data 10**) showed a predominance of terminal CMA^+ bands, with few interstitial CMA^+ bands in *R. holoschoenoides* (**Supplementary data 10E, F**). It is important to highlight that *R. corymbosa* from sect. *Longirostres* accumulated several terminal and interstitial $CMA^+/DAPI^+$ bands (**Supplementary data 10C, D** and **Figure 5D**). Species from sect. *Pluriflorae* presented variable banding profiles, specially samples of *R. globosa* (**Figures 3L–O**, **5C**, and

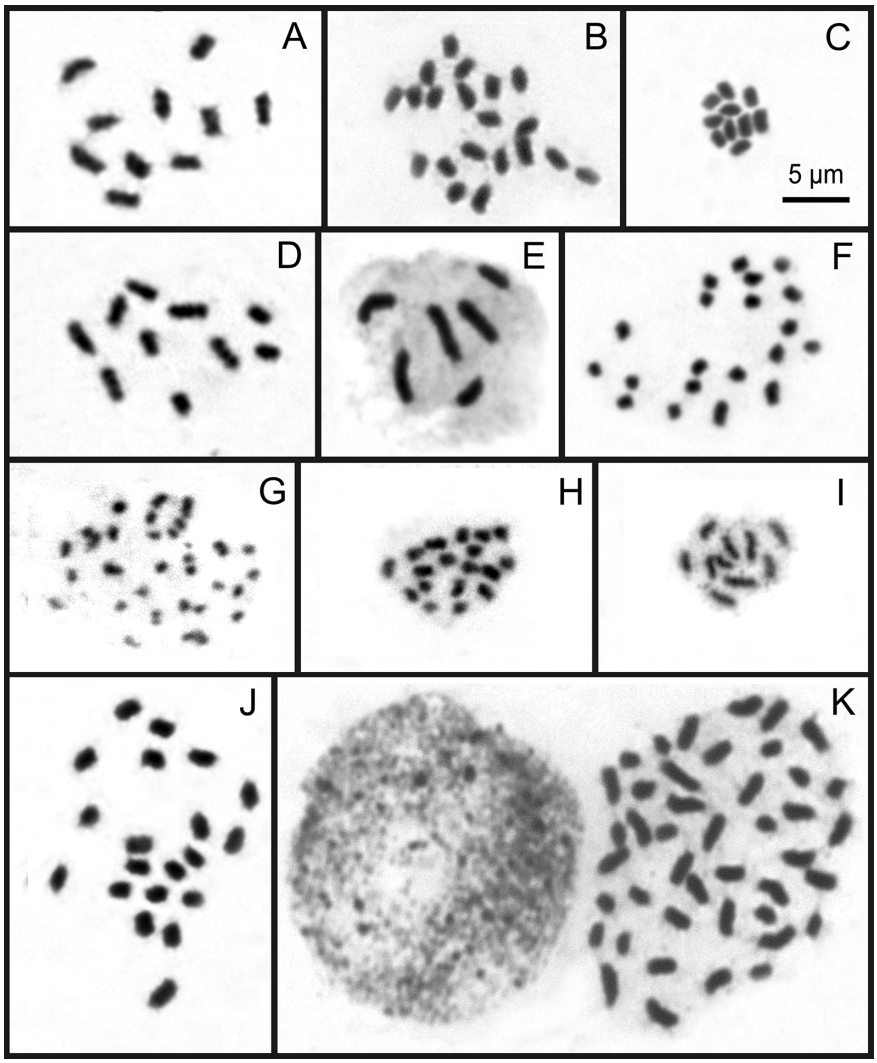

FIGURE 2 | Overview of karyotype diversity in holocentric species of *Rhynchospora*. Mitotic chromosomes of *R. nervosa* with 2*n* = 10 **(A)**, 2*n* = 20 **(B),** and *R. setigera* with 2*n* = 10 **(C)** from sect. *Dichromena.* Note a relative symmetry among of them. Mitotic chromosomes of *R. riparia* with 2*n* = 10 **(D)** and *R. tenuis* with 2*n* = 6 **(E)**, both from sect. *Tenues* and *R. tenerrima* with 2*n* = 20 **(F)** from sect. *Spermodontes.* Note that chromosomes of the polyploid *R. tenerrima* are smaller than those of closer species. The mitotic chromosomes of polyploid *R. dissitispicula* with 2*n* = 36 in **(G)** represents an asymmetrical karyotype with the smaller chromosomes sampled here. Metaphase of *R. pedersenii* (sect. *Longirostres*) with 2*n* = 18 **(H)**. Images in **(I–K)**, from sect. *Pluriflorae*, comprise the most variable karyotypes observed here. Prometaphase in *Rhynhcoposra terminalis* var. *rosemariana* with 2*n* = 10 **(I)** and metaphases in *R. albobracteata* with 2*n* = 20 **(J)** and *R. globosa* with 2*n* = 43 **(K)**, exhibit variability not only in number, but also in the sizes and symmetry of karyotypes. Note the accumulation of chromocenters in *R. globosa* interphase.

Supplementary data 11). The absence of $DAPI^+$ bands in *R. terminalis*, the accumulation of CMA^+ and $CMA^+/DAPI^+$ in *R. albobracteata* (**Supplementary data 10G–J**), and a highly variable banding distribution in *R. globosa* (see **Figure 5C** and **Supplementary data 11**) were evident.

In relation to intraspecific diversity and polymorphisms in C-CMA/DAPI bands, we compared diploid/polyploid samples of *R. nervosa* (**Figure 4A**) with cytotypes of *R. globosa* (**Figure 5C**), which is a complex of polyploid/dysploid taxa. The amount and position of CMA^+ and $CMA^+/DAPI^+$ bands varied among diploid samples of *R. nervosa*. Besides, the band profile observed in the polyploid did not represent the exact duplication of band profiles of diploids. Polymorphisms were observed in all three cases (**Figure 4A**). Populations of *R. globosa* were the most variable regarding band profiles, which often made chromosome pairing impossible. We can highlight the absence of bands on the two largest chromosomes and the presence of bands on the smallest chromosomes in the population with 2*n* = 36, and the inverse situation in populations with 2*n* = 43, 49, and 61, where the smallest ones have no bands and the largest ones accumulate more heterochromatic bands (**Figure 5C**). Polymorphisms in the occurrence and location of CMA and/or DAPI bands were

