CYTOGENOMICS: STRUCTURAL ORGANIZATION AND EVOLUTION OF GENOMES

EDITED BY: Ricardo Utsunomia, Francisco J. Ruiz-Ruano and Magdalena Vaio
PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-491-4
DOI 10.3389/978-2-88976-491-4

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

CYTOGENOMICS: STRUCTURAL ORGANIZATION AND EVOLUTION OF GENOMES

Topic Editors:
Ricardo Utsunomia, Federal Rural University of Rio de Janeiro, Brazil
Francisco J. Ruiz-Ruano, University of East Anglia, United Kingdom
Magdalena Vaio, Universidad de la República, Uruguay

Citation: Utsunomia, R., Ruiz-Ruano, F. J., Vaio, M., eds. (2022). Cytogenomics: Structural Organization and Evolution of Genomes. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-491-4

Table of Contents

frontiers | Frontiers in Genetics

EDITORIAL
published: 09 June 2022
doi: 10.3389/fgene.2022.938513

Editorial: Cytogenomics: Structural Organization and Evolution of Genomes

Ricardo Utsunomia[1,2]*, *Magdalena Vaio*[3] and *Francisco J. Ruiz-Ruano*[4,5]

[1]Department of Genetics, Institute of Biological Sciences and Health, Federal Rural University of Rio de Janeiro, Seropédica, Brazil, [2]Faculty of Sciences, São Paulo State University, Bauru, Brazil, [3]Department of Plant Biology, Faculty of Agronomy, University of the Republic, Montevideo, Uruguay, [4]Department of Organismal Biology—Systematic Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden, [5]School of Biological Sciences, Norwich Research Park University of East Anglia, Norwich, United Kingdom

Keywords: repetitive (repeated) DNA, satellite DNA, transposable element (TE), sex chromosomes, supernumerary chromosome, karyotype variability

Editorial on the Research Topic

Cytogenomics: Structural Organization and Evolution of Genomes

INTRODUCTION

Cytogenetics is a pioneer field within genetics and emerged when chromosomes were revealed as the gene carriers in the early twentieth century, long before the discovery of the DNA structure. For several decades, cytogenetic studies provided information on the karyotype structure and genome organization of numerous species, revealing a variety of chromosomal polymorphisms in the intra- and interspecies levels.

In the past decade, the power of next-generation sequencing technologies and bioinformatic protocols have become increasingly available to the community of cytogeneticists, allowing the integration of chromosomal and genomic data, even in non-model species of animals and plants, giving rise to the discipline of cytogenomics. The main objective of this Research Topic was to bring together a collection of research and review articles to advance our understanding about karyotype diversification in abroad organism spectrum. In summary, this Topic consists of 11 original papers and one review.

Repeatomes

Much of the variation in DNA content between eukaryotic species is related to the differential accumulation of a heterogeneous collection of repetitive sequences, defined as the "repeatome". Here, Pellicer et al. analyzed congeneric species of *Heloniopsis* (Melanthiaceae) that share the same chromosome number, but differ nearly twofold in genome size. Genome skimming data revealed that the differential amplification of existing and distinct LTR-elements and a single satellite DNA are the main drivers of genome amplification in this case. Following this topic, Marino et al. analyzed five already published Cephalopod genomes, which are known to consist in at least 50% of repetitive sequences, to characterize the catalogue of shared repeats between the species. Data revealed the apomorphic nature of retroelement expansion in octopus, while several DNA transposons were found to be conserved in this lineage.

Although phylogenetic analyses are usually based on single-copy genes, the integration of repetitive DNA sequences data can provide some insights into genome evolution, especially in

Edited and reviewed by:
Aditya Pratap,
Indian Institute of Pulses Research (ICAR), India

***Correspondence:**
Ricardo Utsunomia
ut_ricardo@ufrrj.br
ricardo.utsunomia@unesp.br

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *07 May 2022*
Accepted: *23 May 2022*
Published: *09 June 2022*

Citation:
Utsunomia R, Vaio M and Ruiz-Ruano FJ (2022) Editorial: Cytogenomics: Structural Organization and Evolution of Genomes. Front. Genet. 13:938513. doi: 10.3389/fgene.2022.938513

hybrid taxa. In this context, Oliveira et al. performed a comparative analysis of several diploids and allopolyploids *Stylosanthes* species (Leguminosae) with short-read sequencing data. After assembling and characterizing organelle genomes and repeatomes, assembly-free phylogenetic analyses were performed and allowed the recognition of parental genomes in two allopolyplid species, providing a phylogenetic approach for understanding the genome evolution in this group.

Satellite DNAs

Satellite DNAs (satDNAs) consist of a variety of abundant simple tandemly repeated sequences that are usually located on the peri- and subtelomeric regions. Since its description in the early 60s (Kit 1961), the enzymatic restriction of genomes was the most applied approach to characterize satDNAs. However, this method is not thoroughly efficient, because it is a chance-based approach, since these sequences are highly dynamic and susceptible to quick changes between species, which did not allow a systematic study of multiple satDNAs, even in closely-related species. This scenario changed with the advent and expansion of next generation sequencing technologies and the development of specialized bioinformatic pipelines (Novák et al., 2013), which allowed for a rapid and massive identification of satDNAs from non-model species by taking advantage of low-coverage sequencing.

Conifers are unevenly distributed plant species that exhibit higher abundance on Europe and Asia. Their genomes are largely expanded, mostly exceeding 10 Gb. Here, Heitkam et al. performed a comparative repeat profiling between two conifer species of the *Larix* genus and found that its transposable elements and tandem repeats content were very similar. However, a young satDNA was exclusively found in the European species comparing to the Japanese one, illustrating that the generation of novel repeat families can also play a role in the diversification of conserved conifer genomes.

On the other hand, González et al. analyzed the distribution of four satDNAs in 12 species of *Deschampsia* (Poaceae) and showed that, despite the number of loci, sequence conservation in the monomers of these satDNAs and their chromosomal distribution are quite maintained. Thus, these results suggest that changes in array size and loci number of satDNA, associated with their karyotype and genome diversification, are more marked in these *Deschampsia* species, rather than changes at sequence level.

Since the centromeres are one of the most important structures in cell biology, unveiling their nucleotide composition is a key finding. In this context, Valeri et al. discovered the first centromeric satDNA in two Sirenia species. This satDNA is 684 bp-long and originated after the divergence of Sirenia from Proboscidea and Hyracoidea. Interestingly, no species-specific polymorphisms were found, when comparing several Sirenia species, which is not in accordance with the predictions of concerted evolution and could possibly be related with a centromeric function

Satellite DNAs and Sex Chromosomes

As largely known, sex chromosomes evolved independently multiple times in the history of life. Once established, a common pattern is observed for these elements and usually includes an intense accumulation of repetitive DNA and heterochromatin, due to its non-recombining nature (Charlesworth et al., 2005). In this context, searching for which repeated DNAs accumulated and where, in the chromosomes, can help us to track the origin and evolution of sex chromosomes in different clades.

Lepidoptera is one of the most diverse groups in nature with the vast majority of species exhibiting an ancestral ZZ/ZW sex chromosome system. Here, Cabral-de-Mello et al. characterized the catalogues of satDNAs in three species within the Crambidae moths and performed comparative analyses between males and females' genomic libraries for each species. As a whole, they showed a low abundance of satDNAs in this group, but highly differentiated, which was also reflected on the analyzed W chromosomes, that are each following their own evolutionary path.

Contrary to moths, the fish species *Megaleporinus elongatus* exhibit one of the biggest catalogs of satDNAs to date, with 140 different families but comprising only 5% of the genome. Using an integrated approach, Crepaldi et al. tracked the chromosomal clustering of some satDNAs in the sex chromosomes of *M. elongatus* and *M. macrocephalus* and found relevant differences between these species originated recently and that this genome fraction is strongly related to the female sex chromosome differentiation.

B and Germline-Restricted Chromosomes

Supernumerary, or B, chromosomes are dispensable elements that can perpetuate in natural populations in a parasitic way. Early observations of these elements occurred in the 40s (Östergren 1945) and they were considered inert elements for a long time. However, recent findings, suggest that sometimes B chromosomes can play a significant role by being coopted for essential functions, like sex determination, pathogenicity and others. In this context, Pokorná and Reifová reviewed all such cases of cellular domestication of B chromosomes and showed that, supernumerary elements can be important players with a significant evolutionary impact.

Germline-restricted chromosomes (GRCs) are present in all songbirds studied to date and constitute interesting genomic elements, as they absent in the somatic cells. Torgasheva et al. analyzed the behavior of GRC in male and female meiosis of the great tit and found that GRC was ejected from most male germ cells, corroborating the idea of exclusively maternal inheritance. In addition, chromosome painting analyses revealed that GRCs differ substantially in their genetic content, despite similarities in its behavior during meiosis.

Molecular Cytogenetics and Karyotype Diversification Patterns

Comparative molecular cytogenetics is still a powerful tool to detect major chromosomal rearrangements. Poignet et al.

compared the karyotypes of two passerine species from the genus *Luscinia* by using multiple cytogenetic approaches. Results obtained indicated that diploid chromosome numbers are conserved, as well as main karyotype features, including the presence of similar GRCs. However, comparative genomic hybridization experiments revealed that centromeric repeats in most chromosomes have already diverged, which could theoretically cause meiotic drive or reduced fertility in interspecific hybrids. Following this line, Moraes et al. analyzed five miniature *Pyrrhulina* fishes, which exhibited variation in diploid numbers, as well as differential distribution of repetitive DNAs, suggesting that karyotype diversification in this group has been driven by major structural rearrangements.

CONCLUSION AND PERSPECTIVES

This Research Topic presented studies using a wide variety of approaches and covered several topics in the cytogenetics field. As a whole, this collection demonstrates that the integration of genomic and chromosomal data, and soon, other layers of information, will accelerate our understanding about various aspects of genome evolution.

REFERENCES

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

Kit, S. (1961). Equilibrium Sedimentation in Density Gradients of DNA Preparations from Animal Tissues. *J. Mol. Biol.* 3, 711–IN2. doi:10.1016/s0022-2836(61)80075-2

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

Östergren, G. (1945). Parasitic Nature of Extra Fragment Chromosomes. *Bot. Not.* 2, 157–163.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

RU prepared the draft editorial. MV and FR-R revised the manuscript. All authors contributed to the article and approved the submitted version

ACKNOWLEDGMENTS

We hope the cytogeneticists community will find this Research Topic to be a useful collection of articles. As editors, we would like to thank all authors for their contribution and we are also grateful to all reviewers for their careful evaluation of the papers. Finally, we acknowledge the Frontiers in Genetics team for supporting this Research Topic.

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 Genetics

ORIGINAL RESEARCH
published: 30 March 2021
doi: 10.3389/fgene.2021.661417

The Role of Satellite DNAs in Genome Architecture and Sex Chromosome Evolution in Crambidae Moths

Diogo C. Cabral-de-Mello[1,2], Magda Zrzavá[2,3], Svatava Kubíčková[4], Pedro Rendón[5] and František Marec[2]**

[1] Departamento de Biologia Geral e Aplicada, Instituto de Biociências/IB, UNESP—Univ Estadual Paulista, Rio Claro, Brazil, [2] Biology Centre, Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czechia, [3] Faculty of Science, University of South Bohemia, České Budějovice, Czechia, [4] Veterinary Research Institute, Brno, Czechia, [5] IAEA-TCLA—Consultant—USDA-APHIS-Moscamed Program Guatemala, Guatemala City, Guatemala

OPEN ACCESS

Edited by:
Ricardo Utsunomia, Federal Rural University of Rio de Janeiro, Brazil

Reviewed by:
André Marques, Max Planck Institute for Plant Breeding Research, Germany
Pablo Mora Ruiz, University of South Bohemia in České Budějovice, Czechia
Duílio M. Z. A. Silva, São Paulo State University, Brazil

***Correspondence:**
Diogo C. Cabral-de-Mello
cabral.mello@unesp.br
František Marec
marec@entu.cas.cz

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: 30 January 2021
Accepted: 04 March 2021
Published: 30 March 2021

Citation:
Cabral-de-Mello DC, Zrzavá M, Kubíčková S, Rendón P and Marec F (2021) The Role of Satellite DNAs in Genome Architecture and Sex Chromosome Evolution in Crambidae Moths. Front. Genet. 12:661417. doi: 10.3389/fgene.2021.661417

Tandem repeats are important parts of eukaryotic genomes being crucial e.g., for centromere and telomere function and chromatin modulation. In Lepidoptera, knowledge of tandem repeats is very limited despite the growing number of sequenced genomes. Here we introduce seven new satellite DNAs (satDNAs), which more than doubles the number of currently known lepidopteran satDNAs. The satDNAs were identified in genomes of three species of Crambidae moths, namely *Ostrinia nubilalis*, *Cydalima perspectalis*, and *Diatraea postlineella*, using graph-based computational pipeline RepeatExplorer. These repeats varied in their abundance and showed high variability within and between species, although some degree of conservation was noted. The satDNAs showed a scattered distribution, often on both autosomes and sex chromosomes, with the exception of both satellites in *D. postlineella*, in which the satDNAs were located at a single autosomal locus. Three satDNAs were abundant on the W chromosomes of *O. nubilalis* and *C. perspectalis*, thus contributing to their differentiation from the Z chromosomes. To provide background for the *in situ* localization of the satDNAs, we performed a detailed cytogenetic analysis of the karyotypes of all three species. This comparative analysis revealed differences in chromosome number, number and location of rDNA clusters, and molecular differentiation of sex chromosomes.

Keywords: Lepidoptera, repetitive DNAs, holocentric chromosomes, tandem repeat, W chromatin

INTRODUCTION

Sex chromosomes are extremely dynamic components of the genomes with high inter- and intraspecific variability. They evolved *de novo* multiple times from an ordinary pair of autosomes across various Eukaryote taxa (Charlesworth et al., 2005; Wei and Barbash, 2015; Abbott et al., 2017; Furman et al., 2020) *via* acquisition of a master sex-determining locus in one of the two homologs

of an autosomal pair (Bull, 1983; Charlesworth, 1991; Wright et al., 2016; Furman et al., 2020). This event is followed by suppression of recombination, which leads to accumulation of repetitive DNAs and gene pseudogenization. Ultimately, this gradual genetic erosion may result in the disappearance of Y or W chromosomes (Wei and Barbash, 2015; Abbott et al., 2017).

Accumulation of repeats leads to heterochromatinization of Y or W in many species (Wei and Barbash, 2015; Abbott et al., 2017). Repetitive DNAs that accumulate on sex chromosomes often include satellite DNAs (satDNAs). These non-coding sequences occur in the genomes in hundreds to thousands of copies arranged in tandem in a head-to-tail manner. In general, satDNAs are highly dynamic and their nucleotide sequence, copy number, length of monomers, and chromosomal location can change quickly (Garrido-Ramos, 2017). Empirical data obtained from various animal taxa, such as orthopterans (Palacios-Gimenez et al., 2017; Ferretti et al., 2020), mammals (Acosta et al., 2007; Escudeiro et al., 2019), fishes (Crepaldi and Parise-Maltempi, 2020; Serrano-Freitas et al., 2020), reptiles (Giovannotti et al., 2018), and anurans (Gatto et al., 2018), have demonstrated the role of satDNAs in sex chromosome differentiation.

Lepidoptera (moths and butterflies) with about 160,000 described species (Van Nieukerken et al., 2011) is the largest animal group with female heterogamety (Traut et al., 2007). Most lepidopterans harbor a simple WZ sex chromosome system in females, although Z0 or multiple sex chromosomes have been reported in some species (Traut et al., 2007; Šíchová et al., 2015; Hejníčková et al., 2019). Extensive efforts to understand the sex chromosome composition and evolution in Lepidoptera have been made using variable approaches, from classical cytogenetics to Z-linked gene mapping, comparative genomic hybridization (CGH) of male and female genomic DNAs (gDNAs), genomic *in situ* hybridization (GISH) of female gDNA, W-chromosome painting, and sequencing of male and female genomes (Abe et al., 2005; Yoshido et al., 2005, 2020; Traut et al., 2007, 2013; Vítková et al., 2007; Nguyen et al., 2013; Šíchová et al., 2013; Fraïsse et al., 2017; Zrzavá et al., 2018). These studies documented that the Z is a gene-rich, autosome-like chromosome with conserved synteny (Van't Hof et al., 2013; Dalíková et al., 2017a). In contrast, the W chromosome is largely heterochromatic, composed of repetitive sequences and mostly lacking protein-coding genes (Traut et al., 2007; Sahara et al., 2012). The heterochromatin nature of the W chromosomes leads to the formation of sex chromatin, a roundish body in interphase nuclei of females (Smith, 1945; Traut and Marec, 1997).

Due to the prevalence of repeats, attempts to sequence the lepidopteran W chromosome are scarce (Abe et al., 2005; Fuková et al., 2007; Traut et al., 2013). These efforts, however, have shown that the W consists mainly of mobile elements. Other studies that focused directly on satDNA detected a few satellites located on the W chromosome in several species, namely in *Plodia interpunctella* (Dalíková et al., 2017b), *Mamestra brassicae* (Mandrioli et al., 2003), and *Spodoptera frugiperda* (Lu et al., 1994). However, low amounts of satDNA appear to be a general feature of lepidopteran genomes, which are generally rather small, ranging from 0.29 pg in *Danaus plexippus* (Danaidae) to 1.94 pg in *Euchlaena irraria* (Geometridae) (Gregory, 2020, Animal Genome Size Database). So far, only a total of five satDNAs have been identified in all lepidopteran species investigated (Lu et al., 1994; Mandrioli et al., 2003; Mahendran et al., 2006; Věchtová et al., 2016; Dalíková et al., 2017b). They showed variable patterns of chromosomal distribution including W specific satDNA (Dalíková et al., 2017b), satDNA shared exclusively by Z and W chromosomes (Mandrioli et al., 2003), and satDNAs spread on multiple chromosomes (Mahendran et al., 2006; Věchtová et al., 2016).