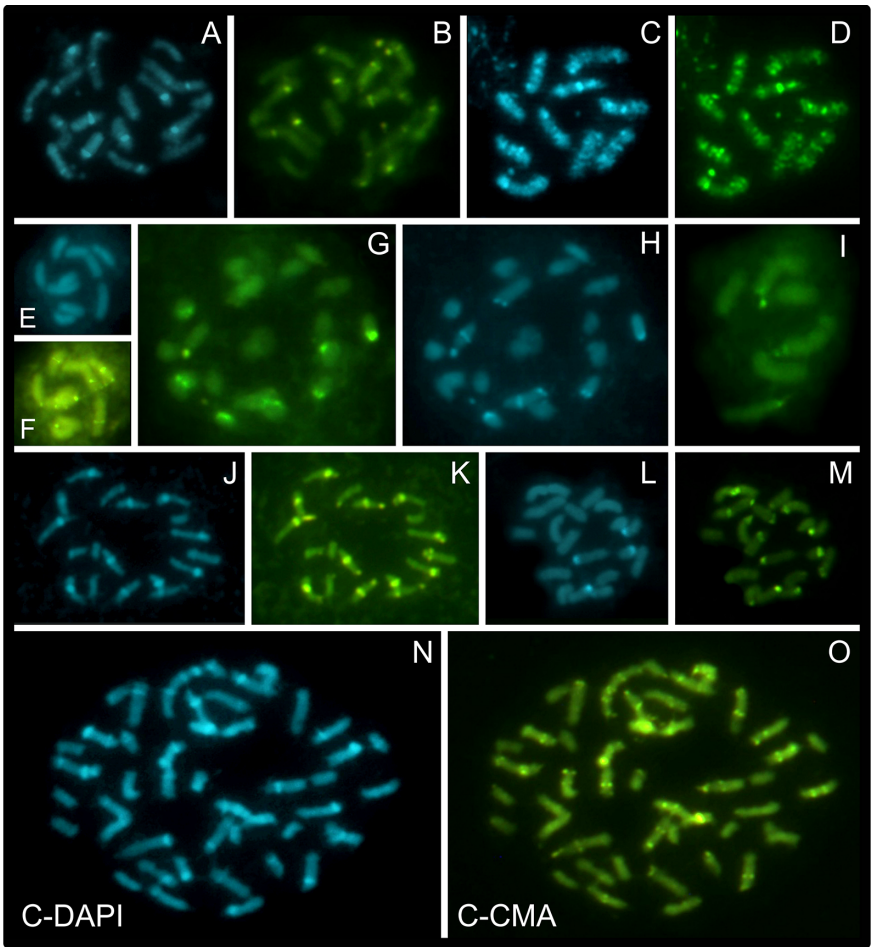

FIGURE 3 | Overview of C-CMA/DAPI banding profiles in holocentric species of *Rhynchospora* containing different chromosome numbers and sizes. Prometaphases of *R. nervosa* from Carrancas with 2*n* = 20 **(A, B)**, *R. ciliata* with 2*n* = 10 **(C, D)**, and *R. breviuscula* with 2*n* = 10 **(E, F)**, from sect. *Dichromena*. Note the accumulation of interstitial $CMA^+/DAPI^+$ bands in *R. ciliata* and the lack of $DAPI^+$ signals in the chromosomes of *R. breviuscula*. Prometaphases in *R. tenerrima* (2*n* = 20; sect. *Spermodontes*) **(G, H)**, and in *R. tenuis* (2*n* = 6; sect. *Tenues*) **(I)**. While *R. tenerrima* showed both CMA^+ and $DAPI^+$ bands, a closely related species, *R. tenuis*, only has CMA^+ signals. Prometaphase of *R. corymbosa* with large interstitial $DAPI^+/CMA^+$ blocks in several chromosomes, and some smaller terminal $DAPI^0/CMA^+$ ones **(J, K)**. Metaphase of *R. albobracteata*, with many terminal CMA^+ and fewer terminal $DAPI^+$ bands **(L, M)**, and prometaphase of *R. globosa* with 2*n* = 43, from Jaguariaíva **(N, O)**, both from sect. *Pluriflorae*. Note that band distribution in *R. globosa* is more diverse that in *R. albobracteata*.

detected in nine of the 15 species compared (indicated as red stars in **Figures 4** and **5**).

Phylogenetic Relationships and Ancestral Chromosome Number (ACN) Reconstruction

Species relationships were obtained for 37 taxa from a plastid concatenated alignment >50 Kb in length comprising >1,300 sequences. Despite the limited amount of species, a large number (54) of chloroplast coding loci were used, and the tree showed high support (>95 bootstrap) for most clades. All *Rhynchospora* taxa for which chromosome numbers and genome sizes are available were organized according to Kükenthal's classification and are presented beside the phylogenetic tree obtained from chloroplast sequences (**Figure 6**). Comparing **Figures 6A, B**, the majority of taxa in Kükenthal's sections were assigned to the same phylogenetic clade. In general, the phylogenetic analysis was in agreement with the previous taxonomic proposition regarding pars Diplostylae and Haplostylae, except for sect. *Pseudocapitatae* (Haplostylae) that grouped with sections belonging to pars Diplostylae. *Rhynchospora holoschoenoides*, *R. riedeliana*, and *R. barbata* were grouped with different taxa than Kükenthal had proposed. Also, sect. *Dichromena* and *Pseudocapitatae* were assigned to the same phylogenetic clade. Within every section/clade, taxa share the same basic chromosome number, however, their distribution along the phylogenetic tree does not seem to follow a particular order, neither does the occurrence of dysploidy or polyploidy, which are common in almost every clade, i.e., numerical and structural rearrangements appeared in all the clades.

To gain insights into major drivers of chromosome evolution in *Rhynchospora* we analyzed the frequency of chromosome number change events and performed an ancestral chromosome

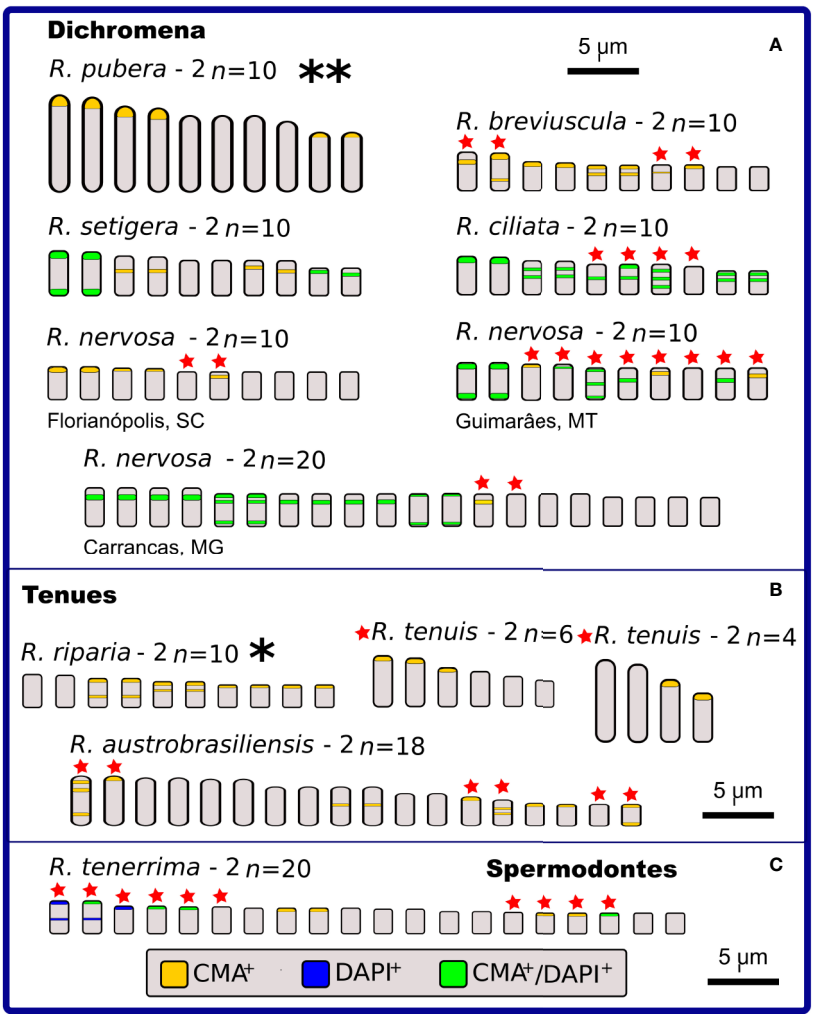

FIGURE 4 | Idiograms representing the physical mapping of C-CMA/DAPI banding in karyotypes of 12 species of *Rhynchospora*, including two data obtained from literature (Vanzela and Guerra, 2000), which are highlighted with * and **. Species are grouped in sections according to Kükenthal's classification. In sect. *Dichromena* **(A)**, five cytotypes exhibited polymorphisms in the banding location (indicated by red stars). Observe the large C-CMA/DAPI banding diversity among populations of *R. nervosa*. Only *R. pubera* and *R. setigera* showed regular distribution of bands. From the other two closest sections (*Tenues* and *Spermodontes*, **(B, C)**, respectively), only *R. riparia* exhibited a regular band distribution. *Rhynchospora tenuis* (2*n* = 4 and 2*n* = 6) presents karyotypes completely involved in chromosome fusion (symploidy) and fission (agmatoploidy) and the other two, *R. austrobasiliensis* and *R. tenerrima*, showed part of their chromosomes with heteromorphisms.

number reconstruction based on the phylogenetic tree with PastML (**Figure 7**) and ChromEvol (**Supplementary data 12**). Based on the ancestral character reconstruction with PastML (Ishikawa et al., 2019) karyotypes with $n = 5$ or 10 were the most likely ancestral complement, based on maximum likelihood (ML) and maximum parsimony (MP), respectively (**Figure 7**). The ChromEvol model reported an ACN of $n = 5$ and variations were mostly attributed to polyploidization (1.39), chromosome fusion (1.13) and less frequently to fission (0.76) (**Supplementary data 12** shows much more support for $n = 5$), according to the optimal model selected by means of the Akaike information Criterion (AIC) (Akaike, 1974). N = 5 was found as the most likely ACN in most clades with both ChromEvol and PastML ML JOINT+F81 models, except the species rich clade corresponding to sections *Glaucae-Albae-Fuscae-Valderugosae*, for which the ACN was $n = 9$ and $n = 18$ for ChromEvol and ML JOINT+F8, respectively. Results obtained with PastML MP showed $n = 10$ as the most likely ACN in most clades, again only the clade *Glaucae-Albae-Fuscae-Valderugosae* did not converge to $n = 10$.

DISCUSSION

Karyotype Differentiation Versus DNA Content Variation

Genomes size analyses allow us to understand the DNA gain and loss influence among related species, and can explain some aspects of evolutionary differentiation among taxa (Bennett

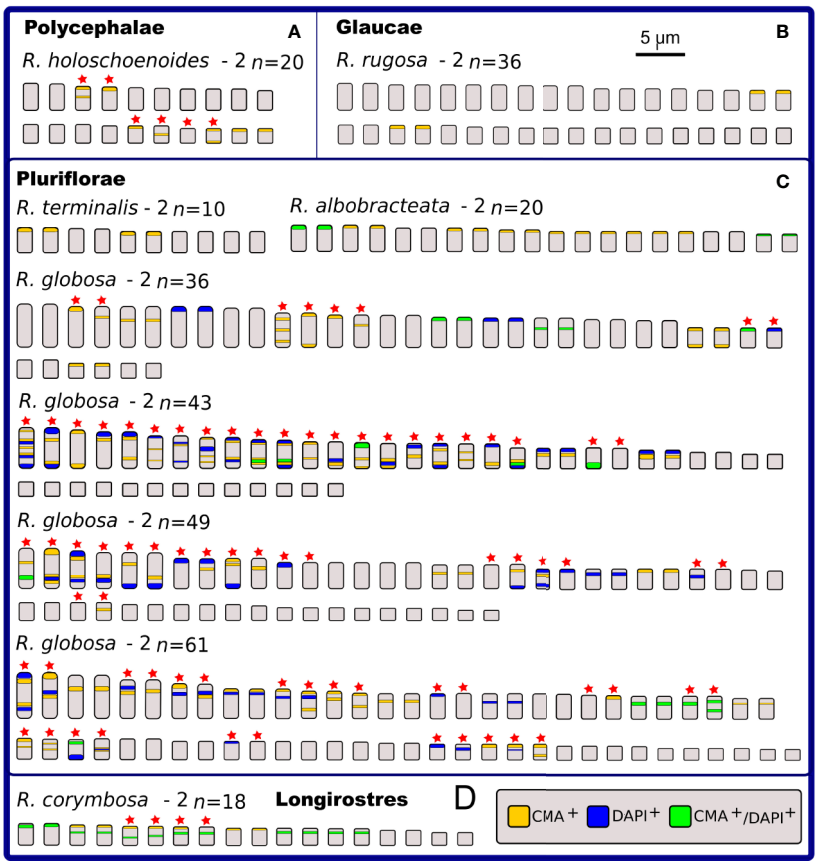

FIGURE 5 | Idiograms representing the physical mapping of C-CMA/DAPI banding in karyotypes of nine species of *Rhynchospora* from sections *Polycephalae* **(A),** *Glaucae* **(B),** *Pluriflorae* **(C),** and *Longirostres* **(D),** according to Kükenthal's classification. Except for *R. rugosa* **(B),** *R. terminalis* and *R. albobracteata* **(C),** the remaining species exhibit polymorphisms in the banding location (indicated by red stars). The most striking situation occurs in the populations of *R. globosa*, which vary in chromosomal number (polyploidy and dysploidy) and in the banding profiles, i.e. number of terminal and interstitial bands, in addition to large differences in the occurrence and number of CMA^+, $DAPI^+$, and $CMA^+/DAPI^+$.