The small sizes of lepidopteran genomes and their repetitive composition, mainly consisting of transposable elements (TEs), may have hindered the isolation of satDNAs using classical methods, i.e., restriction endonuclease digestion of gDNA, which depends on high abundance of a particular repeat (Camacho et al., 2015). Only recently, this difficulty has been overcome by bioinformatic analysis of low-coverage sequenced genomes using RepeatExplorer (Novák et al., 2013). This approach allowed the characterization of multiple satDNAs and generated valuable chromosomal and genomic information, for example in crickets (Palacios-Gimenez et al., 2017), *Drosophila* species (Silva et al., 2019), fishes (Silva et al., 2017; Crepaldi and Parise-Maltempi, 2020), amphibians (da Silva et al., 2020), and some plants (Mata-Sucre et al., 2020; Zwyrtková et al., 2020). However, this approach has not yet been applied to Lepidoptera.

The lepidopteran family Crambidae has more than 10,000 species described worldwide (Nuss et al., 2003–2020), some of which are serious pest of agricultural crops such as sugarcane, maize, rice, and sorghum, and are economically important (Solis, 1997; Munroe and Solis, 1999; Meissle et al., 2010). Despite the overall significance of crambid moths, very little is known about their genome architecture, because cytogenetics of the Crambidae was poorly explored. In a few crambid species, only chromosome numbers are known (Robinson, 1971), including the reduced $n = 17$ in the sugarcane borer *Diatraea saccharalis* (Virkki, 1963) and the ancestral $n = 31$ in *Ostrinia* representatives (Guthrie et al., 1965; Kageyma and Traut, 2004; Yasukochi et al., 2016). Sex chromosomes have only been studied in *O. scapulalis* (Kageyma and Traut, 2004) and *O. nubilalis* (Yasukochi et al., 2016). In *O. nubilalis*, a more detailed chromosomal analysis was performed with the assignment of the 31 chromosomes by gene-based fluorescence *in situ* hybridization (FISH) mapping (Yasukochi et al., 2016). To expand our knowledge of this group, we analyzed karyotypes and genomes of three representatives of Crambidae, namely the European corn borer (*Ostrinia nubilalis*), one of the most important pests causing economical losses to corn growers (Lynch, 1980), the box tree moth (*Cydalima perspectalis*), an Asian species that has been introduced in Europe, causing defoliation of the ornamental shrub *Buxus* spp. (Nacambo et al., 2013), and the Guatemalan sugarcane borer (*Diatraea postlineella*), a pest which causes significant damage to sugarcane in Guatemala (Solis and Metz, 2016; Solis et al., accepted). Our aim was to investigate the role of tandem repeats in general architecture of genomes and in sex chromosome differentiation in studied species, as little is known about these sequences in Lepidoptera, which are otherwise important components of eukaryotic genomes. To

accomplish this, we used for the first time in Lepidoptera clustering analysis performed by RepeatExplorer on male and female genomic sequences of all three species to identify and map satDNAs. Further, we characterized the chromosomes of the studied species in order to provide a cytogenetic background not only to our FISH experiments but also to existing or upcoming sequencing projects. In particular, we focused on sex chromosomes, knowledge of which is important for potential pest control, such as the sterile insect technique, as *O. nubilalis* is one of the most important pests causing economical losses to corn growers (Lynch, 1980).

MATERIALS AND METHODS

Animals

We studied three species of the family Crambidae, the box tree moth (*C. perspectalis*), the Guatemalan sugarcane borer (*D. postlineella*), and the European corn borer (*O. nubilalis*). A small laboratory colony of *C. perspectalis* was established from the last instar larvae collected from infested shrubs of *Buxus* spp. in Valtice (Czechia) in May 2018. The colony was kept for several generations on fresh leaves of *Buxus* spp. Eggs of *D. postlineella* were obtained from a mass-reared colony at the Santa Ana sugarcane farm, Escuintla, Guatemala City, Guatemala. Larvae of this species were reared on artificial diet prepared according to the supplier's recipe. *O. nubilalis* (Z and E strains) was obtained from a laboratory colony kept at the Department of Entomology, Max Planck Institute of Chemical Ecology, Jena, Germany. For genomic analysis, we used sequenced genomes from both *O. nubilalis* strains, but only Z-strain was used for further chromosomal studies. Larvae were reared on an artificial wheat germ-based diet (Lewis and Lynch, 1969) with some modifications, and adults fed a 10% solution of honey in water. Cultures of all three species were maintained at 20–22°C and 12-h light/12-h dark regime.

Polyploid Nuclei Preparation and Microdissection of the W Chromatin

Polyploid interphase nuclei were prepared from Malpighian tubules of male and female larvae of the last instar (see Marec and Traut, 1994). Malpighian tubules were dissected in physiological solution, fixed in Carnoy's fixative (ethanol, chloroform, acetic acid, 6:3:1) on a slide for about 1 min, stained with 1.25% lactic acetic orcein for 3–5 min, and mounted in the staining solution. The slides were then examined under a light microscope for the presence of W chromatin in females and its absence in males (reviewed in Traut and Marec, 1996).

For laser microdissection of W chromatin bodies, we followed the procedure described in Fuková et al. (2007) with some modifications. Malpighian tubules were dissected from last instar female larvae, swollen for 10 min in a hypotonic solution (75 mM KCl) and fixed in Carnoy's fixative for 15 min. The tubules were then transferred into a drop of 60% acetic acid on a glass slide coated with a polyethylene naphthalate membrane (Goodfellow, Huntingdon, United Kingdom) and torn to pieces using tungsten needles. The cell suspension was spread at 40°C using a heating plate and stained with 5% Giemsa for 7 min. Microdissection of W chromatin bodies was performed using the PALM MicroLaser System (Carl Zeiss MicroImaging GmbH, Munich, Germany) as described in Kubickova et al. (2002).

Chromosome Preparations

Spread chromosome preparations were made as described previously (Mediouni et al., 2004; Šíchová et al., 2013). Mitotic chromosomes were obtained from wing imaginal disks of the last instar larvae of both sexes and meiotic chromosomes in the pachytene stage from testes of the penultimate and last instar larvae and ovaries of the last instar larvae and early pupae. Wing imaginal disks and testes were dissected in a physiological solution, hypotonized for 10 min in 75 mM KCl and fixed in Carnoy's solution for at least 15 min, whereas the ovaries were transferred to Carnoy's solution for 15 min immediately after dissection. The fixed tissues were macerated on a slide in a drop of 60% acetic acid and then spread at 45°C using a heating plate. The slides were inspected under a phase contrast microscope and preparations of sufficient quality were passed through an ethanol series (70, 80, and 100%, 30 s each) and stored at –20°C until use.

Genome Sequencing, Satellite DNA Identification and Analysis

Genomic DNA was extracted from one male larva and one female larva of each species using the CTAB (hexadecyltrimethylammonium bromide) DNA isolation procedure according to Winnepenninckx et al. (1993) with the following modifications. The concentration of EDTA in the extraction buffer was doubled, i.e., 40 mM, the homogenized material was incubated overnight, the centrifugation steps were performed at 14,000 *g* and prior to isopropanol precipitation, the samples were treated with 62.25 μg/mL RNase A for 30 min at 37°C.

Paired-end sequencing (2 × 150 bp) using the Illumina HiSeq 4000 system was performed by Novogene (HK) Co., Ltd., (Hong Kong, China). We checked the quality of reads with FastQC (Andrews, 2010) and processed the reads though Toolkit suit using default options (Gordon and Hannon, 2010) to be used as input for RepeatExplorer (available at[1]) clustering. We made separate identification of satDNAs for each three species. A comparative analysis of satDNAs between the sexes for *D. postlineella* and *C. perspectalis* was performed using 250,000 reads of each sex as input. For *O. nubilalis*, genomes of both sexes from the two strains (E and Z) were used, being 125,000 reads per genome. After clustering by RepeatExplorer (Novák et al., 2010, 2013) using default options, the satDNAs identified as satDNAs by TAREAN tool (Novák et al., 2017) were considered for subsequent analysis. The tandem arrangement was checked by dotplot and Tandem Repeats Finder, TRF (Benson, 1999).

For similarity analysis, we performed all-against-all comparison of monomers of the recovered satDNAs using RepeatMasker (Smit et al., 2013–2015) and the

[1]http://repeatexplorer.org

"rm_homology.py" script[2]. We used RepeatMasker to calculate divergence and abundance of each satDNA in the female and the male genomes at intraspecific level. For this purpose, we randomly selected 7.5 million read pairs per library obtained by seqtk tool[3] and aligned them against dimers of consensus satDNA sequences. To estimate the average Kimura 2-parameter distances (K2P) for each satDNA family, we used the calcDivergenceFromAlign.pl script from the RepeatMasker utility tool. Genomic abundance for each satDNA family was calculated according to the proportion of nucleotides aligned with the reference consensus sequence divided by the library size. Finally, we compared the divergence of satDNAs between sexes generating repeat landscapes showing the relative abundance of repeat elements on the Y-axis and 1% intervals of K2P distance from the consensus on the X-axis. In addition, we also performed the RepeatMasker analysis at interspecific level to check the occurrence of different satDNAs in the species studied and to check their abundance. Similarity of the satDNAs identified with previously characterized sequences was checked by blast search in NCBI and Repbase.

DNA Probes and Fluorescence *in situ* Hybridization

For each studied species, we prepared specific gDNA probes, W-chromosome painting probes, and satDNA probes. We also prepared non-species specific probes to localize conserved sequences including telomeric repeats and 18S rDNA. These probes were used for different FISH experiments following different protocols, as specified below.

Female and male gDNAs extracted by the CTAB procedure were labeled by nick translation with Cy3-dUTP for female gDNA or fluorescein-12-dUTP (both Jena Bioscience, Jena, Germany) for male gDNA, or vice versa. The nick translation mixture was composed of 500 ng of gDNA, 25 μM of each dATP, dCTP, and dGTP, 9 μM of dTTP, 16 μM of labeled nucleotides, nick translation buffer (50 mM Tris–HCl, pH 7.5, 5 mM $MgCl_2$, 0.005% BSA), 10 mM β-mercaptoethanol, 20 U DNA Polymerase I (ThermoFisher Scientific, Waltham, MA, United States), and 0.005 U DNase I (ThermoFisher Scientific). The mixture was incubated for 2 h at 15°C. These labeled gDNA probes were used for CGH following the protocol described in Traut et al. (1999) with modifications proposed by Dalíková et al. (2017a).

W-chromosome painting probes were prepared as described in Fuková et al. (2007). Briefly, samples of microdissected W chromatin were amplified by PCR using a GenomePlex Single Cell Whole Genome Amplification Kit, WGA4 (Sigma-Aldrich, St. Louis, MO, United States). The amplified product was re-amplified and labeled with either Cy3-dUTP or fluorescein-12-dUTP by PCR using a GenomePlex WGA Reamplification Kit, WGA3 (Sigma-Aldrich). These probes were used in W-chromosome painting experiments at intraspecific level and for cross-species W-painting following the CGH protocol (Traut et al., 1999). For intraspecific experiments, we combined the W-painting probe with a $(TTAGG)_n$ telomeric probe obtained by non-template PCR according to Ijdo et al. (1991) and using primers $(TTAGG)_5$ and $(CCTAA)_5$. The telomeric probe was labeled with Cy3-dUTP by using the improved nick translation procedure of Kato et al. (2006) with some modifications (see Dalíková et al., 2017a).

18S rDNA and satDNA probes were amplified by PCR from female gDNA and labeled by PCR or nick translation with biotin-16-dUTP (Roche Diagnostics, Mannheim, Germany). The 18S rDNA was obtained from codling moth (*Cydia pomonella*) gDNA using the primers described by Fuková et al. (2005). This probe was labeled by nick translation. For satDNA amplification we designed primers using primer3 (Untergasser et al., 2012) or manually (**Supplementary Table 1**). PCR was carried out in 20-μl reaction volume containing 1.5 μM of each primer, 1 × Ex *Taq* buffer (TaKaRa), 0.5 U Ex *Taq* polymerase (TaKaRa), 200 μM of each nucleotide and either 50–100 ng gDNA. The thermal cycling profile consisted of initial denaturation at 95°C for 5 min, 30 cycles of denaturation at 94°C for 30 s, annealing at variable temperature (see **Supplementary Table 1**) for 30 s, elongation at 72°C for 1 min and final elongation 72°C for 3 min. After PCR, the fragments of satDNAs were visualized on a 1.5% agarose gel and monomers were extracted from the gel and re-amplified by PCR. The re-amplified products were sequenced by the Sanger method using the service of SEQme (Dobříš, Czechia) to verify their reliability. For satDNAs probe labeling, a second PCR was done with incorporation of biotin-16-dUTP. For FISH with these probes, we followed the protocol described in Cabral-de-Mello and Marec (2021). The probes were detected by Cy3-conjugated streptavidin (Jackson Immuno Res. Labs. Inc., West Grove, PA, United States).

In experiments in which the W-chromosome painting probe and satDNAs were mapped on the same slide, we applied two rounds of FISH. We first performed W-chromosome painting using the CGH protocol (Traut et al., 1999), then FISH with satDNA probe (Cabral-de-Mello and Marec, 2021). In all FISH experiments, chromosomes were counterstained with DAPI (4′,6-diamidino-2-phenylindole, dihydrochloride; Sigma-Aldrich) and slides mounted with antifade based on DABCO (1,4-diazabicyclo(2.2.2)-octane; Sigma-Aldrich).

RESULTS

Karyotypes and Chromosomal Location of 18S rDNA

Diploid chromosome numbers were determined by analysis of mitotic metaphase from wing imaginal disks of females and males stained with DAPI. In *C. perspectalis* and *O. nubilalis* we observed $2n = 62$ chromosomes [**Supplementary Figure 1A**; cf. Yasukochi et al. (2016) for chromosome number in *O. nubilalis*]. However, in *D. postlineella* we found a reduced number of chromosomes to $2n = 42$ (**Supplementary Figure 1B**). We also counted bivalents in pachytene of females or males, confirming

[2]https://github.com/fjruizruano/ngs-protocols

[3]https://github.com/lh3/seqtk

the number of chromosomes in all three species (**Supplementary Figures 2A–C**). In *D. postlineella*, meiotic bivalents often formed clumps at the pachytene stage (**Supplementary Figure 2B**), probably due to their longer length compared to the other two species. None of the studied species showed large heterochromatin blocks.

A single sex chromatin body, deeply stained with orcein and of regular spherical shape, was observed in each polyploid nucleus of Malpighian tubules from females, indicating the presence of the W chromosome in all three species (**Supplementary Figures 3A,C,E**). As expected, sex chromatin was absent in the polyploid nuclei of Malphigian tubules from males (**Supplementary Figures 3B,D,E**). This finding suggests that the three crambid species have a WZ/ZZ sex chromosome system, which was confirmed by further analysis.

Fluorescence *in situ* hybridization with 18S rDNA probe applied to male pachytene nuclei revealed significant differences between the species studied in the number and location of rDNA clusters (**Figures 1a–c**). In *C. perspectalis*, two clusters were found, each located in the terminal position of a different bivalent (**Figure 1a**), whereas in *D. postlineella* only one interstitial cluster was observed (**Figure 1b**). In contrast, four terminal clusters in four different bivalents were found in *O. nubilalis* (**Figure 1c**).

Differentiation of Sex Chromosomes by CGH and W-Painting Probes

To investigate the gross molecular differentiation of W and Z chromosomes, we performed CGH. In all three species, the W chromosome showed stronger binding of probes derived from both female and male gDNAs compared to the other chromosomes. In addition, all species showed more intense labeling of female-derived probes in comparison with male-derived probes on the W chromosome, indicating that this chromosome consists mainly of female-specific or female-enriched sequences (**Figures 2a–l**).

The W-painting probe specifically highlighted the entire W chromosome at the intraspecies level, allowing easy identification of the sex chromosome bivalent and confirming considerable molecular differentiation of the W chromosome from other chromosomes in all three species. The telomeric probe hybridized only to the end of the chromosomes, which also confirmed that the sex chromosome bivalent is composed of only two elements (WZ) (**Figures 3a,d,g**). Using cross-species W-painting, we investigated the molecular differentiation of W chromosomes between the three species. Cross-hybridization of W-painting probes resulted in scattered signals on the W chromosome of another species in all cases, but in varying numbers and intensities, depending on the species (**Figures 3b,c,e,f,h,i**). In general, after FISH with a W-probe from another species, W chromosomes were decorated with several to multiple clusters of hybridization signals along the entire length. The highest intensity of labeling was recorded in the W chromosome of *D. postlineella* using the W-painting probe of *O. nubilalis* (**Figure 3e**). In contrast, the lowest signal intensity in cross-species W-painting experiments was observed using the W-probe from *C. perspectalis* in the W chromosome of *O. nubilalis* (**Figure 3h**). Only weak hybridizations signals were observed on some autosomes (results not shown).

satDNAs Genomic Characterization and Chromosomal Mapping

By clustering analysis, separately for each species, in RepeatExplorer using the TAREAN report, we identified a total of seven putative satDNAs in the three species studied, some of them with high confidence and others with low confidence (a classification provided by TAREAN) (**Supplementary Figure 4**). One satDNA was identified in *C. perspectalis* (Cper-Sat01), two in *D. postlineella* (Dpos-Sat01 and Dpos-Sat02), and four in *O. nubilalis* (Onub-Sat01, Onub-Sat02, Onub-Sat03, and Onub-Sat04). Subsequent analyses were performed for these seven putative satDNAs, demonstrating their tandem arrangement by dotplot and TRF and also by PCR, which showed a typical ladder-like pattern. Monomers of these satDNAs were highly variable in size, ranging from 123 to 2,244 bp. These satDNAs were A + T-rich with G + C content ranging from 36.44 to 49.11%. They represented only a small portion of the male and female genomes in each species, ranging from 0.01909 to 0.15464%, showing similar abundance between the sexes. The divergences for each satDNA family ranged from 1.32 to 11.34% and were similar between the sexes. These data are summarized in **Table 1** and the sequence logos are shown in **Supplementary Figure 4**. SatDNA landscapes, which were generated to compare the repeats between sexes, aiming to check the possible amplification of variants with distinct degrees of divergence, did not reveal any sex-specific differences (**Figure 4**). We also searched for all satDNA families at the interspecific level by RepeatMasker and found that they are present in all three genomes, but with a lower proportion compared to the species in which they were identified (**Supplementary Table 2**). Finally, our Repbase and NCBI searching did not reveal relevant similarity with any described sequences. The sequences were deposited in GenBank under the accession numbers MW369067–MW369073.

Fluorescence *in situ* hybridization mapping of satDNAs revealed a variable distribution pattern with clusters on autosomes and sex chromosomes (**Figures 5–7**). Hybridization signals of Cper-Sat01 were scattered in all chromosomes of *C. perspectalis*, with no apparent enrichment in specific chromosomal regions (**Figure 5a**). However, analysis of interphase nuclei revealed enrichment of this satDNA on W chromatin (**Figure 5b**). At the pachytene stage, in combined analysis with the W-probe, clusters of Cper-Sat01 were enriched and distributed along the entire length of the W chromosome, but were almost absent on the Z chromosome (**Figure 5c**). In *D. postlineella*, two clusters of Dpos-Sat01 were located interstitially in a pair of autosomes carrying the nucleolar organizer region (NOR) formed by the major rDNA (**Figure 6a**). FISH

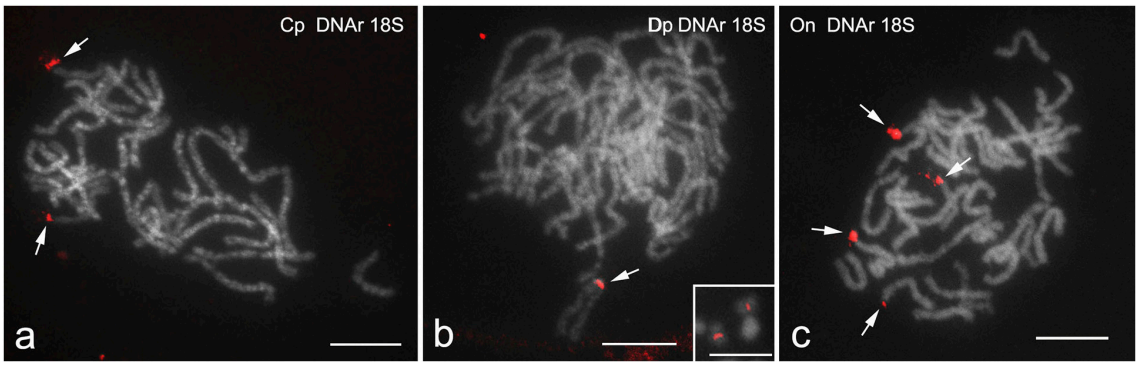

FIGURE 1 | FISH mapping of 18S rDNA clusters in male pachytene bivalents of *Cydalima perspectalis* **(a)**, *Diatraea postlineella* **(b)**, and *Ostrinia nubilalis* **(c)**. Chromosomes were counterstained with DAPI. Arrows indicate hybridization signals of the probes (red). The inset in panel **(b)** shows two homologous mitotic chromosomes from a *D. postlineella* spermatogonium with signals of the 18S rDNA in the interstitial position. Bar = 10 μm.

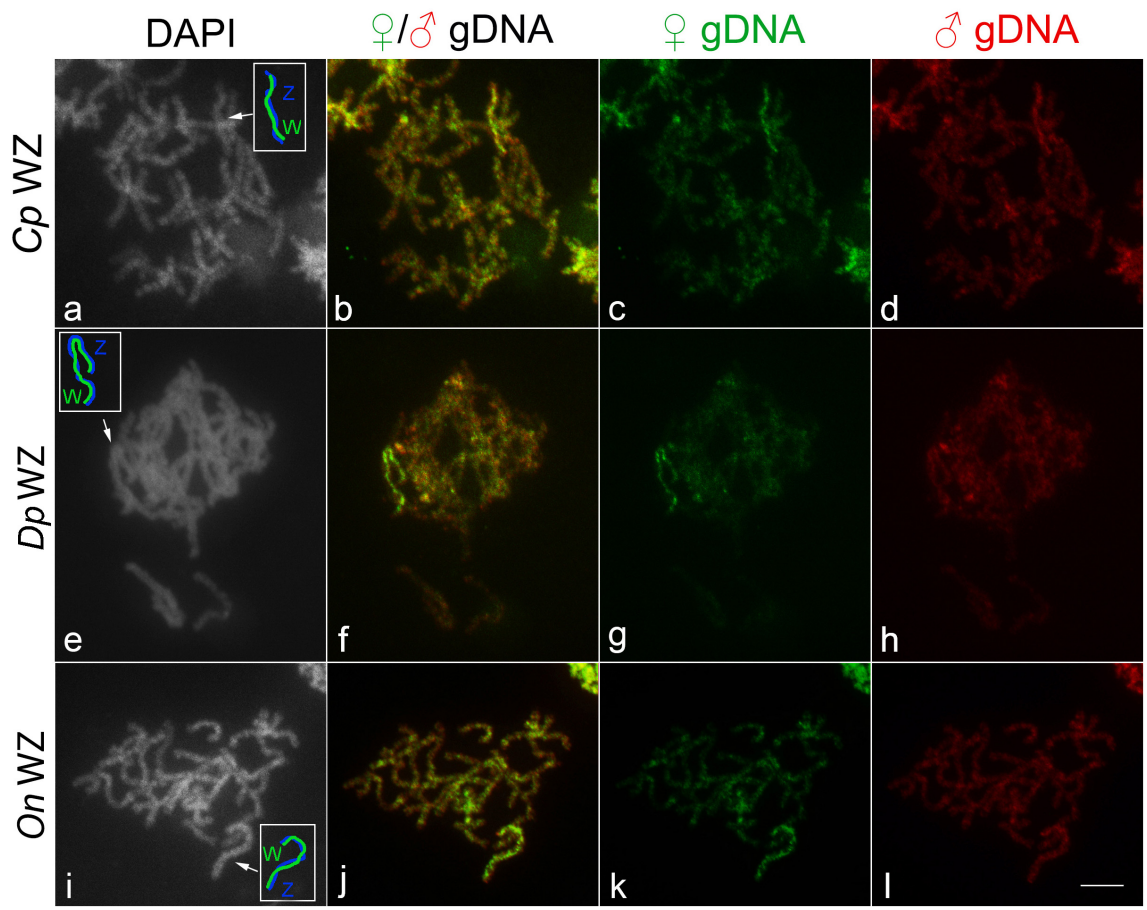

FIGURE 2 | WZ sex chromosome bivalents identified by comparative genomic hybridization (CGH) in female pachytene oocytes of **(a–d)** *Cydalima perspectalis*, **(e–h)** *Diatraea postlineella*, and **(i–l)** *Ostrinia nubilalis*. **(a,e,i)** DAPI staining, **(b,f,j)** merged images of male-derived and female-derived probes, **(c,g,k)** female-derived genomic DNA probe, and **(d,h,l)** male-derived genomic DNA probe. Sex chromosome bivalents are indicated by arrows and schematized in white boxes in panels **(a,e,i)**. Bar = 10 μm.

with Dpos-Sat02 probe identified a single cluster that was located on another autosomal pair in the interstitial position (**Figure 6b**).

Fluorescence *in situ* hybridization mapping of four satDNAs identified in the genome of *O. nubilalis* showed scattered hybridization signals on most chromosomes.

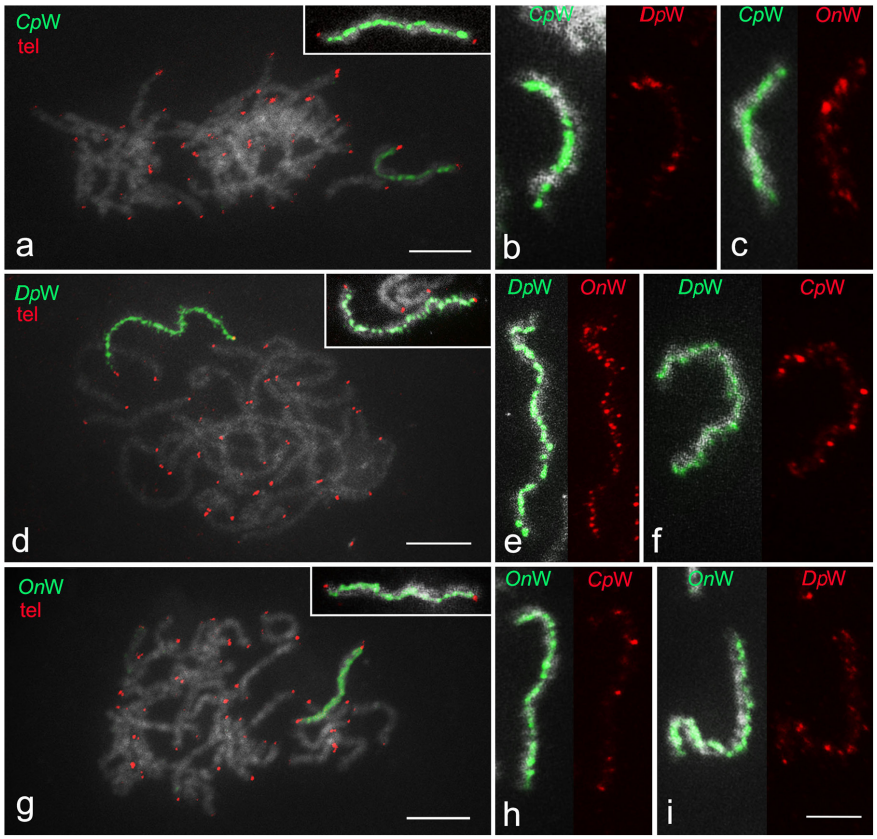

FIGURE 3 | W-chromosome painting in three species of Cambridae. **(a,d,g)** Intraspecies W-chromosome painting (green) and FISH with $(TTAGG)_n$ telomeric probe (red). Pachytene oocyte complements are shown in panel **(a)** *Cydalima perspectalis*, **(d)** *Diatraea postlineella*, and **(g)** *Ostrinia nubilalis*. In panels **(b,c,e,f,h,i)**, intraspecies and cross-species paintings of the W chromosome in the WZ bivalent of **(b,c)** *C. perspectalis*, **(e,f)** *D. postlineella*, and **(h,i)** *O. nubilalis* are compared. The probes are indicated directly in the images. Bar = 10 μm.

TABLE 1 | Main characteristics of SatDNA families identified in Crambidae genomes.

SpeciessatDNA family	Monomer size (bp)	G + Ccontent (%)	Abundance (%)		Divergence (%)	
			Female	Male	Female	Male
Cydalima perspectalis						
Cper-Sat01	2,244	43.36	0.14476	0.12624	7.39	7.14
Diatraea postlineella						
Dpos-Sat01	1,391	36.44	0.04980	0.04895	10.1	11.23
Dpos-Sat02	1,380	40.57	0.01929	0.02431	1.32	1.37
Ostrinia nubilalis						
Onub-Sat01	528	39.77	0.15464	0.11740	9.38	9.55
Onub-Sat02	453	40.39	0.05552	0.07105	11.34	11.31
Onub-Sat03	957	49.11	0.02807	0.02993	2.05	2.08
Onub-Sat04	123	37.39	0.02839	0.02604	8.86	9.28

However, these satDNAs were distributed differently in various chromosomal regions. Onub-Sat02 and Onub-Sat03 were virtually scattered and formed clusters of small dots along the entire length of most chromosomes, while Onub-Sat01 and Onub-Sat04 were enriched primarily in the terminal regions of the chromosomes (**Figures 7a,b**). Concerning the WZ sex chromosomes, Onub-Sat01 (**Figure 7c**), Onub-Sat02 (**Figure 7d**), and Onub-Sat03 (**Figure 7e**) showed a similar distribution in both sex chromosomes. Remarkably, large terminal clusters occurred in Onub-Sat01, in addition to several smaller clusters located mainly on the W chromosome (**Figure 7c**). In contrast, Onub-Sat04 was located exclusively at one termini of the Z chromosome (**Figure 7f**).

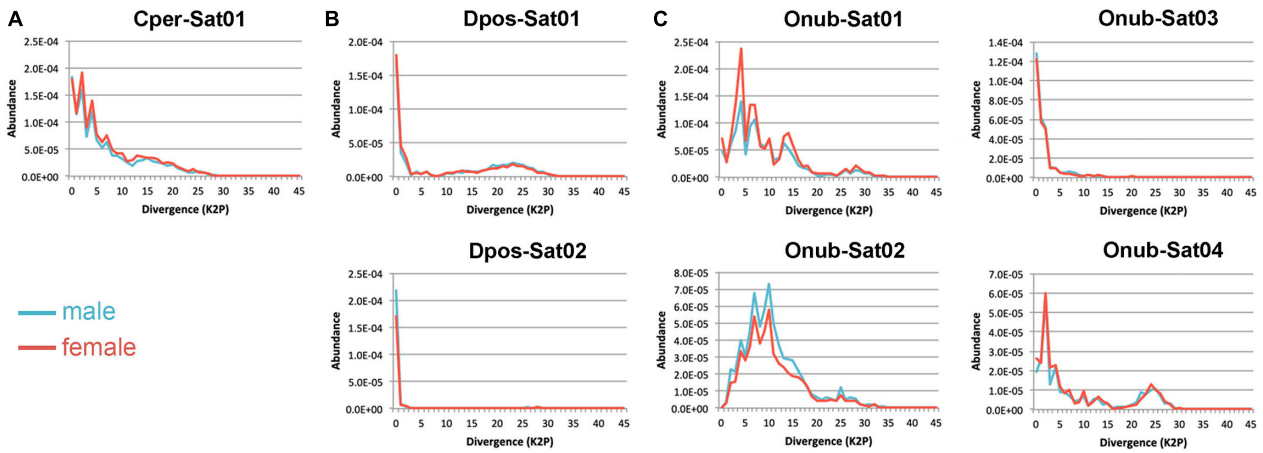

FIGURE 4 | SatDNA landscapes of male (blue) and female (red) repeats for each satDNA identified in **(A)** *Cydalima perspectalis*, **(B)** *Diatraea postlineella*, and **(C)** *Ostrinia nubilalis*.

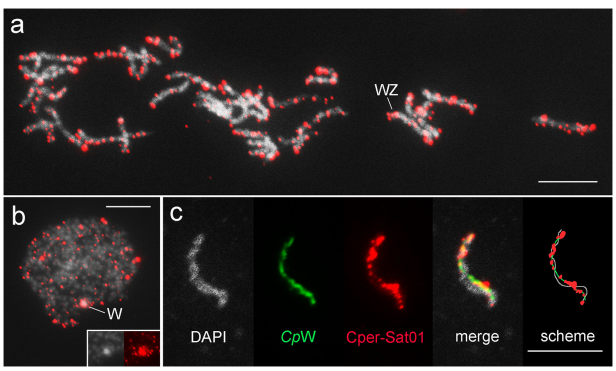

FIGURE 5 | Chromosomal mapping of Cper-Sat01 in pachytene oocyte bivalents **(a,c)** and interphase nucleus **(b)** of *Cydalima perspectalis*. In panel **(c)**, a detail of the WZ bivalent with the W chromosome identified by the W-painting probe (green) is shown, which allows accurate mapping of the distribution of Cper-Sat01 clusters (red). Note the enrichment of the clusters on the W chromosome relative to the Z chromosome. In panel **(b)**, note the strong hybridization signals concentrated on the W chromatin body; the inset shows a detail of the W chromatin body stained with DAPI (gray on the left) and with the probe (red on the right). Bar = 10 μm.

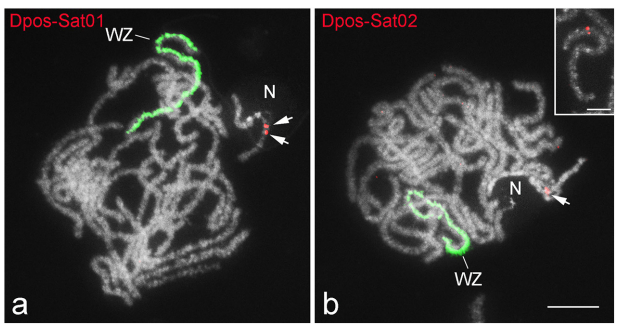

FIGURE 6 | FISH mapping of two satDNAs (red) and W-painting probe (green) in pachytene oocytes of *Diatraea postlineella*. Chromosomes were counterstained with DAPI. Note that both satDNAs show an autosomal location. Dpos-Sat01 **(a)** is located on bivalent carrying the nucleolus (N), while Dpos-Sat02 **(b)** is located on another bivalent. The inset in panel **(b)** shows in detail the interstitial position of Dpos-Sat02. Bar = 10 μm.

DISCUSSION

Karyotypes and Molecular Differentiation of Sex Chromosomes in Crambidae

In this study, we made progress in understanding the basic features of karyotypes and molecular differentiation of sex chromosomes in the economically important family of Lepidoptera, Crambidae. Although the number of Crambidae species karyotyped so far is low, a high variability of haploid numbers, ranging from $n = 10$ to $n = 41$ has been reported (Virkki, 1963; Robinson, 1971; Kageyma and Traut, 2004; Thakur and Gautam, 2013; Yasukochi et al., 2016). These variations in the number of chromosomes below and above the ancestral number of $n = 31$ (Van't Hof et al., 2013; Ahola et al., 2014) suggest a dynamic karyotype evolution by chromosome fusions and fissions, respectively. Fusions probably also played a major role in reducing the number of chromosomes to $n = 21$ found in this study in *D. postlineella*. The occurrence of large chromosomes in *D. postlineella*, which were difficult to individualize in pachytene compared to *O. nubilalis* and *C. perspectalis* with $n = 31$, also supports the fusion hypothesis. The fact that a related species, the sugarcane borer *D. saccharalis*, also has a reduced chromosome number ($n = 17$; Virkki, 1963), suggests that this is an evolutionary trend in the genus *Diatraea*. In lepidopteran species, a reduction in the chromosome number by fusion is generally more common compared to an increase by fission, which could be attributed to more deleterious effects of fissions in meiotic products (Pringle et al., 2007; de Vos et al., 2020).