and Leitch, 2005; Leitch et al., 2010). The study of DNA content in holocentrics can lead to interesting results, as genome sizes can be maintained, or not, in clades with regular or rearranged karyotypes (Roalson, 2008), such as observed in *R. tenuis* with $2n = 4$ (0.80 pg) and $2n = 6$ (0.83 pg). DNA amount varied by approximately 22× in 2C-values and 27× in hypothetical 1C*x*-values. However, a comparison between three sections showed that the predominantly hexaploid sect. *Glaucae* ($2n = 36$ and 0.51–0.56 pg) proved to be very stable in relation to sect. *Dichromena* (diploid and polyploid) ($2n = 10$, 20 or 30, and 0.78–3.3 pg) and sect. *Tenues* (diploid, dysploid, and polyploid). The latter two sections have two species (*R. pubera* and *R. austrobrasiliensis*, respectively) that stood out for their substantial genomic DNA accumulation. The great diversity in genome sizes becomes clearer when the putative diploid species *R. terminalis* and *R. ridleyi*, and the assumed polyploids *R. pilosa* and *R. globosa*, are contrasted. A noteworthy case was the ~17% contrast in genome size between assumed tetraploid populations of *R. holoschoenoides*, with 1.25 pg (Ribeiro et al., 2018) and 1.06–1.10 pg. This may be an indication of extensive genome differentiation between populations sampled over 3,000 kilometers apart. Fluctuations in DNA content in Cyperaceae have already been linked to the activity of the transposable elements (Bureš and Zedek, 2014; Souza et al., 2018). In addition, there is evidence of differential accumulation of 35S rDNA sites and heterochromatic bands in *Rhynchospora* and *Eleocharis* (Vanzela et al., 1998; Vanzela and Guerra, 2000; Da Silva et al., 2008; Da Silva et al., 2010). This allows us to suggest that variations in the repetitive DNA fraction can contribute to karyotype differentiation in *Rhynchospora*.

Regarding karyotype organization, numerical diversification in sect. *Tenues* seems to have happened *via* a set of events, starting from $2n = 10$ ($n = 5$), reducing to $2n = 4$ ($n = 2$; descending dysploidy), and achiasmatic meiosis in *R. tenuis* (Vanzela et al., 2003; Cabral et al., 2014), followed by ascending dysploidy to $2n = 5$, 6, and a possible polyploid with $2n = 8$, also in *R. tenuis*. Besides, potential polyploidy was responsible for $2n = 18$ found in two other species in the section (see Vanzela et al., 1996; Vanzela et al., 2000; Arguelho et al., 2012; Michelan et al., 2012). Except for the potential

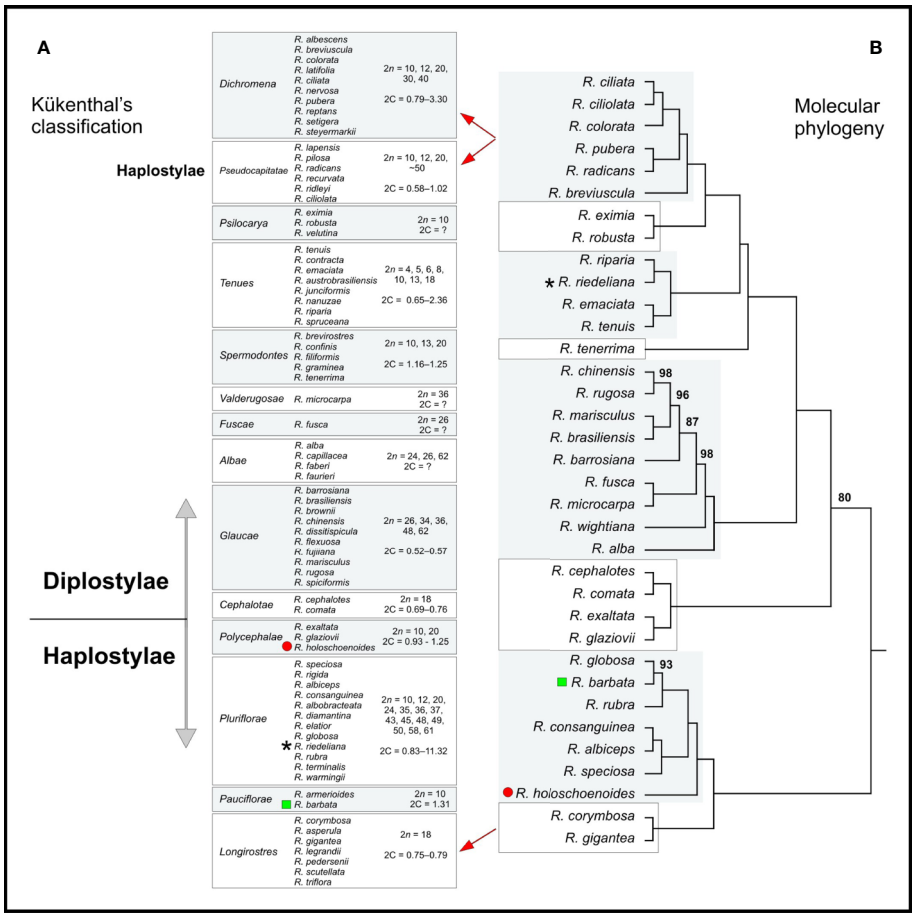

FIGURE 6 | (A) Kükenthal's classification is contrasted with **(B)** the maximum likelihood phylogenetic inference for Rhynchospora based on 54 chloroplast coding sequences (CDS), Length: 51,383 bases. The red circle, the black star and the green square indicate species whose phylogenetic position differed from Kükenthal's classification. Numbers above the branches indicate bootstrap values. Branches without numbers have bootstrap values of 100. The range of 2*n* values for each clade is shown from highest to lowest. Taxa and 2*n* values contain more taxa than are shown in the phylogeny—the clades are considered to be representative for the taxonomic sections.

polyploid cases, we do not see such a sharp contrast in DNA content between diploid and dysploid species of this section. Intraspecific analysis of different populations of *R. globosa* with different chromosome numbers (2*n* = 43, 49, and 61) showed that chromosome number increases are not associated with proportionate genome size increases. Unlike what happened in sect. *Tenues*, polyploidy, dysploidy, and variations in the repetitive fraction of DNA may be acting together in *R. globosa* 2C DNA content variation, and this can be seen in the hypothetical 1C*x*-values. When we compare the DNA amount variation (22–27×) with other Cyperaceae, *Rhynchospora* spp. are more variable than *Carex*, which exhibits a 7× difference among species (Lipnerová et al., 2013), and closer to what happens in *Eleocharis* which exhibited a 21.43× variation (Zedek et al., 2010; Souza et al., 2018). Dysploidy is common in *Rhynchospora* as well as in other Cyperaceae where there is a predisposition for cytotypes and chromosome races in its various clades, including odd numbers due to chromosome fission and fusion, associated or not with polyploidization (Mola and Papeschi, 2006; Roalson, 2008). This could explain the occurrence of different basic numbers (*x* = 5, 6, and 9) in *Rhynchospora* (Luceño et al., 1998a; Vanzela et al., 2000; Ribeiro et al., 2018), as well as in other genera of this family (see Roalson, 2008).