The karyotypes of crambid species differed greatly in the number and chromosomal position of rDNA clusters.

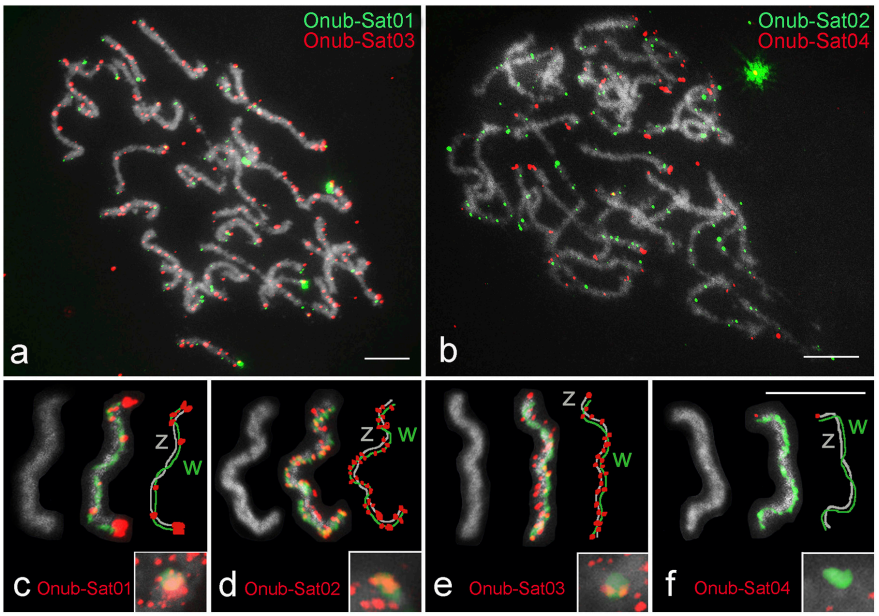

FIGURE 7 | Localization of four satDNAs identified from the genome of *Ostrinia nubilalis*. **(a,b)** Male pachytene. In panels **(c–f)**, selected WZ bivalents from pachytene oocytes stained with DAPI (gray), probed for individual satDNAs (red) and with the W chromosome identified by W-painting probe (green) are shown. Schematic representations of each WZ bivalent showing the W (green) and Z (gray) chromosomes and the satDNA distribution (red) are also given in panels **(c–f)**. The insets in panels **(c–f)** show hybridization signals of the probes on W chromatin from interphase nuclei. Probes are indicated in each image. Bar = 10 μm.

A single rDNA cluster in the interstitial position, as found in *D. postlineella*, is a common pattern for Lepidoptera (Nguyen et al., 2010). Two terminal rDNA clusters found in the karyotype of C. *perspectalis* were also observed in the Mediterranean flour moth, *Ephestia kuehniella*, from the sister family Pyralidae (Nguyen et al., 2010). Interestingly, the Z-strain of *O. nubilalis* showed four terminal rDNA clusters (this study), while specimens from another population showed five terminal clusters per haploid genome (see Nguyen et al., 2010), suggesting intra-specific variability in the number of rDNA clusters. Besides Crambidae, rDNA variability between species has also been reported in the lepidopteran family Tortricidae, but was mainly manifested by different chromosomal positions of rDNA clusters (Šíchová et al., 2013). The data obtained here for Crambidae add to the growing evidence of high rDNA dynamics in Lepidoptera, which is mainly attributed to ectopic recombination not only in Lepidoptera but also in other insects (Nguyen et al., 2010; Cabral-de-Mello et al., 2011; Ferretti et al., 2019).

We have shown that the studied crambid species have a WZ/ZZ (female/male) sex chromosome system, which is common in Lepidoptera (Sahara et al., 2012). In all three species, the W chromosome is highly molecularly differentiated from the Z chromosome and consists mainly of female-specific and/or female-enriched repetitive sequences, as in some other Lepidoptera (Sahara et al., 2003; Dalíková et al., 2017a,b; Zrzavá et al., 2018). Despite this apparent similarity in gross molecular composition, cross-hybridization experiments revealed that W chromosomes exhibit distinct molecular divergence between species. Nevertheless, some degree of homology has been demonstrated, suggesting partial conservation of repeats on the W chromosomes. However, this degree of homology is not consistent with the phylogeny of Crambidae subfamilies (Regier et al., 2012; Zhu et al., 2018; Léger et al., 2021), since the W probe from *O. nubilalis* (Pyraustinae) showed the highest number of signals in the W chromosome of distant *D. postlineella* (Crambinae), but a lower number of signals in the W chromosome of more closely related *C. perspectalis* (Spilomelinae). These results support the hypothesis of fast independent molecular divergence of W chromosomes with occasional conservation of some repetitive sequences. Partial homology between W chromosomes was also observed in species of the sister family Pyralidae, but the extent of homology was consistent with phylogenetical relationships (Vítková et al., 2007). In contrast, a high molecular divergence was found between the W chromosomes of two congeneric moths of the genus *Abraxas*, Geometridae (Zrzavá et al., 2018).

satDNAs in Crambidae

The interspecies occurrence of the seven satDNAs identified herein is consistent with the library hypothesis (Fry and Salser, 1977). A common feature of these newly identified satDNAs is the enrichment in A + T base pairs (light satDNAs), similar to the other five satDNAs previously described in Lepidoptera (Lu et al., 1994; Mandrioli et al., 2003; Mahendran et al., 2006; Věchtová et al., 2016; Dalíková et al., 2017b). Interestingly, the monomer length of some of the satDNAs identified herein is high, reaching 2,244 bp for Cper-Sat01 in *C. perspectalis* and more than 1,000 bp for two satDNAs in *D. postlineella*. This is an unusual feature in insects, which generally tend to have satDNA monomer sizes less than 600 bp (Palomeque and Lorite, 2008). Exceptions have

been found, for example, in the ant *Monomorium subopacum* with a monomer size of 2.5 kb (Lorite et al., 2004), the kissing bug *Triatoma infestans* with about 1 kb repeat unit (Pita et al., 2017), and in the beetles *Misolampus goutodii* and *Hippodamia variegata* with 1.2 and 2 kb repeat units, respectively (Pons et al., 1993; Mora et al., 2020). It is worth noting that the use of different methods for satDNA prospection could affect the capacity of discovering satDNA families with greater or smaller monomer lengths.

The absence of large heterochromatic blocks in most studied lepidopterans suggests the low abundance of tandem repeats forming long arrays in the chromosomes. This assumption is supported by the extremely low abundance of satDNAs in the three crambid species, in which satDNAs reached a maximum of 0.255% of the genome (mean value between the male and female genomes), as found here in *O. nubilalis*. Such a small proportion of satDNAs is in line with the generally small amount of heterochromatin in Lepidoptera, which is mostly observed only in the W chromosome (Traut et al., 2007) or associated with NOR regions (Nguyen et al., 2010; Šíchová et al., 2013). Only in exceptional cases were numerous heterochromatin blocks observed in other chromosomes (Šíchová et al., 2015, 2016; Zrzavá et al., 2018). This contrasts with other insects, such as hemipterans and grasshoppers, in which a high abundance of satDNAs was revealed by bioinformatics tools (Bardella et al., 2020; Palacios-Gimenez et al., 2020). This may be due to the large genome size of the latter insect groups compared to the generally small size of lepidopteran genomes (Gregory, 2020, Animal Genome Size Database) and also due to predominance of interspersed repeats in Lepidoptera (Talla et al., 2017).

Five of the seven satDNAs identified in Crambidae show a scattered distribution on chromosomes. This appears to be a recurrent pattern in Lepidoptera, as it was also observed in two other satDNAs mapped on chromosomes of moth species from distant families, Saturniidae and Tortricidae (Mahendran et al., 2006; Věchtová et al., 2016). The remaining two mapped satDNAs, MBSAT1 in *M. brassicae* (Noctuidae) and PiSAT1 in *P. interpunctella* (Pyralidae), were limited to sex chromosomes only (Mandrioli et al., 2003; Dalíková et al., 2017b). The predominant scattered distribution is in contrast to patterns observed in organism with large heterochromatin-rich genomes (e.g., Camacho et al., 2015; Pita et al., 2017). This might be caused by the fact that the formation of long arrays of tandem repeats requires a special environment, which is usually lacking in Lepidoptera. Vondrak et al. (2020) recently studied satDNAs of the legume plant *Lathyrus sativus* using ultra-long nanopore reads. They discovered that besides the long arrays of tandem repeats located in centromeric and subtelomeric heterochromatin blocks, satDNAs also formed short arrays in other parts of the genome associated with and probably originated from LTR transposons. The authors speculated that the formation of a long array of satDNA requires restriction of recombination, a process which might reduce the length of the array *via* homologous recombination. This condition is fulfilled, for example, in centromeres and sex chromosomes (Navrátilová et al., 2008; Vondrak et al., 2020). In holokinetic chromosomes of Lepidoptera, which lack a primary constriction (i.e., centromere), there are usually no suitable regions for long arrays of tandem repeats other than the W chromosome, which might be the reason why satDNAs often show a scattered distribution of small clusters. The occurrence of satDNAs in euchromatin, as found in Lepidoptera, has been reported in some insects (Brajković et al., 2012; Palacios-Gimenez et al., 2020; Sproul et al., 2020). Such satDNAs could be located near genes and modulate their expression (Brajković et al., 2012; Feliciello et al., 2015). This aspect deserves more research in Lepidoptera, due to the frequent location of satDNAs in euchromatin and transcription of some satDNAs (Věchtová et al., 2016; Dalíková et al., 2017b).

Distribution of satDNAs on Sex Chromosomes

Although it is well documented that the lepidopteran W chromosome is mostly composed of heterochromatin rich in repetitive DNA sequences, especially TEs, its specific molecular composition is largely unknown (Abe et al., 2005; Traut et al., 2007, 2013; Sahara et al., 2012). Our study contributed to this knowledge and demonstrated the occurrence of satDNAs on the W chromosomes of *C. perspectalis* and *O. nubilalis*, but not in *D. postlineella*. It should be noted that, unlike Lepidoptera, satDNAs are widely reported in the Y chromosomes of various animals (Giovannotti et al., 2018; Escudeiro et al., 2019; Utsunomia et al., 2019; Ferretti et al., 2020) and plants (Hobza et al., 2015, 2017). Among the tandem sequences identified on the lepidopteran W chromosomes are rDNA in few species (Yoshido et al., 2005; Van't Hof et al., 2008; Šíchová et al., 2013; Zrzavá et al., 2018) and two satDNAs, PiSAT1 in *P. interpunctella* (Dalíková et al., 2017b) and MBSAT1 in a *M. brassicae* cell line (Mandrioli et al., 2003). However, these satDNAs form conspicuous blocks on the W chromosomes, differing from the satDNAs of *C. perspectalis* and *O. nubilalis*, which are arranged as scattered clusters on the W chromosomes, similar to autosomes. Interestingly, AT-1 of *C. pomonella* is underrepresented on the W chromosome (Věchtová et al., 2016) and two satDNAs identified in *D. postlineella* are located exclusively on autosomes (this study). This demonstrates the great plasticity of satDNA arrangement in lepidopteran genomes, from absence to abundance on the W chromosomes and from a single cluster to a high number of scattered clusters.

The results of our bioinformatic analysis of satDNAs in both sexes of the crambid species studied suggest that in female genomes, ultimately in the W chromosome, there was no extensive amplification of these repeats, as is often reported in the Y or W chromosome in other animals (Palacios-Gimenez et al., 2017; Utsunomia et al., 2019; Ferretti et al., 2020). However, our complementary chromosomal analysis revealed enrichment of some satDNAs on the W (*C. perspectalis* and *O. nubilalis*) or Z chromosome (*O. nubilalis*). Because satDNAs are highly dynamic in copy number (Garrido-Ramos, 2017), the difference in abundance could be attributed to interindividual variability of autosomal clusters, masking the difference in abundance between sex chromosomes.

As observed for euchromatic autosomes, the absence of long arrays surprisingly also applies to the heterochromatic

W chromosome in the crambid species studied. Of the seven satDNAs analyzed, four were present on the W chromosome, three of which were abundant there. The patterns of the W-signals of the respective satDNA resembled those observed on the other chromosomes, indicating that the clusters were relatively small and scattered. The only satDNA that formed slightly larger subtelomeric clusters was the Onub-Sat01 (**Figure 7c**). Of all the lepidopteran satDNAs known today, the only one which forms long arrays in the genome is MBSAT1 in the cabbage moth, *M. brassicae*, which has been shown to be vastly abundant on both the W and Z chromosomes, occupying ca 1.9% of the genome (Mandrioli et al., 2003). However, it would be worth to reanalyze this case, as these results have been obtained from a cell line that showed a drastically reduced chromosome number (n = 11) compared to the number reported for the wild population of this species (n = 31) (Saitoh, 1959; Sahara et al., 2013). This suggests that the cell line may have undergone genome reconstruction accompanied by amplification of this satDNA, which may not occur in wild populations. Thus, the fact that there are no long arrays of satDNA on the W chromosome that is non-recombining because lepidopteran females lack recombination on all of their chromosomes remains unexplained.

Our CGH results suggest a high abundance of W-enriched and/or W-specific sequences on the W chromosome of all three crambid species. Combining CGH with satDNA mapping helped us to better understand nature of these sequences. In *D. postlineella*, the W-enriched and/or W-specific sequences are not satellites because the only two satDNAs we detected are low-copy autosomal repeats. Hence, the hybridization signals of the female gDNA probe, which highlighted the W chromosome, were generated by other types of sequences, such as mobile elements or microsatellites. In the case of *C. perspectalis*, the only identified Cper-Sat01 is abundant on both autosomes and the W chromosome, while it is underrepresented or absent on the Z chromosome. Thus, it is enriched in females and contributes to the W-highlighting using the female gDNA probe. In *O. nubilalis*, satDNAs mapped to autosomes as well as both sex chromosomes, except Onub-Sat04, which was absent on the W and formed a small cluster at one Z-chromosome end. Hence, the W-enriched and/or W-specific sequences are in fact other repeats than the satDNAs studied.

CONCLUSION

Our study significantly contributed to the understanding of karyotype diversification, genome architecture, and sex chromosome evolution in Lepidoptera. The karyotypes of the studied species of the family Crambidae differ in some aspects, namely in the gross architecture (i.e., diploid numbers), probably due to chromosomal fusions, and in fine structures, as evidenced by variability in the distribution of repetitive DNAs. We revealed low abundance and high variability of satDNAs in these Crambidae species, which also contributed to the plasticity of sex chromosomes. The W chromosomes are highly differentiated between the three species due to independent evolution, although some degree of random conservation (not consistent with species phylogeny) of some anonymous repeats has been found. Finally, the combination of genomic and chromosomal data allowed the number of satDNAs identified in Lepidoptera to be doubled, opening new avenues for the study of this genome fraction.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

FM and DC-D-M contributed to conception and design of the study. DC-D-M and MZ performed the experiments and analyzed the data. DC-D-M, MZ, and FM interpreted the data and wrote the first draft of the manuscript. SK prepared the W chromosome probes. PR obtained samples of *Diatraea postlineella* and ensured the correct determination of this species. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This study was supported by grants 17-13713S and 20-13784S of the Czech Science Foundation, CNPq-Conselho Nacional de Pesquisa e Desenvolvimento, and FAPESP-Fundação de Amparo à Pesquisa do Estado de São Paulo (BE.PQ, process number 2017/097319, Regular Research Grant, process number 2019/19069-7). SK acknowledges support from grant RO 0520 of the Ministry of Education, Youth, and Sports of the Czechia. This work was conducted as part of the IAEA National project Using the Sterile Insect Technique (SIT) to Assess it as an Alternative for the Control of the Sugarcane Borer in Pilot Areas (GUA5017).

ACKNOWLEDGMENTS

We would like to thank David G. Heckel and Melanie Unbehend for samples of *Ostrinia nubilalis* and Luis C. Arroyo from the Santa Ana sugarcane farm in Escuintla, Guatemala, for preparing *Diatraea postlineella* samples from the rearing facility. We would also like to thank Alma Solis (ARS, USDA) for access to an unpublished manuscript about *D. postlineella* and relevant pyraloid references. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.661417/full#supplementary-material

Supplementary Figure 1 | Mitotic chromosome spreads obtained from wing imaginal disks of female larvae of **(A)** *Cydalima perspectalis* (prometaphase) and **(B)** *Diatraea postlineella* (early anaphase) and stained with DAPI. Bar = 10 μm.

Supplementary Figure 2 | DAPI-stained pachytene bivalents obtained from testes (top images). **(A)** *Cydalima perspectalis* (*n* = 31), **(B)** *Diatraea postlineella* (*n* = 21) and **(C)** *Ostrinia nubilalis* (*n* = 31). A schematic drawing of the same pachytene nucleus is shown at the bottom of each panel; this way of schematic representation was used to determine the number of bivalents. Bar = 10 μm.

Supplementary Figure 3 | Orcein-stained polyploid nuclei obtained from Malpighian tubule cells of female **(A,C,E)** and male **(B,D,F)** larvae showing sex chromatin status in **(A,B)** *Cydalima perspectalis*, **(C,D)** *Diatraea postlineella*, and **(E,F)** *Ostrinia nubilalis*. Note the presence of a deeply stained W chromatin body (arrows) in females, while it is absent in males. Bar = 10 μm.

Supplementary Figure 4 | Cluster layouts and sequence logos for satDNAs. Graphical layouts were obtained from RepeatExplorer output. Sequence logos were generated from full-length consensus sequences from the most frequent k-mers for each satDNA.

Supplementary Table 1 | DNA sequences of primers used for PCR amplification of seven satDNAs and 18S rDNA in Crambidae species. Cper, *Cydalima perspectalis*; Dpos, *Diatraea postlineella*; Onub, *Ostrinia nubilalis*.

Supplementary Table 2 | Analysis of abundance of identified satDNAs at the interspecific level in the genomes of Crambidae species.

REFERENCES

Abbott, J. K., Nordén, A. K., and Hansson, B. (2017). Sex chromosome evolution: historical insights and future perspectives. *Proc. Biol. Sci.* 284:20162806. doi: 10.1098/rspb.2016.2806

Abe, H., Mita, K., Yasukochi, Y., Oshiki, T., and Shimada, T. (2005). Retrotransposable elements on the W chromosome of the silkworm, *Bombyx mori*. *Cytogenet. Genome Res.* 110, 144–151. doi: 10.1159/000084946

Acosta, M. J., Marchal, J. A., Martínez, S., Puerma, E., Bullejos, M., Díaz de la Guardia, R., et al. (2007). Characterization of the satellite DNA Msat-160 from the species *Chionomys nivalis* (Rodentia, Arvicolinae). *Genetica* 130, 43–51. doi: 10.1007/s10709-006-0018-1

Ahola, V., Lehtonen, R., Somervuo, P., Salmela, L., Koskinen, P., Rastas, P., et al. (2014). The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat. Commun.* 5:4737. doi: 10.1038/ncomms5737

Andrews, S. (2010). *FastQC: a Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babrahamac.uk/projects/fastqc/. (Accessed March 18, 2019).

Bardella, V. B., Milani, D., and Cabral-de-Mello, D. C. (2020). Analysis of *Holhymenia histrio* genome provides insight into the satDNA evolution in an insect with holocentric chromosomes. *Chromosome Res.* 28, 369–380. doi: 10.1007/s10577-020-09642-1

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

Brajković, J., Feliciello, I., Bruvo-Mađarić, B., and Ugarković, Đ (2012). Satellite DNA-like elements associated with genes within euchromatin of the beetle *Tribolium castaneum*. *G3 (Bethesda)* 2, 931–941. doi: 10.1534/g3.112.003467

Bull, J. J. (1983). *Evolution of Sex Determining Mechanisms*. San Francisco, CA: The Benjamin/Cummings Publishing Company, Inc.

Cabral-de-Mello, D. C., and Marec, F. (2021). Universal fluorescence in situ hybridization (FISH) protocol for mapping repetitive DNAs in insects and other arthropods. *Mol. Genet. Genom.* doi: 10.1007/s00438-021-01765-2 Online ahead of print.

Cabral-de-Mello, D. C., Moura, R. C., and Martins, C. (2011). Cytogenetic mapping of rRNAs and histone H3 genes in 14 species of *Dichotomius* (Coleoptera, Scarabaeidae, Scarabaeinae) beetles. *Cytogenet. Genome Res.* 134, 127–135. doi: 10.1159/000326803

Camacho, J. P., Ruiz-Ruano, F. J., Martín-Blázquez, R., López-León, M. D., Cabrero, J., Lorite, P., et al. (2015). A step to the gigantic genome of the desert locust: chromosome sizes and repeated DNAs. *Chromosoma* 124, 263–275. doi: 10.1007/s00412-014-0499-0

Charlesworth, B. (1991). The evolution of sex chromosomes. *Science* 251, 1030–1033. doi: 10.1126/science.1998119

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

Crepaldi, C., and Parise-Maltempi, P. P. (2020). Heteromorphic sex chromosomes and their DNA content in fish: an insight through satellite DNA accumulation in *Megaleporinus elongatus*. *Cytogenet. Genome Res.* 160, 38–46. doi: 10.1159/000506265

da Silva, M. J., Fogarin Destro, R., Gazoni, T., Narimatsu, H., Pereira, dos Santos, P. S., et al. (2020). Great abundance of satellite DNA in *Proceratophrys* (Anura, Odontophrynidae) revealed by genome sequencing. *Cytogenet. Genome Res.* 160, 141–147. doi: 10.1159/000506531

Dalíková, M., Zrzavá, M., Hladová, I., Nguyen, P., Šonský, I., Flegrová, M., et al. (2017a). New insights into the evolution of the W chromosome in Lepidoptera. *J. Hered.* 108, 709–719. doi: 10.1093/jhered/esx063

Dalíková, M., Zrzavá, M., Kubíčková, S., and Marec, F. (2017b). W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). *Chromosome Res.* 25, 241–252. doi: 10.1007/s10577-017-9558-8

de Vos, J. M., Augustijnen, H., Bätscher, L., and Lucek, K. (2020). Speciation through chromosomal fusion and fission in Lepidoptera. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375, 20190539. doi: 10.1098/rstb.2019.0539

Escudeiro, A., Adega, F., Robinson, T. J., Heslop-Harrison, J. S., and Chaves, R. (2019). Conservation, divergence, and functions of centromeric satellite DNA families in the Bovidae. *Gen. Biol. Evol.* 11, 1152–1165. doi: 10.1093/gbe/evz061

Feliciello, I., Akrap, I., and Ugarković, Đ (2015). Satellite DNA modulates gene expression in the beetle *Tribolium castaneum* after heat stress. *PLoS Genet.* 11:e1005466. doi: 10.1371/journal.pgen.1005466

Ferretti, A. B. S. M., Milani, D., Palacios-Gimenez, O. M., Ruiz-Ruano, F. J., and Cabral-de-Mello, D. C. (2020). High dynamism for neo-sex chromosomes: satellite DNAs reveal complex evolution in a grasshopper. *Heredity* 125, 124–137. doi: 10.1038/s41437-020-0327-7

Ferretti, A. B. S. M., Ruiz-Ruano, F. J., Milani, D., Loreto, V., Martí, D. A., Ramos, E., et al. (2019). How dynamic could be the 45S rDNA cistron? An intriguing variability in a grasshopper species revealed by integration of chromosomal and genomic data. *Chromosoma* 128, 165–175. doi: 10.1007/s00412-019-00706-8

Fraïsse, C., Picard, M. A. L., and Vicoso, B. (2017). The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nat. Commun.* 8:1486. doi: 10.1038/s41467-017-01663-5

Fry, K., and Salser, W. (1977). Nucleotide sequences of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* 12, 1069–1084. doi: 10.1016/0092-8674(77)90170-2

Fuková, I., Nguyen, P., and Marec, F. (2005). Codling moth cytogenetics: karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. *Genome* 48, 1083–1092. doi: 10.1139/g05-063

Fuková, I., Traut, W., Vítková, M., Kubíčková, S., and Marec, F. (2007). Probing the W chromosome of the codling moth, *Cydia pomonella*, with sequences from microdissected sex chromatin. *Chromosoma* 116, 135–145. doi: 10.1007/s00412-006-0086-0

Furman, B. L. S., Metzger, D. C. H., Darolti, I., Wright, A. E., Sandkam, B. A., Almeida, P., et al. (2020). Sex chromosome evolution: so many exceptions to the rules. *Genome Biol. Evol.* 12, 750–763. doi: 10.1093/gbe/evaa081

Garrido-Ramos, M. A. (2017). Satellite DNA: an evolving topic. *Genes* 8:230. doi: 10.3390/genes8090230

Gatto, K. P., Mattos, J. V., Seger, K. R., and Lourenço, L. B. (2018). Sex chromosome differentiation in the frog genus *Pseudis* involves satellite DNA

and chromosome rearrangements. *Front. Genet.* 9:301. doi: 10.3389/fgene.2018.00301

Giovannotti, M., Cerioni, P. N., Rojo, V., Olmo, E., Slimani, T., Splendiani, A., et al. (2018). Characterization of a satellite DNA in the genera *Lacerta* and *Timon* (Reptilia, Lacertidae) and its role in the differentiation of the W chromosome. *J. Exp. Zool. Part B Mol. Develop. Evol.* 330, 83–95. doi: 10.1002/jez.b.22790

Gordon, A., and Hannon, G. J. (2010). *Fastx-toolkit. FASTQ/A Short-reads Pre-processing Tools*. Available online at: http://hannonlab.cshl.edu/fastx_toolkit. (Accessed April 8, 2020).

Gregory, T. R. (2020). *Animal Genome Size Database*. Available online at: http://www.genomesize.com. (Accessed November 26, 2020).

Guthrie, W. D., Dollinger, E. J., and Stetson, J. F. (1965). Chromosome studies of the European corn borer, smartweed borer, and lotus borer (Pyralidae). *Ann. Entomol. Soc. Amer.* 58, 100–105. doi: 10.1093/aesa/58.1.100

Hejníčková, M., Koutecký, P., Potocký, P., Provazníková, I., Voleníková, A., Dalíková, M., et al. (2019). Absence of W chromosome in Psychidae moths and implications for the theory of sex chromosome evolution in Lepidoptera. *Genes* 10:1016. doi: 10.3390/genes10121016

Hobza, R., Cegan, R., Jesionek, W., Kejnovsky, E., Vyskot, B., and Kubat, Z. (2017). Impact of repetitive elements on the Y chromosome formation in plants. *Genes* 8:302. doi: 10.3390/genes8110302

Hobza, R., Kubat, Z., Cegan, R., Jesionek, W., Vyskot, B., and Kejnovsky, E. (2015). Impact of repetitive DNA on sex chromosome evolution in plants. *Chromosome Res.* 23, 561–570. doi: 10.1007/s10577-015-9496-2

Ijdo, J. W., Wells, R. A., Baldini, A., and Reeders, S. T. (1991). Improved telomere detection using a telomere repeat probe (TTAGGG)$_n$ generated by PCR. *Nucleic Acids Res.* 19:4780. doi: 10.1093/nar/19.17.4780

Kageyma, D., and Traut, W. (2004). Opposite sex-specific effects of *Wolbachia* and interference with the sex determination of its host *Ostrinia scapulalis*. *Proc. Biol. Sci.* 271, 251–258. doi: 10.1098/rspb.2003.2604

Kato, A., Albert, P. S., Veja, J. M., and Birchler, J. A. (2006). Sensitive fluorescence in situ hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. *Biotechnic Histoch.* 81, 71–78. doi: 10.1080/10520290600643677

Kubickova, S., Cernohorska, H., Musilova, P., and Rubes, J. (2002). The use of laser microdissection for the preparation of chromosome-specific painting probes in farm animals. *Chromosome Res.* 10, 571–577. doi: 10.1023/A:1020914702767

Léger, T., Mally, R., Neinhuis, C., and Nuss, M. (2021). Refining the phylogeny of Crambidae with complete sampling of subfamilies (Lepidoptera, Pyraloidea). *Zool. Scripta.* 50, 84–99. doi: 10.1111/zsc.12452

Lewis, L. C., and Lynch, R. E. (1969). Rearing the European corn borer, *Ostrinia nubilalis* (Hubner), on diets containing corn leaf and wheat germ. *Iowa State J. Sci.* 44, 9–14.

Lorite, P., Carrillo, J. A., Aguilar, J. A., and Palomeque, T. (2004). Isolation and characterization of two families of satellite DNA with repetitive units of 135 bp and 2.5 kb in the ant *Monomorium subopacum* (Hymenoptera, Formicidae). *Cytogenet. Genome Res.* 105, 83–92. doi: 10.1159/000078013

Lu, Y. J., Kochert, G. D., Isenhour, D. J., and Adang, M. J. (1994). Molecular characterization of a strain-specific repeated DNA sequence in the fall armyworm *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Insect Mol. Biol.* 3, 123–130. doi: 10.1111/j.1365-2583.1994.tb00159.x

Lynch, R. E. (1980). European corn borer: yield losses in relation to hybrid and stage of corn development. *J. Econ. Entomol.* 73, 159–164. doi: 10.1093/jee/73.1.159

Mahendran, B., Acharya, C., Dash, R., Ghosh, S. K., and Kundu, S. C. (2006). Repetitive DNA in tropical tasar silkworm *Antheraea mylitta*. *Gene* 370, 51–57. doi: 10.1016/j.gene.2005.11.010

Mandrioli, M., Manicardi, G. C., and Marec, F. (2003). Cytogenetic and molecular characterization of the MBSAT1 satellite DNA in holokinetic chromosomes of the cabbage moth, *Mamestra brassicae* (Lepidoptera). *Chromosome Res.* 11, 51–56. doi: 10.1023/A:1022058032217

Marec, F., and Traut, W. (1994). Sex chromosome pairing and sex chromatin bodies in W-Z translocation strains of *Ephestia kuehniella* (Lepidoptera). *Genome* 37, 426–435. doi: 10.1139/g94-060

Mata-Sucre, Y., Sader, M., Van-Lume, B., Gagnon, E., Pedrosa-Harand, A., Leitch, I. J., et al. (2020). How diverse is heterochromatin in the Caesalpinia group? Cytogenomic characterization of *Erythrostemon hughesii* Gagnon & G.P. Lewis (Leguminosae: Caesalpinioideae). *Planta* 252:49. doi: 10.1007/s00425-020-03453-8

Mediouni, J., Fuková, I., Frydrychová, R., Dhouibi, M. H., and Marec, F. (2004). Karyotype, sex chromatin and sex chromosome differentiation in the carob moth, *Ectomyelois ceratoniae* (Lepidoptera: Pyralidae). *Caryologia* 57, 184–194. doi: 10.1080/00087114.2004.10589391

Meissle, M., Mouron, P., Musa, T., Bigler, F., Pons, X., Vasileiadis, V. P., et al. (2010). Pests, pesticide use and alternative options in European maize production: current status and future prospects. *J. Appl. Entomol.* 134, 357–375. doi: 10.1111/j.1439-0418.2009.01491.x

Mora, P., Vela, J., Ruiz-Ruano, F. J., Ruiz-Mena, A., Montiel, E. E., Palomeque, T., et al. (2020). Satellitome analysis in the ladybird beetle *Hippodamia variegata* (Coleoptera, Coccinellidae). *Genes* 11:783. doi: 10.3390/genes11070783

Munroe, E., and Solis, M. A. (1999). "The Pyraloidea," in *Handbook of Zoology IV: Lepidoptera, Moths and Butterflies, Vol. 1, Arthropoda, Insect, Vol.4, Part 35*, ed. N. Kristensen (Berlin: Walter de Gruyter & Co.), 491.

Nacambo, S., Leuthardt, F. L. G., Wan, H., Li, H., Haye, T., Baur, B., et al. (2013). Development characteristics of the box-tree moth *Cydalima perspectalis* and its potential distribution in Europe. *J. Appl. Entomol.* 138, 14–26. doi: 10.1111/jen.12078

Navrátilová, A., Koblížková, A., and Macas, J. (2008). Survey of extrachromosomal circular DNA derived from plant satellite repeats. *BMC Plant Biol.* 8:90. doi: 10.1186/1471-2229-8-90

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

Nguyen, P., Sýkorová, M., Šíchová, J., Kůta, V., Dalíková, M., Čapková Frydrychová, R., et al. (2013). Neo-sex chromosomes and adaptive potential in tortricid pests. *Proc. Natl. Acad. Sci. U S A* 110, 6931–6936. doi: 10.1073/pnas.1220372110

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. *Bioinformatics* 29, 792–793. doi: 10.1093/bioinformatics/btt054

Novák, P., Robledillo, L. A., 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

Nuss, M., Landry, B., Mally, R., Vegliante, F., Trankner, A., Bauer, F., et al. (2003–2020). *Global informationsystem on Pyraloidea*. Available online at: http://www.pyral oidea.org. (Accessed December 16, 2020).

Palacios-Gimenez, O. M., Dias, G. B., Lima, L. G., Kuhn, G. C. S., Ramos, E., Martins, C., et al. (2017). High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis*. *Sci. Rep.* 7:6422. doi: 10.1038/s41598-017-06822-6828

Palacios-Gimenez, O. M., Milani, D., Song, H., Martí, D. A., López-León, M. D., Ruiz-Ruano, F. J., et al. (2020). Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. *Genome Biol. Evol.* 12, 88–102. doi: 10.1093/gbe/evaa018

Palomeque, T., and Lorite, P. (2008). Satellite DNA in insects: a review. *Heredity* 100, 564–573. doi: 10.1038/hdy.2008.24

Pita, S., Panzera, F., Mora, P., Vela, J., Cuadrado, A., Sánches, A., et al. (2017). Comparative repeatome analysis on *Triatoma infestans* Andean and non-Andean lineages, main vector of Chagas disease. *PLoS One* 12:e0181635. doi: 10.1371/journal.pone.0181635

Pons, J., Petitpierre, E., and Juan, C. (1993). Characterization of the heterochromatin of the darkling beetle *Misolampus goudoti*: cloning of two satellite DNA families and digestion of chromosomes with restriction enzymes. *Hereditas* 119, 179–185. doi: 10.1111/j.1601-5223.1993.00179.x

Pringle, E. G., Baxter, S. W., Webster, C. L., Papanicolaou, A., Lee, S. F., and Jiggins, C. D. (2007). Synteny and chromosome evolution in the Lepidoptera: evidence from mapping in *Heliconius melpomene*. *Genetics* 177, 417–426. doi: 10.1534/genetics.107.073122

Regier, J. C., Mitter, C., Solis, M. A., Hayden, J. E., Landry, B., Nuss, M., et al. (2012). A molecular phylogeny for the pyraloid moths (Lepidoptera: Pyraloidea) and

its implications for higher-level classification. *Syst. Entomol.* 37, 635–656. doi: 10.1111/j.1365-3113.2012.00641.x

Robinson, R. (1971). *Lepidoptera Genetics*. Oxford: Pergamon.

Sahara, K., Marec, F., Eickhoff, U., and Traut, W. (2003). Moth sex chromatin probed by comparative genomic hybridization (CGH). *Genome* 46, 339–342. doi: 10.1139/G03-003

Sahara, K., Yoshido, A., Shibata, F., Fujikawa-Kojima, N., Okabe, T., Tanaka-Okuyama, M., et al. (2013). FISH identification of *Helicoverpa armigera* and *Mamestra brassicae* chromosomes by BAC and fosmid probes. *Insect Biochem. Mol. Biol.* 43, 644–653. doi: 10.1016/j.ibmb.2013.04.003

Sahara, K., Yoshido, A., and Traut, W. (2012). Sex chromosome evolution in moths and butterflies. *Chromosome Res.* 20, 83–94. doi: 10.1007/s10577-011-9262-z

Saitoh, K. (1959). The chromosome numbers of some species of moths. *Jap. J. Genet.* 34, 84–87.