DNA Content Diversity Versus CMA/DAPI Band Gain and Loss

Results focusing on the C-CMA/DAPI bands allowed us to access the dynamics of holocentric genomes from another point of view. Our work with *Rhynchospora* provides one more line of evidence that heterochromatic band accumulation and elimination have an important role in genome differentiation and DNA content fluctuations in plants generally, and specifically in Cyperaceae (Kellogg and Bennetzen, 2004; Grover and Wendel, 2010; Bureš and Zedek, 2014). The comparison among C-CMA/DAPI band profiles suggests an association between band accumulation and increase of DNA 2C-values in some species, but not in others. This is evident in *R. breviuscula*, *R. ciliata*, and *R. pubera* (see also Vanzela and Guerra, 2000; Ribeiro et al., 2018), since *R. ciliata* exhibits many C-CMA/DAPI bands scattered throughout its

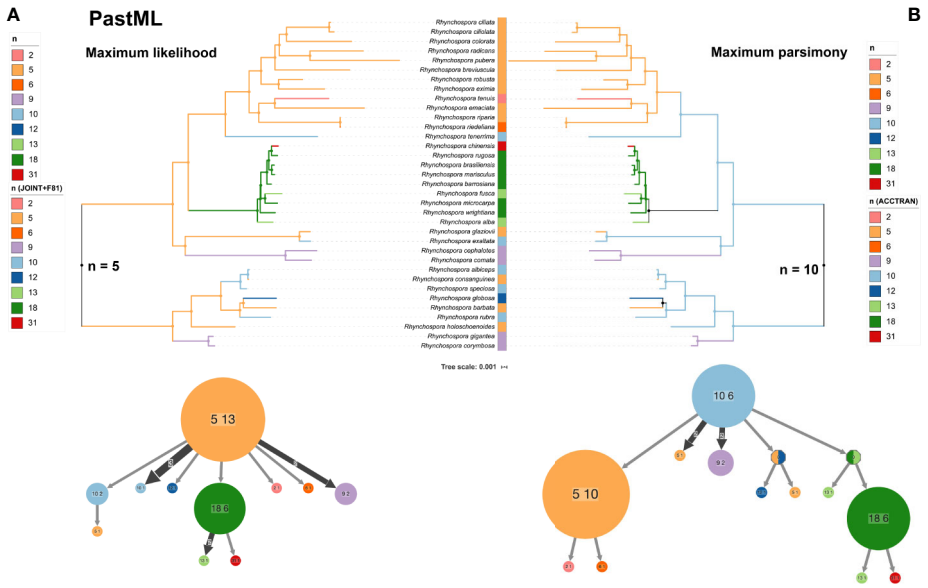

FIGURE 7 | Ancestral chromosome number reconstruction with PastML along the same phylogeny shown in with the two equally best prediction methods. **(A)** Left panel shows the maximum likelihood JOINT+F81 and **(B)** the right panel shows the maximum parsimony accelerated transformation prediction methods, respectively. Bottom graphs show the PastML compressed nodes view for each ancestral character reconstruction method.

chromosomes and unlike *R. pubera* which has threefold larger genome and few bands. Although these suggest that accumulation of repetitive sequences is not detectable by banding techniques in *R. pubera*, the data obtained here do not provide much insight about the nature of these sequences. It is known that not every family of satellite DNA is detected by banding. A good example of a region unlikely to be detected is the centromeric satDNA Tyba that specifically associates with CENH3 (the centromeric histone H3 variant) within the groove along the holocentric chromatids of *R. pubera* (Marques et al., 2015). Only non-centromeric satellite sequences were identified as heterochromatic blocks, as the ones found in *R. globosa* and *R. ciliata* (Ribeiro et al., 2017). When we compared karyotypes in the same taxonomic section, but with large genome size variation (*R. tenuis* and *R. austrobrasiliensis*), or polyploid species from distant sections (*R. nervosa*, *R. tenerrima*, *R. holoschoenoides* and *R. albobracteata*), it is possible to note that there is not always CMA/DAPI band amplification when chromosome number increase.

Polymorphisms were found after C-CMA/DAPI banding in nine out of the 15 species analyzed. Indeed, this peculiarity was observed before in *R. ciliata* and *R. tenuis* (Ribeiro et al., 2017). In several cases, homologous chromosome pairs could not be identified as they did not present identical banding patterns, even when there is normal bivalent formation at meiosis, such as *R. ciliata* (Luceño et al., 1998a), *R. breviuscula*, and *R. nervosa* (Arguelho et al., 2012). These observations suggest that a reduced level of meiotic recombination or unequal recombination could favor accumulation of heterozygosity between homologs. In addition, the apparent regular axis formation observed in the meiosis of some *Rhynchospora* species (Cabral et al., 2014), could be enough to enable regular pairing within diploid and polyploid species, even in those with differences in sequence collinearity. However, this does not seem to apply to our samples of *R. globosa* that accumulated many band polymorphisms, associated with intense numerical rearrangements, and in which there is evidence of irregular meiosis (Luceño et al., 1998a; Arguelho et al., 2012).

Insights on Genome Evolution of Holocentric Species of *Rhynchospora*

The phylogenetic analysis using a large fraction of chloroplast genomes, combined with chromosome and genome size data, showed a wide diversity among and within sections of *Rhynchospora*, similar to the findings of Ribeiro et al. (2018). Species were placed in equivalent clades, except for *R. cephalotes* and *R. exaltata*, which were grouped together in a single clade. The ChromEvol analysis indicated polyploidization as the major driver of genome evolution in *Rhynchospora*, followed by dysploidy. Ancestral chromosome number reconstruction gave either $x = 5$ or $x = 10$ as potential ACN in contrast to the primary number $x = 5$, reported by classic cytogenetic studies (Luceño et al., 1998a; Vanzela et al., 2000; Arguelho et al., 2012; Michelan et al., 2012). If $x = 5$ is indeed the ACN for *Rhynchospora* (as suggested by the Bayesian ChromEvol method), then polyploidy could be more relevant than fission/fusion events, although we could not discard other mechanisms. If $x = 10$ is the ACN, dysploidy would be the most active mechanism. However, the low number of populations analyzed until now could actually "mask" the intraspecific variability and the importance of dysploidy in *Rhynchospora* evolutionary history. Some chromosome pairing reports in species with $2n = 18$ and $2n = 20$ always showed bivalents (Luceño et al., 1998a;

Luceño et al., 1998b; Vanzela et al., 2000). Because 10 appeared as a secondary ACN in our maximum parsimony analysis, this could suggest a diploid meiotic behavior or then point toward a possible paleopolyploidy. In this case, we cannot exclude $n = 10$ as a candidate for ACN in *Rhynchospora*.