Serrano-Freitas, E. A., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Utsunomia, R., Araya-Jaime, C., Oliveira, C., et al. (2020). Satellite DNA content of B chromosomes in the characid fish *Characidium gomesi* supports their origin from sex chromosomes. *Mol. Genet. Genom.* 295, 195–207. doi: 10.1007/s00438-019-01615-2

Šíchová, J., Nguyen, P., Dalíková, M., and Marec, F. (2013). Chromosomal evolution in tortricid moths: conserved karyotypes with diverged features. *PLoS One* 8:e64520. doi: 10.1371/journal.pone.0064520

Šíchová, J., Ohno, M., Dincă, V., Watanabe, M., Sahara, K., and Marec, F. (2016). Fissions, fusions, and translocations shaped the karyotype and multiple sex chromosome constitution of the northeast-Asian wood white butterfly, *Leptidea amurensis*. *Biol. J. Linnean Soc.* 118, 457–471. doi: 10.1111/bij.12756

Šíchová, J., Voleníková, A., Dincă, V., Nguyen, P., Vila, R., Sahara, K., et al. (2015). Dynamic karyotype evolution and unique sex determination systems in *Leptidea* wood white butterflies. *BMC Evol. Biol.* 15:89. doi: 10.1186/s12862-015-0375-4

Silva, B. S. M. L., Heringer, P., Dias, G. B., Svartman, M., and Kuhn, G. C. S. (2019). De novo identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines. *PLoS One* 14:e0223466. doi: 10.1371/journal.pone.0223466

Silva, D. M. Z. A., Utsunomia, R., Ruiz-Ruano, F. J., Daniel, S. N., Porto-Foresti, F., Hashimoto, D. T., et al. (2017). High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Sci. Rep.* 7:12726. doi: 10.1038/s41598-017-12939-7

Smit, A. F. A., Hubley, R., and Green, P. (2013–2015). *RepeatMasker Open-4.0*. Available online at: http://www.repeatmasker.org. (Accessed Dec 14, 2018).

Smith, S. G. (1945). Heteropycnosis as a means of diagnosing sex. *J. Hered.* 36, 195–196. doi: 10.1093/oxfordjournals.jhered.a105498

Solis, M. A. (1997). "Snout moths: unraveling the taxonomic diversity of a speciose group in the neotropics," in *Biodiversity II: Understanding and Protecting our Biological Resources*, eds M. L. Reaka-Kudla, D. Wilson, and E. O. Wilson (Washington, D. C: Joseph Henry Press), 231–242.

Solis, M. A., and Metz, M. (2016). An illustrated guide to the identification of the known species of *Diatraea* Guilding (Lepidoptera: Crambidae: Crambinae) based on genitalia. *Zookeys.* 565, 73–121. doi: 10.3897/zookeys.565.6797

Solis, M. A., Scheffer, S. J., Lewis, M. L., and Rendon, P. (accepted). *Diatraea postlineella* Schaus (Lepidoptera: Crambidae) from Guatemala: molecular identity and host plant. *Proc. Entomol. Soc. Wash.*

Sproul, J. S., Khost, D. E., Eickbush, D. G., Negm, S., Wei, X., Wong, I., et al. (2020). Dynamic evolution of euchromatic satellites on the X chromosome in *Drosophila melanogaster* and the *simulans* clade. *Mol. Biol. Evol.* 37, 2241–2256. doi: 10.1093/molbev/msaa078

Talla, V., Suh, A., Kalsoom, F., Dincă, V., Vila, R., Friberg, M., et al. (2017). Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (*Leptidea*) butterflies. *Genome Biol. Evol.* 9, 2491–2505. doi: 10.1093/gbe/evx163

Thakur, R., and Gautam, D. C. (2013). Chromosome studies on four species of moths. *Cytologia* 78, 327–331. doi: 10.1508/cytologia.78.327

Traut, W., and Marec, F. (1996). Sex chromatin in Lepidoptera. *Quar. Rev. Biol.* 71, 239–256. doi: 10.1086/419371

Traut, W., and Marec, F. (1997). Sex chromosome differentiation in some species of Lepidoptera (Insecta). *Chromosome Res.* 5, 283–291. doi: 10.1023/B:CHRO.0000038758.08263.c3

Traut, W., Sahara, K., and Marec, F. (2007). Sex chromosomes and sex determination in Lepidoptera. *Sex. Dev.* 1, 332–346. doi: 10.1159/000111765

Traut, W., Sahara, K., Otto, T. D., and Marec, F. (1999). Molecular differentiation of sex chromosomes probed by comparative genomic hybridization. *Chromosoma* 108, 173–180. doi: 10.1007/s00412005036 6

Traut, W., Vogel, H., Glöckner, G., Hartmann, E., and Heckel, D. G. (2013). High-throughput sequencing of a single chromosome: a moth W chromosome. *Chromosome Res.* 110, 491–505. doi: 10.1007/s10577-013-9376-6

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596

Utsunomia, R., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Goes, C. A. G., Melo, S., Ramos, L. P., et al. (2019). Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* 9:5856. doi: 10.1038/s41598-019-42383-8

Van Nieukerken, E. J., Kaila, L., Kitching, I. J., Kristensen, N. P., Lees, D. C., Minet, J., et al. (2011). "Order Lepidoptera Linnaeus, 1758," in *Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness*, ed. Z. Q. Zhang (Auckland: Magnolia Press).

Van't Hof, A. E., Marec, F., Saccheri, I. J., Brakefield, P. M., and Zwaan, B. J. (2008). Cytogenetic characterization and AFLP-based genetic linkage mapping for the butterfly *Bicyclus anynana*, covering all 28 karyotyped chromosomes. *PLoS One* 3:e3882. doi: 10.1371/journal.pone.0003882

Van't Hof, A. E., Nguyen, P., Dalíková, M., Edmonds, N., Marec, F., and Saccheri, I. J. (2013). Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism. *Heredity* 110, 283–295. doi: 10.1038/hdy.2012.84

Věchtová, P., Dalíková, M., Sýkorová, M., Žurovcová, M., Füssy, Z., and Zrzavá, M. (2016). CpSAT-1, a transcribed satellite sequence from the codling moth, *Cydia pomonella*. *Genetica* 144, 385–395. doi: 10.1007/s10709-016-9907-0

Virkki, N. (1963). Gametogenesis in the sugarcane borer moth, *Diatraea saccharalis* (F.) (Crambidae). *J. Agricult. University Puerto Rico* 47, 102–137. doi: 10.46429/jaupr.v47i2.12944

Vítková, M., Fuková, I., Kubíčková, S., and Marec, F. (2007). Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera). *Chromosome Res.* 15, 917–930. doi: 10.1007/s10577-007-1173-7

Vondrak, T., Robledillo, L. A., Novák, P., Koblížková, A., Neumann, P., and Macas, J. (2020). Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. *Plant J.* 101, 484–500. doi: 10.1111/tpj.14546

Wei, K. H. C., and Barbash, D. A. (2015). Never settling down: frequent changes in sex chromosomes. *PLoS Biol.* 13:e1002077. doi: 10.1371/journal.pbio.1002077

Winnepenninckx, B., Backeljau, T., and de Wachter, R. (1993). Extraction of high molecular weight DNA from molluscs. *Trends Genet.* 12:407. doi: 10.1016/0168-9525(93)90102-n

Wright, A. E., Dean, R., Zimmer, F., and Mank, J. E. (2016). How to make a sex chromosome. *Nat. Commun.* 7:12087. doi: 10.1038/ncomms12087

Yasukochi, Y., Ohno, M., Shibata, F., Jouraku, A., Nakano, R., Ishikawa, Y., et al. (2016). A FISH-based chromosome map for the European corn borer yields insights into ancient chromosomal fusions in the silkworm. *Heredity* 116, 75–83. doi: 10.1038/hdy.2015.72

Yoshido, A., Marec, F., and Sahara, K. (2005). Resolution of sex chromosome constitution by genomic in situ hybridization and fluorescence in situ hybridization with $(TTAGG)_n$ telomeric probe in some species of Lepidoptera. *Chromosoma* 114, 193–202. doi: 10.1007/s00412-005-0013-9

Yoshido, A., Šíchová, J., Pospíšilová, K., Nguyen, P., Voleníková, A., Šafář, J., et al. (2020). Evolution of multiple sex-chromosomes associated with dynamic

genome reshuffling in *Leptidea* wood-white butterflies. *Heredity* 125, 138–154. doi: 10.1038/s41437-020-0325-9

Zhu, W., Yan, J., Song, J., and You, P. (2018). The first mitochondrial genomes for Pyralinae (Pyralidae) and Glaphyriinae (Crambidae), with phylogenetic implications of Pyraloidea. *PLoS One* 13:e0194672. doi: 10.1371/journal.pone.0194672

Zrzavá, M., Hladová, I., Dalíková, M., Šíchová, J., Õunap, E., Kubíčková, S., et al. (2018). Sex chromosomes of the iconic moth *Abraxas grossulariata* (Lepidoptera, Geometridae) and its congener *A. sylvata*. *Genes* 9:279. doi: 10.3390/genes9060279

Zwyrtková, J., Němečková, A., Čížková, J., Holušová, K., Kapustová, V., Svačina, R., et al. (2020). Comparative analyses of DNA repeats and identification of a novel Fesreba centromeric element in fescues and ryegrasses. *BMC Plant Biol.* 20:280. doi: 10.1186/s12870-020-02495-0

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

The reviewers DS and PR declared a shared affiliation with several of the authors, DC-D-M and MZ respectively, to the handling editor at the time of review.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 July 2021
doi: 10.3389/fgene.2021.683668

Comparative Repeat Profiling of Two Closely Related Conifers (*Larix decidua* and *Larix kaempferi*) Reveals High Genome Similarity With Only Few Fast-Evolving Satellite DNAs

Tony Heitkam[1], Luise Schulte[1,2‡], Beatrice Weber[1], Susan Liedtke[1], Sarah Breitenbach[1], Anja Kögler[1], Kristin Morgenstern[3], Marie Brückner[4], Ute Tröber[4], Heino Wolf[4], Doris Krabel[3] and Thomas Schmidt[1†]*

[1]Institute of Botany, Technische Universität Dresden, Dresden, Germany, [2]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany, [3]Institute of Forest Botany and Forest Zoology, Technische Universität Dresden, Tharandt, Germany, [4]Staatsbetrieb Sachsenforst, Pirna, Germany

In eukaryotic genomes, cycles of repeat expansion and removal lead to large-scale genomic changes and propel organisms forward in evolution. However, in conifers, active repeat removal is thought to be limited, leading to expansions of their genomes, mostly exceeding 10 giga base pairs. As a result, conifer genomes are largely littered with fragmented and decayed repeats. Here, we aim to investigate how the repeat landscapes of two related conifers have diverged, given the conifers' accumulative genome evolution mode. For this, we applied low-coverage sequencing and read clustering to the genomes of European and Japanese larch, *Larix decidua* (Lamb.) Carrière and *Larix kaempferi* (Mill.), that arose from a common ancestor, but are now geographically isolated. We found that both *Larix* species harbored largely similar repeat landscapes, especially regarding the transposable element content. To pin down possible genomic changes, we focused on the repeat class with the fastest sequence turnover: satellite DNAs (satDNAs). Using comparative bioinformatics, Southern, and fluorescent *in situ* hybridization, we reveal the satDNAs' organizational patterns, their abundances, and chromosomal locations. Four out of the five identified satDNAs are widespread in the *Larix* genus, with two even present in the more distantly related *Pseudotsuga* and *Abies* genera. Unexpectedly, the EulaSat3 family was restricted to *L. decidua* and absent from *L. kaempferi*, indicating its evolutionarily young age. Taken together, our results exemplify how the accumulative genome evolution of conifers may limit the overall divergence of repeats after speciation, producing only few repeat-induced genomic novelties.

Keywords: *Larix decidua* (Mill.), *Larix kaempferi* (Lamb.) Carrière, conifer, satellite DNA, tandem repeat, retrotransposon, repetitive DNA, fluorescent *in situ* hybridization

OPEN ACCESS

Edited by:
Francisco J. Ruiz-Ruano, University of East Anglia, United Kingdom

Reviewed by:
Alexander Belyayev, Institute of Botany (ASCR), Czechia
Visnja Besendorfer, University of Zagreb, Croatia
Laura Avila Robledillo, Academy of Sciences of the Czech Republic (ASCR), Czechia

****Correspondence:***
Tony Heitkam
tony.heitkam@tu-dresden.de

[†]Deceased

[‡]Present address:
Luise Schulte, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *21 March 2021*
Accepted: *25 May 2021*
Published: *12 July 2021*

Citation:
Heitkam T, Schulte L, Weber B, Liedtke S, Breitenbach S, Kögler A, Morgenstern K, Brückner M, Tröber U, Wolf H, Krabel D and Schmidt T (2021) Comparative Repeat Profiling of Two Closely Related Conifers (Larix decidua and Larix kaempferi) Reveals High Genome Similarity With Only Few Fast-Evolving Satellite DNAs. Front. Genet. 12:683668. doi: 10.3389/fgene.2021.683668

INTRODUCTION

Ranging in size between 0.002 and nearly 150 Gb, eukaryotic genomes vary by several orders of magnitude (Hidalgo et al., 2017). Among those, conifer genomes are especially large with sizes up to 37 Gb (Ahuja and Neale, 2005). As new reference genome sequences are generated – among them conifers such as spruces, pines, and recently firs and larches – new insights into the composition of conifer genomes are brought forward (Nystedt et al., 2013; Wegrzyn et al., 2014; Stevens et al., 2016; Kuzmin et al., 2019; Mosca et al., 2019). It is becoming obvious that the conifer karyotypes are highly conserved with $2n = 2x = 24$ chromosomes, that their genome sizes can be huge, ranging between 8 and 72 Gb, and that polyploidy can be largely excluded as a source of genome growth (Zonneveld, 2012; Neale and Wheeler, 2019a). Instead, one of the main takeaways is that the steady accumulation of repeats is the main driver for conifer genome expansion, presumably due to limited elimination of transposable elements (TEs; Nystedt et al., 2013; Prunier et al., 2016).

As the large conifer genomes have accumulated repeats over long periods of time with only slow removal and turnover of repetitive sequences, we wondered whether species-specific repeat profiles were able to evolve in closely related conifers. Regarding repetitive sequence classes, it is already hypothesized that TE families likely persist in conifers over long evolutionary timeframes (Zuccolo et al., 2015). In contrast, satellite DNAs (satDNAs) have much faster sequence turnovers than TEs. They form one of the major repeat groups, constituting up to 36% of some plant genomes (Ambrozová et al., 2011; Garrido-Ramos, 2017). SatDNAs are composed of short monomers with individual lengths often between 160 and 180 bp as well as 320 and 360 bp (Hemleben et al., 2007; Melters et al., 2013), and are arranged in long tandemly repeated arrays. As they confer important functions with roles in cell division, chromatid separation, and chromosome stability (Jagannathan et al., 2018), they often occupy specific chromosomal regions, such as the centromeres and the (sub-) telomeres (Schmidt and Heslop-Harrison, 1998; Melters et al., 2013; Oliveira and Torres, 2018). Due to their fast evolution and defined chromosomal localization, satDNAs may represent valuable targets to trace repeat evolution and divergence over long, evolutionary timeframes in conifers.

As models, we investigate two related conifers within the genus *Larix*, the deciduous European and Japanese larches, i.e., *Larix decidua* (Lamb.) Carrière and *Larix kaempferi* (Mill.). According to the fossil record, larches have been widespread in Asia and North America, and only reached Europe in the last million years (LePage and Basinger, 1995). Their genome size estimates range between 9.7 and 13 Gb (*L. decidua*) as well as 12.9 and 13.3 Gb (*L. kaempferi*; Zonneveld, 2012; Bennett and Leitch, 2019), with their huge genomes likely being the result of many divergent and ancient repeats (Nystedt et al., 2013; Pellicer et al., 2018). Larches frequently hybridize, leading to an unclear genetic basis with debated phylogenetic positions of individual species (Wei and Wang, 2003; Lu et al., 2014). From a breeding perspective, the interspecific hybrid *Larix* × *eurolepis* (with parental contributions of *L. decidua* and *L. kaempferi*) offers interesting possibilities for larch cultivation outside the natural range, especially in Europe (Pâques et al., 2013); however, determining the parental contributions to the traits of larch hybrids remains difficult.

Consistent with other Pinaceae species, all larches have $2n = 2x = 24$ chromosomes with conserved sizes, divided into six meta- and six submetacentric chromosome pairs (Hizume et al., 1993; Prunier et al., 2016). In larches, the 18S-5.8S-26S (35S) and 5S rDNAs are physically separated (Garcia and Kovařík, 2013), and fluorescent *in situ* hybridization (FISH) with respective rDNA probes clearly mark three 35S and one 5S rDNA-labeled chromosome pairs for *L. decidua* (Lubaretz et al., 1996) and two 35S and one 5S rDNA-labeled chromosome pairs for *L. kaempferi* (Liu et al., 2006; Zhang et al., 2010). Similarly, a single satDNA family is known ("LPD"), marking a heterochromatic chromosomal band on 22 chromosomes in *L. kaempferi* (Hizume et al., 2002). However, how the accumulative genome evolution mode of conifers affects the landscapes of larch repeats after speciation is not understood by far.

To test this, we sequenced *L. decidua* and *L. kaempferi* in low coverage to quantify, classify, and compare the respective repeat fractions. As satDNAs are typically marked by high sequence turnovers, we expect the highest differences for this repeat class. Using comparative bioinformatics, Southern, and fluorescent *in situ* hybridization, we deeply profiled five selected satDNAs, focusing on their abundance, their higher order arrangements, and chromosomal location. Assessment of their genomic distribution over a wider range of gymnosperms may give insight into the evolutionary age of satDNAs, may allow pinpointing how conifer repeat landscapes have diverged after speciation, and may be used to gather information regarding the parentage of larch hybrids.