Chromosome numbers and genome sizes were negatively correlated overall, though there was a lack of consistence between closely related species. This could be explained by changes associated with differential repetitive DNA accumulation associated with numerical rearrangements. Although transposable elements were not evaluated in this study, changes in genome size mediated by proliferation of retrotransposons were reported as driving genomic changes in sedges (Zedek et al., 2010; Lipnerová et al., 2013; Souza et al., 2018; Johnen et al., 2020), and this could be the focus of future studies in *Rhynchospora*. This idea is supported by the fact that other non-heterochromatic repetitive sequences played a role in differentiating some genomes, such as *R. ciliata* and *R. pubera* (Marques et al., 2015; Ribeiro et al., 2017), besides it appears that C-CMA/DAPI heterochromatin changes independently of phylogenetic relationships. Perhaps the best indication of this diversity are differences between the large genome of *R. pubera* with $2n = 10$ and 320 Mbp of average chromosome size and species of section *Glaucae*, with $2n = 36$ and 14 Mbp of chromosome size and *R. pilosa* with $2n = 50$ and 11 Mbp of chromosome size. An exception to this rule is *R. globosa*, in which many large chromosomes are found, likely representing a true polyploid.

Indeed, such scenarios also seem to happen in other holocentric organisms, and it could be, at least in part, explained by the holokinetic drive model (Bureš and Zedek, 2014). The proposed model works similarly to the centromere drive model of Henikoff et al. (2001), but instead of facilitating evolution of centromere size (number and symmetry), it would facilitate changes in chromosome size and number. Under the holokinetic drive model there are two competing tendencies, 1) fission and loss of repetitive elements and 2) fusion and accumulation of repetitive elements. Similar to the centromeric drive, the holokinetic drive would also depend on i) meiotic asymmetry and ii) the asymmetry of the egg and polar body poles (Bureš and Zedek, 2014). *Rhynchospora* and other Cyperaceae show meiotic asymmetry in both female and male meiosis (Rocha et al., 2016), which could potentially amplify the effectiveness of holokinetic drive in the family. Since *Rhynchospora* genomes are composed of repeat-rich holocentromeres, each duplication/unequal crossing-over could potentially generate chromosomes that accumulate (or lose) more centromere units, facilitated by asymmetric meiosis. In the future, comparative genomic analysis will hopefully unveil the mechanisms for genome evolution in holocentric plants.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

AV: directed researches. PB and MG: performed conventional cytogenetic, C-CMA/DAPI, and flow cytometry analyses. PB, CB, and AM performed bioinformatic and phylogenetic analyses. AV and MS: helped with conventional, C-CMA/DAPI, and idiogram analyses. AV, PB, AM, CB, and MG: wrote the manuscript.

ACKNOWLEDGMENTS

The authors thank the Brazilian agencies FINEP, Fundação Araucária, CNPq (number 309902/2018-5), CAPES, and ProPPG-UEL for financial support. AM is supported by the Max Planck Society. The authors declare they have no conflicts of interest in this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.536507/full#supplementary-material

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE T. Automat. Contr.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Araújo, A. C., Longhi-Wagner, H. M., and Tomas, W. M. (2012). A synopsis of *Rhynchospora* sect. *Pluriflorae* (Cyperaceae). *Brittonia* 64, 381–393. doi: 10.1007/s12228-012-9252-2

Arguelho, E. G., Michelan, V. S., Nogueira, F. M., Silva, C. R. M., Rodriguez, C., Trevisan, R., et al. (2012). New chromosome counts in Brazilian species of *Rhynchospora* (Cyperaceae). *Caryologia* 65, 140–146. doi: 10.1080/00087114.2012.711675

Baez, M., Kuo, Y. T., Dias, Y., Souza, T., Boudichevskaia, A., Fuchs, J., et al. (2020). Analysis of the small chromosomal Prionium serratum (Cyperid) demonstrates the importance of a reliable method to differentiate between mono- and holocentricity. *bioRxiv*. doi: 10.1101/2020.07.08.193714

Bennett, M. D., and Leitch, I. J. (2005). Plant genome research: a field in focus. *Ann. Bot.* 951, 5. doi: 10.1093/aob/mci001

Buddenhagen, C., Lemmon, A. R., Lemmon, E. M., Bruhl, J., Cappa, J., Clement, W. L., et al. (2016). Anchored phylogenomics of Angiosperms I: assessing the robustness of phylogenetic estimates. *bioRxiv* 086298 [Preprint]. doi: 10.1101/086298

Buddenhagen, C. E., Thomas, W. W., and Mast, A. R. (2017). A first look at diversification of beaksedges (Tribe *Rhynchosporeae*; Cyperaceae) in habitat, pollination, and photosynthetic features. *Memoirs New Y. Bot. Garden* 118, 111–124. doi: 10.21135/893275341.002

Buddenhagen, C. E. (2016). *A view of Rhynchosporeae (Cyperaceae) diversification before and after the application of anchored phylogenomics across the angiosperms* (Tallahassee: Florida State University).

Bureš, P., and Zedek, F. (2014). Holokinetic drive: centromere drive in chromosomes without centromeres. *Evolution* 68, 2412–2420. doi: 10.1111/evo.12437

Cabral, G., Marques, A., Schubert, V., Pedrosa-Harand, A., and Schlögelhofer, P. (2014). Chiasmatic and achiasmatic inverted meiosis of plants with holocentric chromosomes. *Nat. Commun.* 5, 5070. doi: 10.1038/ncomms6070

Carrano, A. V., and Heddle, J. A. (1973). The fate of chromosome aberrations. *J. Theor. Biol.* 38, 289–304. doi: 10.1016/0022-5193(73)90176-8

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Da Silva, C. R. M., González-Elizondo, M. S., and Vanzela, A. L. L. (2005). Reduction of chromosome number in *Eleocharis subarticulata* (Cyperaceae) by multiple translocations. *Bot. J. Linn. Soc.* 149, 457–464. doi: 10.1111/j.1095-8339.2005.00449.x

Da Silva, C. R. M., González-Elizondo, M. S., Rego, L. N. A. A., Torezan, J. M. D., and Vanzela, A. L. L. (2008). Cytogenetical and cytotaxonomical analysis of some Brazilian species of *Eleocharis* (Cyperaceae). *Aust. J. Bot.* 56, 82–90. doi: 10.1071/BT07017

Da Silva, C. R. M., Trevisan, R., González-Elizondo, M. S., Ferreira, J. M., and Vanzela, A. L. L. (2010). Karyotypic diversification and its contribution to the taxonomy of *Eleocharis* (Cyperaceae) from Brazil. *Aust. J. Bot.* 58, 49–60. doi: 10.1071/BT09185

Doležel, J., and Göhde, W. (1995). Sex determination in dioecious plants *Melandrium album* and *M. rubrum* using high-resolution flow cytometry. *Cytometry* 19, 103–106. doi: 10.1002/cyto.990190203

Doležel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysák, M. A., Nardi, L., et al. (1998). Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Ann. Bot.* 82 (Suppl. A), 17–26. doi: 10.1093/oxfordjournals.aob.a010312

Doležel, J., Sgorbati, S., and Lucretti, S. (1992). Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plantarum* 85, 625–631. doi: 10.1111/j.1399-3054.1992.tb04764.x