MATERIALS AND METHODS

Plant Material and DNA Isolation

Needles and seeds of eleven gymnosperm accessions have been obtained from the Forest Botanical Garden of Tharandt (Technische Universität Dresden) and the Staatsbetrieb Sachsenforst (**Table 1**). DNA was isolated from 2 g of homogenized material from frozen needles using the DNeasy Plant Maxi Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. To allow for a more efficient elution of conifer DNA, the incubation time during the elution step was increased to 10 min. Purified DNA was eluted into water instead of the provided AE buffer.

For cytogenetics, we have used primary root tips from seeds of *L. decidua* (obtained as selected material for propagation from Staatsdarre Flöha, Partie number 1846, ELA/83704) and *L. kaempferi* seeds (obtained from Niedersächsische Landesforsten, provenance number 83901), as well as root tips from *L.* × *eurolepis* plantlets (clone 56.012.15) obtained from somatic embryogenic cultures from Madlen Walter and Kurt Zoglauer from the Humboldt Universität zu Berlin.

TABLE 1 | Plant material.

#	Species	Origin[a]	Accession	Plant family
1	*Larix decidua* Mill.	T	50°58'58.0'N 13°23'44.4'E	Pinaceae
2	*Larix kaempferi* (Lamb.) Carrière	G	#1041	
3	*Larix gmelinii* (Rupr.) Kuzen.	G	#868	
4	*Larix sibirica* Ledeb.	G	#1340 (5/18)	
5	*Pseudotsuga menziesii var. viridis* (Schwer.) Franco	T	#1014 Indiv. 2	
6	*Pinus sylvestris* L.	T	[U/1] 504 55	
7	*Picea abies* (L.) H. Karst.	T	[*Pf 1935/35] 217/12	
8	*Abies sibirica* Ledeb.	T	[U/1] 403 a210 Indiv. left	
9	*Taxus baccata* L.	T	[U/7] 401 c19	Taxaceae
10	*Juniperus communis* L.	T	[*2000/117] 837/1684	Cupressaceae
11	*Ginkgo biloba*	T	N 50°58'53.5';E 13°34'27.2'	Ginkgoaceae

[a]*Origin of accession, either from Forest Park Tharandt (T) or Staatsbetrieb Sachsenforst Graupa (G).*

Sequencing, Read Clustering, Repeat Classification, and Characterization of One *L. decidua* and One *L. kaempferi* Individual

For sequencing, we used an individual of *L. decidua* and *L. kaempferi* each, with accessions as indicated in **Table 1**, lines 1–2. Whole genome sequence libraries with 350 to 500 bp fragment sizes have been generated by Macrogen Inc., followed by Illumina paired-end sequencing on Illumina HiSeq2000 and HiSeq2500 machines. The reads were trimmed to the same length (101 bp) using *Trimmomatic* (Bolger et al., 2014). We pre-treated and interlaced the read sequences using the custom scripts accompanying the local *RepeatExplorer 1* installation (*paired_fastq_filtering.R* and *fasta_interlacer*, followed by *SeqClust*). The reads were quality-trimmed to include only sequences with a Phred score ≥10 over 95% of the read length. Overlapping paired ends have been excluded. We randomly selected five million paired reads from each library and subjected those to comparative clustering with the *RepeatExplorer 1* software (Novák et al., 2010, 2013) and *TAREAN* (Novák et al., 2017). The resulting clusters were curated and classified manually, integrating similarity searches against the *Conserved Domain Database* for the functional annotation of proteins (Marchler-Bauer et al., 2011), *RepBase* Update (Jurka et al., 2005), the *REXdb* database (Neumann et al., 2019), as implemented in the local *RepeatExplorer 1* installation, and a custom library containing ribosomal, telomeric, and plastid sequences collected from NCBI, as well as the *PIER* 2.0 transposable element database (Wegrzyn et al., 2013), as downloaded from the *TreeGenes* Web site (Wegrzyn et al., 2008). Each assignment to a repeat type was verified according to different characteristics: For protein-coding transposable elements, most weight was placed on the type of identified protein domain, and that their order and classification were consistent with the transposable element assignment. If one of the repeat databases produced a similar hit, this was seen as further evidence. For non-coding transposable elements, we have relied on the identification of other conserved sequence features, such as long terminal repeats, primer binding sites, or tRNA-derived promoter boxes. Finally, satDNAs were checked by dotplots and by the organization of paired reads to derive tentative candidates. Later, their tandem organization was experimentally verified by Southern hybridization. If we could not unambiguously assign a cluster to a repeat type, we have left it as unassigned. Hence, we consider all our repeat assignments as high-confidence classifications.

Clusters connected by paired reads and sharing a common annotation have been manually combined to superclusters, after making sure that the corresponding repeat classifications were compatible. Graphic representations as bar and pie charts have been produced with *R* using the *ggplot2* library (Wickham, 2016).

Clusters with a satellite-typical star-like and circular graphical representation (Novák et al., 2010) have been selected for further analysis. Putative monomers were manually detected on the *RepeatExplorer*-derived contigs as well as with the software *Tandem Repeats Finder* (Benson, 1999), for which we combined a range of parameters with manual verification. Using the putative circular monomers as a template, we iteratively aligned the paired reads against these template sequences to derive more representative consensus sequences (**Supplementary Data S1**). Repeat sequences have been compared and characterized using multiple sequence alignments and dotplots of monomers with the packages *MAFFT* (Katoh and Standley, 2013) using standard parameters and *FlexiDot* (Seibt et al., 2018) with the parameters -k *18 -S 4 -p 2 -r 0 -x 1*. General sequence investigation (such as the identification of restriction sites, the positioning of primer sequences, and the annotation of smaller, repetitive motifs) was performed with the multi-purpose software *Geneious* 6.1.8 (Kearse et al., 2012).

Repeat Quantification by Comparative Read Mapping

To determine the relative abundance of the selected *Larix* repeat families in other gymnosperms, we complemented our own data (*L. decidua* and *L. kaempferi*) with publicly available whole genome shotgun Illumina reads from twelve gymnosperms (Nystedt et al., 2013; Kuzmin et al., 2019; Zimmermann et al., 2019). The publicly available reads were obtained from NCBI under the following accession numbers: From the Pinaceae, these data sets include other larches (*Larix sibirica*, SRR8555411; *Larix gmelinii*, PRJNA528429), pines (*Pinus taeda*, SRR1054646; *P. sylvestris*, ERR268439; and *P. lambertiana*, SRR2027090), spruces (*Picea abies*, ERR268355; *P. glauca*, SRR1259615; and *P. sitchensis*, SRR3100750), fir (*Abies sibirica*, ERR268418), and Douglas fir (*Pseudotsuga menziesii*, SRR2027118). We also analyzed the genomes of distantly related gymnosperms, such as yew (*Taxus baccata*, ERR268427) and common juniper (*Juniperus communis*, ERR268423). We randomly extracted three million paired reads and iteratively mapped them against the circular satDNA consensus until it remained unchanged. This alignment to the consensus was performed with the *Geneious*

6.1.8 mapping tools (using *medium sensitivity* parameters; Kearse et al., 2012). We graphically represented mapping counts as bubble chart with *R* and *ggplot2* (Wickham, 2016).

Repeat Detection in Genome Assemblies

If we detected EulaSat1 to EulaSat5 presence in additional genomes, we downloaded the corresponding genome assemblies, if available. These included assemblies of *Pseudotsuga menziesii* (Psme v.1.0 from treegenesdb.org, Neale et al., 2017) and *Abies alba* (Abal v.1.1 from treegenesdb.org, Mosca et al., 2019). Using a local *BLAST* search (Altschul et al., 1990), we have retrieved the five scaffolds with the most hits for each of the satDNA consensuses. In order to assess the organization of the satDNA families, we visualized each scaffold as a dotplot. For visualization purposes, we extracted representative 20 kb regions and generated *FlexiDot* dotplots (Seibt et al., 2018) with the parameters -k *18 -S 4 -c n -p 0 -A 1.5 -T 40 -E 16*.

PCR and Cloning

From the monomeric consensus sequences, primer pairs have been designed from the *L. decidua* reference (**Table 2**). For the amplification of satDNA probes for Southern hybridization and FISH, PCR was carried out with the specific primer pairs. PCR reactions with 50 ng plasmid template were performed in 50 µl volume containing 10× DreamTaq buffer and 2.5 units of DreamTaq polymerase (Promega). Standard PCR conditions were 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, primer-specific annealing temperature for 30 s, 72°C for 1 min, and a final incubation time at 72°C for 5 min. The resulting amplicons have been cloned into the pGEM-T vector (Promega), followed by Sanger sequencing. The clones containing inserts most similar to the satDNA consensus have been chosen for hybridization experiments.

Southern Blot Hybridization

For Southern blots, genomic DNA was restricted with enzymes specific for each tandem repeat targeted, separated on 2% agarose gels, and transferred onto Hybond-N+ nylon membranes (GE Healthcare) by alkaline transfer. Hybridizations were performed according to standard protocols using probes labeled with ^{32}P by random priming (Sambrook et al., 1989). Filters were hybridized at 60°C and washed at 60°C for 10 min in 2× SSC/0.1% SDS. Signals were detected by autoradiography.

TABLE 2 | Primer pairs for the generation of satDNA clones.

Primer	Sequence	G/C (%)	Length (bp)	Tm (°C)
EulaSat1_F	GTATGCACATTCTACGTCATAACG	41.7	24	59.3
EulaSat1_R	GAATGCGCAAACTATAGAAAGTCG	41.7	24	59.3
EulaSat2_F	TCAAAGTTGAAAATCGACCGTGC	43.5	23	58.9
EulaSat2_R	ATGTCACATTGGTAGACGAGCG	50.0	22	60.3
EulaSat3_F	GAATTTTTTAGTGTGATTGTTCAGTAG	29.6	27	57.4
EulaSat3_R	GGTCAGAAATGTTAGCATAGTCG	43.5	23	58.9
EulaSat4_F	GGCACAAGCTCAAGGTATAAGC	50.0	22	60.3
EulaSat4_R	ATGGCACAAGATCAAGGAAAGC	45.5	22	58.4
EulaSat5_F	TTCATTCTCGGAGACCTCACG	52.4	21	59.8
EulaSat5_R	GTCCTTAGTGGACAGTTGAGG	52.4	21	59.8

Probe Labeling, Chromosome Preparation, and Fluorescent *in situ* Hybridization

Sequenced satDNA clones have been used as template for PCR-based labeling with biotin-16-dUTP. The probe pZR18S contains a 5,066 bp fragment of the sugar beet 18S rRNA gene within the 35S rDNA (HE578879; Paesold et al., 2012) and was labeled with DY-415 or DY-647-dUTP (Dyomics) by nick translation. The probe pXV1 (Schmidt et al., 1994) for the 5S rRNA gene was labeled with digoxygenin-11-dUTP by nick translation.

We prepared mitotic chromosomes from the meristems of young primary roots, harvested shortly after germination. Prior to fixation in ethanol:chloroform:glacial acetic acid (2:1:1), root tips were incubated either for 16 h in 2 mm 8-hydroxyquinoline or for 1 h in nitrous oxide at 10 bar. Fixed plant material was digested for 0.5 to 1.5 h at 37°C in an enzyme mixture consisting of 2% (w/v) cellulase from *Aspergillus niger* (Sigma C1184), 4% (w/v) cellulase Onozuka R10 (Sigma 16,419), 0.5% (w/v) pectolyase from *Aspergillus japonicus* (Sigma) P-3026, 1% (w/v) cytohelicase from *Helix pomatia* (Sigma) C-8274, 1% hemicellulase from *Aspergillus niger* (Sigma H2125), and 20% (v/v) pectinase from *Aspergillus niger* (Sigma P4716) in citrate buffer (4 mm citric acid and 6 mm sodium citrate). The root tips have been washed and transferred to a slide, before maceration with a needle in 45% glacial acetic acid. Before the slide dried, the chromosomes have been fixed with methanol:glacial acetic acid (3:1).

Prior to FISH, according to the amount of cytoplasm visible under light microscope, we pre-treated the slides with 100 µg/ml RNase in 2× SSC for 30 min, followed by 200 µl of 10 µg/ml pepsin in 10 mm HCl for 15 to 30 min. Slides with abundant cytoplasm were additionally treated for 10 min with proteinase K. FISH was performed according to the protocol of Heslop-Harrison et al. (1991) with modifications as described (Schmidt et al., 1994). Probes were hybridized with a stringency of 76% and subsequently washed with a stringency of 79%. The chromosome preparations were counterstained with DAPI (4', 6'-diamidino-2-phenylindole) and mounted in antifade solution (CitiFluor). Slides were examined with a fluorescence microscope (Zeiss Axioplan 2 imaging) equipped with Zeiss Filter 09 (FITC), Zeiss Filter 15 (Cy3), Zeiss Filter 26 (Cy5), AHF Filter F36-544 (aqua), and Zeiss Filter 02 (DAPI). Images were acquired directly with the Applied Spectral Imaging v. 3.3 software coupled with the high-resolution CCD camera ASI BV300-20A.

RESULTS

L. decidua and *L. kaempferi* Show Very Similar Repeat Profiles

To assess the genome composition of *L. decidua* and *L. kaempferi*, we obtained paired-end Illumina whole genome shotgun

sequences with fragment sizes of 500 bp. Five million reads of each genome have been randomly chosen for comparative low-coverage clustering with *RepeatExplorer*. The software automatically chose 2,124,798 (*L. decidua*) and 2,125,214 (*L. kaempferi*) reads (corresponding to a genome coverage between 1.6 and 1.9%) and yielded estimates of the repetitive fraction of 69.0% for *L. decidua* and 68.1% for *L. kaempferi*. We classified the read clusters according to their repeat class and manually combined clusters connected by read pairs and similar annotation to superclusters. This has led to similar repeat profiles for both genomes (**Figure 1**): In particular, Ty3-*gypsy* retrotransposons (approx. 31% for both genomes) made up the largest fraction, followed by Ty1-*copia* retrotransposons (both approx. 24%). Presentation of the first 214 read superclusters as a two-sided, comparative bar chart illustrates the high degree of genomic similarity between both genomes (**Figure 1**). With only few exceptions, the read clusters are equally abundant in *L. decidua* and *L. kaempferi*, with gaps indicating the absence of the repeat from one of the genomes. We detected most variation in the amount of satellite DNA, with 3.2% for *L. decidua* and 2.0% for *L. kaempferi* (**Figure 1**; marked in red). Read clustering with *TAREAN* delivered similar quantifications for the tandemly repeated sequences.

Larix Tandem Repeats Vary in Abundance and Genome Organization, With Only Punctual Differences Between *L. decidua* and *L. kaempferi*

Six of the analyzed *RepeatExplorer* read clusters produced circular or star-shaped layouts, typical for tandem repeats (**Supplementary Figure S1**), representative of five satDNA families. Using *L. decidua* as reference organism, we extracted sequences of the candidates and named them EulaSat1 to EulaSat5, short for European larch satellite. We refined the monomer consensus sequences by iterative mapping of three million paired reads to generate robust consensus sequences (**Figure 2**; **Supplementary Data S1**), used for quantification and primer generation. In order to verify the consensus sequence and to generate hybridization probes, we amplified and cloned all five candidates from *L. decidua*. SatDNA characteristics are summarized in **Table 3**, whereas a multi-sequence dotplot shows the family and subfamily structure (**Supplementary Figure S2**).

To verify the head-to-tail organization of the five EulaSat repeat families, we transferred restricted genomic *L. decidua* DNA onto Southern membranes. After Southern hybridization of the EulaSat probes, we investigated the resulting autoradiographs for the presence of satDNA-typical ladder hybridization (**Figure 3**), indicating repeat organization in long arrays. In detail, summarizing the computational and molecular data, the five satDNAs are characterized as follows:

1. Comprising approx. 1% of both *Larix* genomes, EulaSat1 is the major satDNA family in larches. First described as LPD, it is an integral part of many *Larix* genomes (Hizume et al., 2002), forming conserved 173 bp monomer with a G/C content of 33%. Six of the eleven enzymes tested produced a satellite-typical restriction ladder for EulaSat1, all supporting the monomer length of 173 bp (**Figure 3A**). Clearest signals up to the tetra- and pentamers have been generated with *Dra*I (lane 4) and *Alu*I (lane 8), whereas the remaining enzymes, such as *Hin*dIII (lane 2), released longer multimers up to the dodecamer. Although the EulaSat1 consensus contains a potential *Hpa*II/*Msp*I restriction site (indicated in **Figure 2**), we detected only high molecular weight signals, indicating a high degree of DNA methylation (**Figure 3A**; lanes 10, 11).
2. With a genomic representation of 0.46 and 0.22%, EulaSat2 is the second-most abundant satDNA family in *L. decidua* and *L. kaempferi*. EulaSat2 has a relatively high G/C content with 44% and consists of monomers with the satDNA-typical length of 148 bp. The EulaSat2 autoradiograph (**Figure 3B**) showed ladder-like patterns for five enzymes. *Fok*I, *Mbo*I, and *Alu*I (lanes 6–8) released the EulaSat2 monomer, supporting its length of 148 bp. *Dra*I (lane 4) only produced weak monomeric signals, whereas *Rsa*I (lane 9) did not generate any monomeric bands, pointing to only weak restriction site conservation. Bands up to the undecamer were released, before falling together in a smear. Hybridization of *Hpa*II/*MspI-restricted* DNA (lanes 10, 11) did not produce any signals below 3 kb.
3. The EulaSat3 family is divided into three subfamilies with similar features: The conserved 345 bp long monomers contain a generally low G/C content between 26 and 31%. Out of all identified repeats, only the three EulaSat3 subfamilies are genome specific, as their clusters contain only reads from *L. decidua* and none from *L. kaempferi*. Consensus sequences of all EulaSat3 subfamilies can be subdivided into a 178 bp and a 167 bp subunit with identities ranging between 45.5 and 48.3% (**Supplementary Figure S3**), suggesting evolution by EulaSat3 reorganization into structures of higher order. EulaSat3 hybridization (**Figure 3C**) generated ladder-like patterns with different intensities in all lanes, with its monomeric length (345 bp) distinguishable in most cases. For two enzymes, *Dra*I (lane 4) and *Alu*I (lane 8), bands below the monomer size were visible. These additional bands can be explained by multiple restriction sites in the monomer (see **Figure 2**), giving rise to 163 and 182 bp fragments (*Dra*I) as well as 36, 176, 196, and 212 bp fragments (*Alu*I). In addition, *Hpa*II and *Msp*I were able to cut EulaSat3, both producing identical, weak ladders (lanes 10–11), pointing to the presence of at least some monomers without DNA methylation in the putative restriction site.
4. Similarly, for EulaSat4, we detected two subfamilies with different monomeric lengths. EulaSat4a has 203 bp monomers and is more abundant, supported by a mapping of 1,104 reads. In contrast, the less frequent EulaSat4b subfamily (supported by 696 reads) has a monomer length of 169 bp. We did not detect clear, canonical ladder patterns after hybridization of EulaSat4 (**Figure 3D**). However, signals as detected for *Bsm*I (lane 3), *Mbo*I (lane 7), and *RsaI* (lane 9) can be explained by the recognition of both EulaSat4 subfamilies by the Southern probe. As observed, a combination