Doležel, J., Bartoš, J., Voglmayr , H., and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry Part A.* 51, 127–128. doi: 10.1002/cyto.a.10013

Doyle, J. J., Flagel, L. E., Paterson, A. H., Rapp, R. A., Soltis, D. E., Soltis, P. S., et al. (2008). Evolutionary genetics of genome merger and doubling in plants. *Annu. Rev. Genet.* 42, 443–461. doi: 10.1146/annurev.genet.42.110807.091524

Escudero, M., Maguilla, E., Loureiro, J., Castro, M., Castro, S., and Luceño, M. (2015). Genome size stability despite high chromosome number variation in *Carex* gr. *Laevigata Am. J. Bot.* 102, 233–238. doi: 10.3732/ajb.1400433

Freyman, W. A., and Höhna, S. (2018). Cladogenetic and Anagenetic Models of Chromosome Number Evolution: A Bayesian Model Averaging Approach. *Systemat. Biol.* 67, 195–215. doi: 10.1093/sysbio/syx065

Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, J. M., Sharma, D. P., and Firoozabady, E. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissue. *Science* 220, 1049– 1051. doi: 10.1126/science.220.4601.1049

Glick, L., and Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* 31, 1914–1922. doi: 10.1093/molbev/msu122

Greilhuber, J. (1995). "Chromosomes of the monocotyledons (general aspects)," in *Monocotyledons: systematics and evolution*. Eds. P. J. Randall, P. J. Cribb, D. F. Cutler and C. J. Humphries (Kew, U.K: Royal Botanic Gardens), 379–414.

Grover, C. E., and Wendel, J. F. (2010). Recent insights into mechanisms of genome size change in plants. *J. Bot.* doi: 10.1155/2010/382732. article ID 382732, 8 p.

Guerra, M. (2000). *Chromosome number variation and evolution in monocots. Monocots: systematics and evolution* (Melbourne: CSIRO), 127–136.

Guerra, M., Ribeiro, T., and Félix, L. (2019). Monocentric chromosomes in *Juncus* (Juncaceae) and ita implications for the chromosome evolution of the family. *Bot. J. Linn. Soc.* 191, 475–483. doi: 10.1093/botlinnean/boz065

Heckmann, S., and Houben, A. (2013). "Holokinetic centromeres," in *Plant centromere Biology*, *1st ed*. Eds. J. Jiang and J. A. Bichler (New York: John Willey and Sons), pp 83–pp 94.

Henikoff, S., Ahmad, K., and Malik, H. S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293, 1098–1102. doi: 10.1126/science.1062939

Hipp, A. L., Rothrock, P. E., and Roalson, E. H. (2009). The evolution of chromosome arrangements in *Carex* (Cyperaceae). *Bot. Rev.* 75, 96–109. doi: 10.1007/s12229-008-9022-8

Hipp, A. L. (2007). Nonuniform processes of chromosome evolution in sedges (*Carex*: Cyperaceae). *Evolution* 61, 2175–2194. doi: 10.1111/j.1558-5646.2007.00183.x

Hoshino, T. (1987). Karyomorphological studies on seven species of Japanese *Rhynchospora* (Cyperaceae). *Kromosomo* 2, 1557–1561.

Ishikawa, S. A., Zhukova, A., Iwasaki, W., and Gascuel, O. (2019). A fast likelihood method to reconstruct and visualize ancestral scenarios. *Mol. Biol. Evol.* 36, 2069–2085. doi: 10.1093/molbev/msz131

Johnen, L., Souza, T., Rocha, D., Parteka, L., Gonzalez-Elizondo, S., Trevizan, R., et al. (2020). Allopolyploidy and genomic differentiation in holocentric species of the *Eleocharis montana* complex (Cyperaceae). *Plant Systemat. Evol.* 306, 39. doi: 10.1007/s00606-020-01666-8

Kellogg, E. A., and Bennetzen, J. L. (2004). The evolution of nuclear genome structure in seed plants. *Am. J. Bot.* 91, 1709–1725. doi: 10.3732/ajb.91.10.1709

Kondo, K., Sheikh, S. A., and Hoshi, Y. (1994). New finding of another $2n = 6$ species in angiosperms, *Drosera roseana* Marchant. *CIS Chromosome Inf. Serv.* 57, 3–4.

Kükenthal, G. (1949). Vorarbeiten zu einer Monographie der *Rhynchosporoideae*. *Bot. Jahrbucher Syst.* 74, 375–509.

Kükenthal, G. (1950a). Vorarbeiten zu einer Monographie der *Rhynchosporoideae*. *Bot. Jahrbucher Syst.* 75, 90–126.

Kükenthal, G. (1950b). Vorarbeiten zu einer Monographie der *Rhynchosporoideae*. *Bot. Jahrbucher Syst.* 75, 127–195.

Kükenthal, G. (1951). Vorarbeiten zu einer Monographie der *Rhynchosporoideae*. *Bot. Jahrbucher Syst.* 75, 273–314.

Leitch, I. J., Beaulieu, J. M., Chase, M. W., Leitch, A. R., and Fay, M. F. (2010). Genome size dynamics and evolution in Monocots. *J. Bot.* doi: 10.1155/2010/862516. article ID 862516. 18 p.

Lipnerová, I., Bureš, P., Horová, L., and Šmarda, P. (2013). Evolution of genome size in *Carex* (Cyperaceae) in relation to chromosome number and genomic base composition. *Ann. Bot.* 111, 79–94. doi: 10.1093/aob/mcs239

Luceño, M., and Guerra, M. (1996). Numerical variations in species exhibiting holocentric chromosomes: a nomenclatural proposal. *Caryologia* 49, 301–309. doi: 10.1080/00087114.1996.10797374

Luceño, M., Vanzela, A. L. L., and Guerra, M. (1998a). Cytotaxonomic studies in Brazilian *Rhynchospora* (Cyperaceae), a genus exhibiting holocentric chromosomes. *Can. J. Bot.* 76, 440–449. doi: 10.1139/b98-013

Luceño, M., Mendes, A. P., Vanzela, A. L. L., and Alves, M. V. (1998b). Agmatoploidy and symploidy in *Rhynchospora cephalotes* (L.) Vahl (Cyperaceae). *Cytologia* 63, 79–81. doi: 10.1508/cytologia.63.79

Luceño, M. (1994). Cytotaxonomic studies in Iberian, Balearic, North African, and Macaronesian species of *Carex* (Cyperaceae): II. *Can. J. Bot.* 72, 587–596. doi: 10.1139/b94-078

Malheiros, N., Castro, D., and Câmara, A. (1947). Cromosomas sem centrómero localizado. O caso de *Luzula purpurea* Link. *Agronom. Lusitana* 9, 51–74.

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., et al. (2015). Holocentromeres in *Rhynchospora are* associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13633–13638. doi: 10.1073/pnas.1521907112

Marques, A., Schubert, V., Houben, A., and Pedrosa-Harand, A. (2016). Restructuring of holocentric centromeres during meiosis in the plant. *Rhynchospora Pubera Genet.* 204, 555–568. doi: 10.1534/genetics.116.191213

Meister, A. (2005). Calculation of binding length of base-specific DNA dyes by comparison of sequence and flow cytometric data. Application to *Oryza sativa* and. *Arabidopsis Thaliana J. Theor. Biol.* 232, 93–97. doi: 10.1016/j.jtbi.2004.07.022

Melters, D. P., Paliulis, L. V., Korf, I. F., and Chan, S. W. L. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations and genomic analysis. *Chromosome Res.* 20, 579–593. doi: 10.1007/s10577-012-9292-1

Michelan, V. S., Trevisan, R., Silva, C. R. M., Souza, R. F., Luceño, M., and Vanzela, A. L. L. (2012). Morphological and genomic characterization of *Rhynchospora tenuis* complex (Cyperaceae) and its taxonomic implications. *Rodriguésia* 63, 775–784. doi: 10.1590/S2175-78602012000400003

Mirarab, S., Nguyen, N., Guo, S., Wang, L. S., Kim, J., and Warnow, T. (2015). PASTA: Ultra-Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences. *J. Comput. Biol.* 22, 377–386. doi: 10.1089/cmb.2014.0156

Mola, L. M., and Papeschi, A. G. (2006). Holokinetic chromosomes at a glance. *J. Basic Appl. Genet.* 17, 17–33.

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Noirot, M., Barre, P., Duperray, C., Hamon, S., and Kochko, A. (2005). Investigation on the causes of stoichiometric error in genome size estimation using heat experiments: consequences on data interpretation. *Ann. Bot.* 95, 111–118. doi: 10.1093/aob/mci006

Otto, F. J. (1990). "DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA," in *Methods in cell biology*, vol. 33 . Eds. Z. Darzynkiewicks and H. A. Crissman (San Diego: Academic), 105–110.

Pazy, B., and Plitmann, U. (1987). Persisting demibivalents: a unique meiotic behaviour in *Cuscuta babylonica* Choisy. *Genome* 29, 63–66. doi: 10.1139/g87-010

Pazy, B., and Plitmann, U. (1994). Holocentric chromosome behaviour in *Cuscuta* (Cuscutoceae). *Plant Systemat. Evol.* 191, 105–109. doi: 10.1007/BF00985345

Rath, S. P., and Patnaik, S. N. (1977). A note on the cytology of *Fimbristylis umbellaris* (Lamk.) Vahl. *Bot. Magazine* 90, 79–81. doi: 10.1007/BF02489471

Ribeiro, T., Marques, A., Novák, P., Schubert, V., Vanzela, A. L. L., Macas, J., et al. (2017). Centromeric and non-centromeric satellite DNA organisation differs in holocentric *Rhynchospora* species. *Chromosoma* 126, 325–335. doi: 10.1007/s00412-016-0616-3

Ribeiro, T., Buddenhagen, C. E., Thomas, W. W., Souza, G., and Pedrosa-Harand, A. (2018). Are holocentrics doomed to change? Limited chromosome number variation in *Rhynchospora* Vahl (Cyperaceae). *Protoplasma* 255, 263–272. doi: 10.1007/s00709-017-1154-4

Roalson, E. H. (2008). A synopsis of chromosome number variation in the Cyperaceae. *Bot. Rev.* 74, 209–393. doi: 10.1007/s12229-008-9011-y

Rocha, D. M., Marques, A., Andrade, C. G., Guyot, R., Chaluvadi, S. R., Pedrosa-Harand, A., et al. (2016). Developmental programmed cell death during asymmetric microsporogenesis in holocentric species of *Rhynchospora* (Cyperaceae). *J. Exp. Bot.* 67, 5391–5401. doi: 10.1093/jxb/erw300

Schubert, I., and Lysak, M. A. (2011). Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* 27, 207–216. doi: 10.1016/j.tig.2011.03.004

Sharma, A. K. (1970). Annual report, 1967–1968. *Res. Bull. Univ. Calcutta (Cytogenetics Lab)* 2, 1–50.

Sousa, A., Barros e Silva, A. E., Cuadrado, A., Loarce, Y., Alves, M. V., and Guerra, M. (2011). Distribution of 5S and 45S rDNA sites in plants with holokinetic chromosomes and the "chromosome field". *Hypothes. Micron.* 42, 625–631. doi: 10.1016/j.micron.2011.03

Souza, T. B., Chaluvadi, S. R., Johnen, L., Marques, A., González-Elizondo, M. S., Bennetzen, J. L., et al. (2018). Analysis of retrotransposon abundance, diversity and distribution in holocentric *Eleocharis* (Cyperaceae) genomes. *Ann. Bot.* 122, 279–290. doi: 10.1093/aob/mcy066

Tejavathi, D. H. (1988). Somatic instability in the populations of *Cyperus cyperoides* (L.) O. Kuntze (Cyperaceae). *Curr. Sci.* 57, 724–728. doi: 10.2307/24091406

Vanzela, A. L. L., and Colaço, W. (2002). Mitotic and meiotic behavior of γ irradiated holocentric chromosomes of *Rhynchospora pubera* (Cyperaceae). *Acta Scienciarum* 24, 611–614.

Vanzela, A. L. L., and Guerra, M. (2000). Heterochromatin differentiation in holocentric chromosomes of *Rhynchospora* (Cyperaceae). *Genet. Mol. Biol.* 23, 453–456. doi: 10.1590/S1415-47572000000200034

Vanzela, A. L. L., Guerra, M., and Luceño, M. (1996). *Rhynchospora tenuis* Link (Cyperaceae), a species with the lowest number of holocentric chromosomes. *Cytobios* 88, 219–228.

Vanzela, A. L. L., Cuadrado, A., Jouve, L., Luceño, M., and Guerra, M. (1998). Multiple locations of the rDNA sites in holocentric chromosomes of *Rhynchospora* (Cyperaceae). *Chromosome Res.* 6, 345–350. doi: 10.1023/A:1009279912631

Vanzela, A. L. L., Luceño, M., and Guerra, M. (2000). Karyotype evolution and cytotaxonomy in Brazilian species of Rhynchospora Vahl (Cyperaceae). Botanical Journal of the Linnean. *Society* 134, 557–566. doi: 10.1006/bojl.2000.0352

Vanzela, A. L. L., Cuadrado, A., and Guerra, M. (2003). Localization of 45S rDNA and telomeric sites on holocentric chromosomes of *Rhynchospora tenuis* Link (Cyperaceae). *Genet. Mol. Biol.* 26:1, 99–201. doi: 10.1590/S1415-47572003000200014

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (New York: Springer-Verlag).

Zedek, F., Šmerda, J., Šmarda, P., and Bureš, P. (2010). Correlated evolution of LTR retrotransposons and genome size in the genus *Eleocharis*. *BMC Plant Biol.* 10:265. doi: 10.1186/1471-2229-10-265

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.

The reviewer TJ declared a past co-authorship with one of the authors, AM.

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