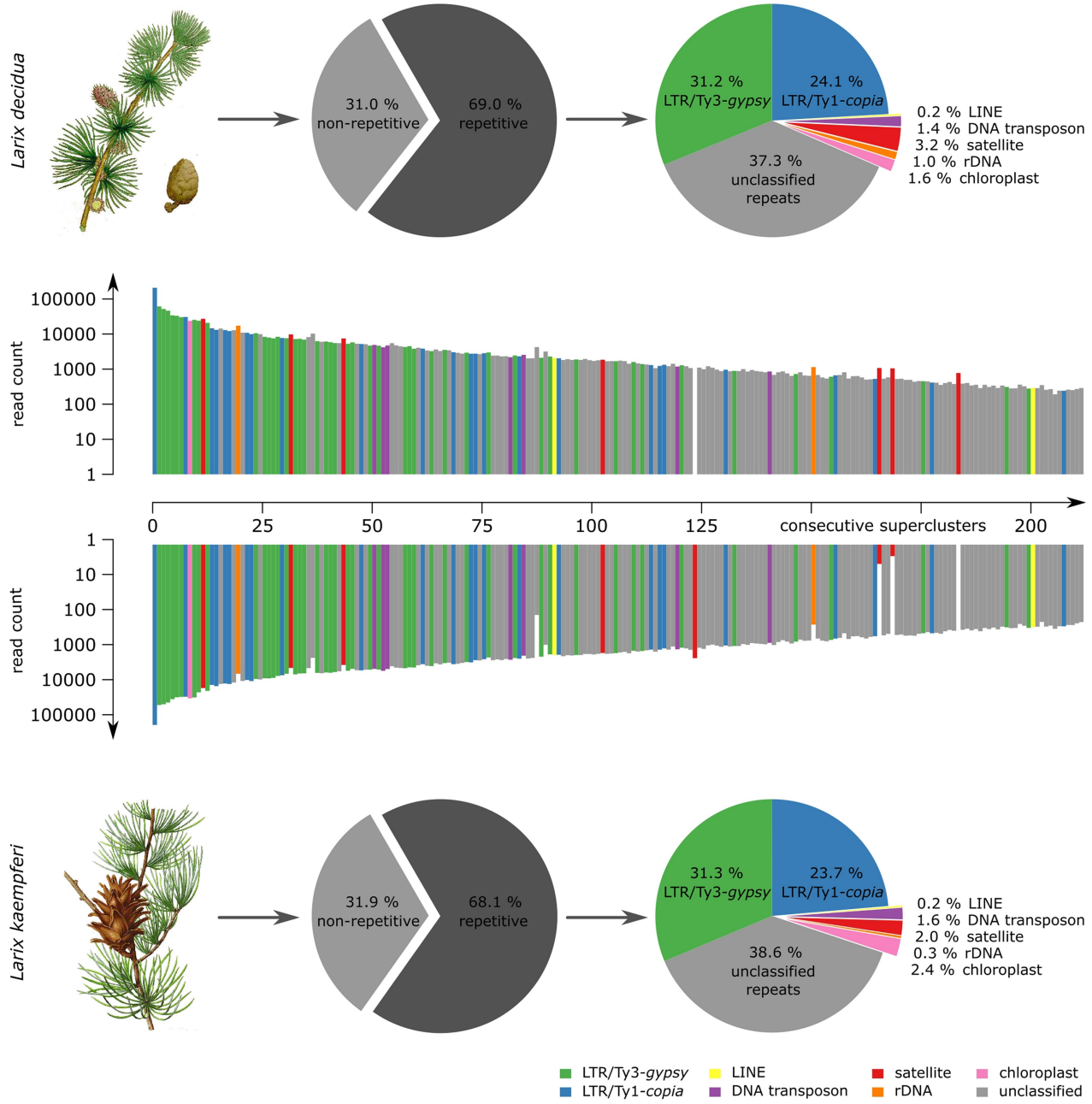

FIGURE 1 | Comparison of repetitive genome fractions reveals high genomic similarity over all repeat types between *Larix decidua* and *Larix kaempferi*. At the center of each figure, a two-sided bar plot shows 214 repeat superclusters with respective read counts in *L. decidua* (top) and *L. kaempferi* (bottom). The read count is presented on a logarithmic scale. The composition of each *Larix* repeat fraction is summarized by pie charts. The plant illustrations are reproduced from Woodville, W., Hooker, W.J., and Spratt, G., Medical Botany, 3th edition, vol. 1: (1832; *L. decidua*) and M. E.-A. Carriere (ed.) Revue Horticole, serié 4, vol. 40: (1868), Paris (*L. kaempferi*).

of 203 and 169 bp fragments leads to the complex ladder patterns with unequal step sizes.

5. Out of all identified satDNA families, EulaSat5 has the shortest monomer (87 bp) and the highest G/C content (50%). Although the monomer is short, this satDNA family makes up 0.35 and 0.18% of the *L. decidua* and the *L. kaempferi* genomes, respectively. EulaSat5 hybridization (**Figure 3E**) yielded ladder patterns for the three enzymes: *AIw*26I, *Fok*I, and *Mbo*I (lanes 5–7). For *Mbo*I, a strong monomeric signal was detected, providing additional support for the monomer size of 87 bp and for the high restriction site conservation within EulaSat5 arrays. Intense signals in the hexa- and heptamer regions indicate arrays with higher order repeat structures. Hybridization of *Hpa*II/*Msp*I-restricted DNA did not reveal bands in the low

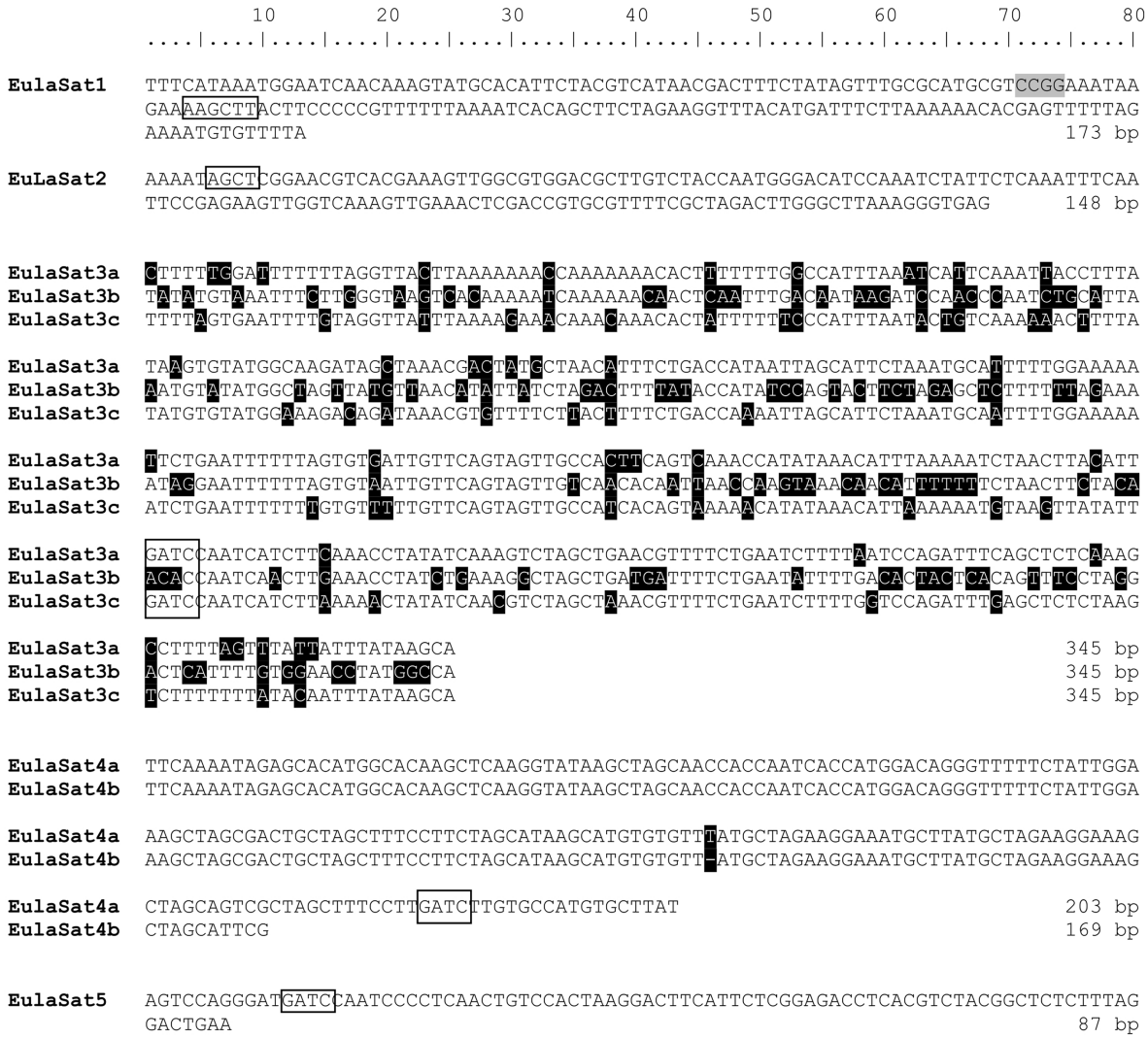

FIGURE 2 | Consensus sequences and subunit structure of the tandem repeat monomers. The monomer consensus sequences of the EulaSat1 to EulaSat5 satellite DNAs are shown. Recognition sites of restriction enzymes used to release the DNA ladder (**Figures 3, 5**) are indicated by rectangles. *Hpa*II/*Msp*I recognition sites are shaded in gray. EulaSat3 and EulaSat4 are divided into the EulaSat3a, EulaSat3b, and EulaSat3c as well as the EulaSat4a and EulaSat4b subfamilies. These sequences are represented as multiple sequence alignment, with ambiguities shaded in black.

molecular weight region, suggesting strong EulaSat5 DNA methylation.

Individual *L. decidua* Chromosomes Show Comparable satDNA Localizations

To determine the position of the satDNA families along *Larix* chromosomes, we prepared mitotic and interphase chromosomes from the *L. decidua* reference and *in situ* hybridized them with biotin-labeled satDNA probes (**Figures 4A–E**):

1. EulaSat1 hybridized to 18 from the 24 *L. decidua* chromosomes, co-localizing with the strongly DAPI-stained heterochromatic proximal bands (**Figure 4A**). EulaSat1's occurrence in the deep heterochromatin was confirmed by co-localization with DAPI-positive regions on interphase nuclei (**Figure 4A**).
2. For EulaSat2, we have observed the presence on all chromosomes. The localization along the centromeric constriction of all chromosomes indicates EulaSat2's suitability to serve as a marker for the centromere (**Figure 4B**). As this position is depleted in DAPI staining, we assume that the EulaSat2 regions are only loosely packaged. At higher resolution, using interphase nuclei, we confirmed that EulaSat2 is largely excluded from the heterochromatin (**Figure 4B**).
3. The three remaining satDNA families, EulaSat3 to EulaSat5, are marked by a dispersed localization along all *L. decidua* chromosomes (**Figures 4C–E**). For EulaSat3, we identified a range of minor signals without exclusion of the centromeres, spread along the chromosomes. At interphases, we noted the EulaSat3 presence in hetero- and euchromatic regions (**Figure 4C**).

TABLE 3 | Characteristics of satDNA in *L. decidua* (*Ld*) and *L. kaempferi* (*Lk*) genomes.

Family	Genome proportion (%)[a]		Monomer length (bp)		G/C content (%)		Mean pairwise identity (%)		Read support[b]	
	Ld	*Lk*	*Ld*	*Lk*	*Ld*	*Lk*	*Ld*	*Lk*	*Ld*	*Lk*
EulaSat1	1.28	0.81	173	173	33	33	91.8	90.3	74,373	46,173
EulaSat2	0.46	0.22	148	148	45	45	84.7	83.9	22,881	10,714
EulaSat3a			345	345	28	28	98.0	–	1,532	0
EulaSat3b	0.05	0.00	345	345	31	31	94.2	–	978	0
EulaSat3c			345	345	26	26	95.2	–	605	0
EulaSat4a	0.09	0.08	203	203	43	43	86.6	85.8	1,104	696
EulaSat4b			169	169	43	43	86.8	86.3	264	119
EulaSat5	0.35	0.18	87	87	50	50	87.1	84.8	4,510	1,398

[a]*RepeatExplorer-based estimate.*
[b]*Number of reads mapping out of three million paired-end reads.*

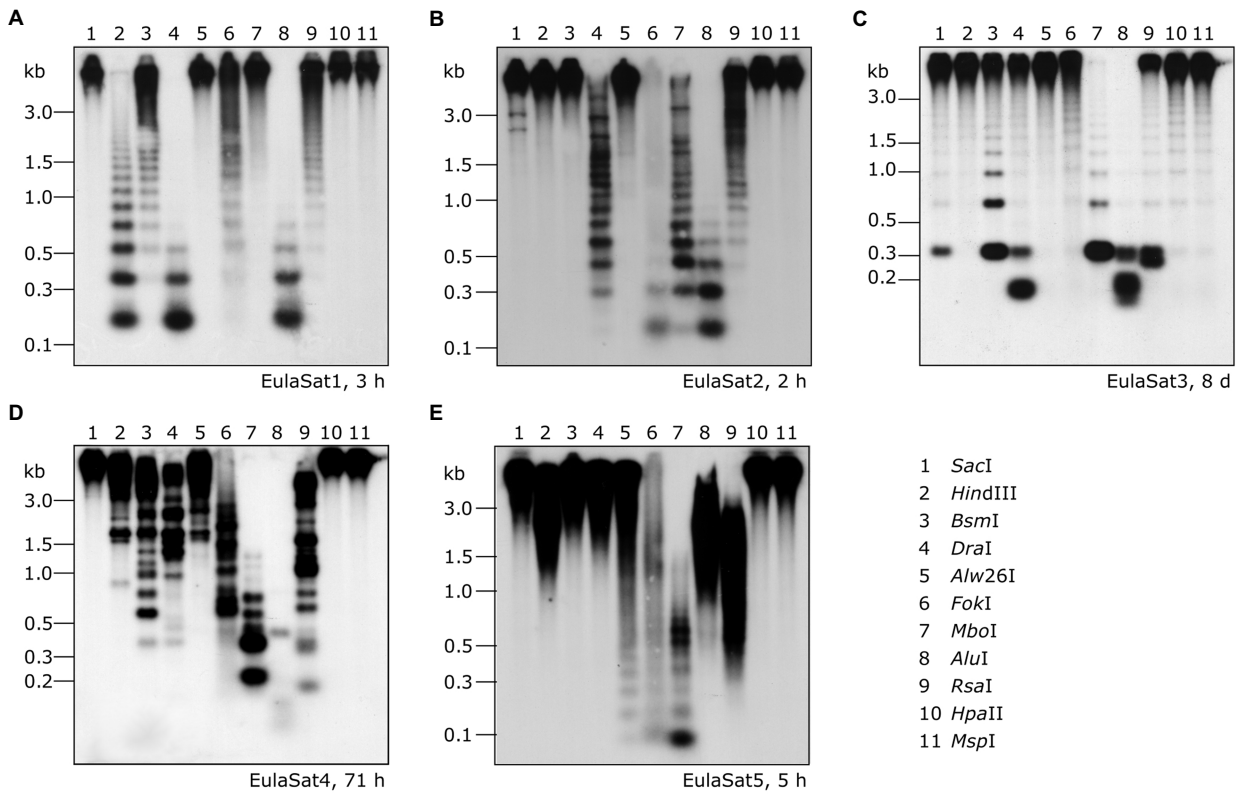

FIGURE 3 | Genomic organization of the *Larix* satDNA families EulaSat1 to EulaSat5. Southern hybridization of restricted genomic *L. decidua* DNA with satDNA probes releases the satellite-typical ladder pattern of EulaSat1 **(A)**, EulaSat2 **(B)**, and EulaSat3 **(C)**, EulaSat4 **(D)**, and EulaSat5 **(E)**. The exposition time is indicated in hours (h) or days (d) for each experiment.

4. This pattern is mirrored for EulaSat4. We found that most of the minor EulaSat4 signals were localized at the intercalary chromosome regions. The distal chromosome regions and the centromeric restrictions were not excluded, but only few chromosomes carried EulaSat4 signals at these regions. At interphases, most signals were localized in the DAPI-positive heterochromatin (**Figure 4D**).
5. EulaSat5 signals were scattered over the whole length of all chromosomes, with frequent enrichments at or near the (peri-) centromeric regions. The signals are often euchromatic, but without exclusion from the DAPI-positive heterochromatin (**Figure 4E**).

Taken together, whereas three of the satDNA probes (EulaSat3 to EulaSat5) are dispersed along all chromosomes, EulaSat1 and EulaSat2 produce distinct signals, limited to the heterochromatic band and the centromeric constriction, and produce clear chromosomal landmarks.

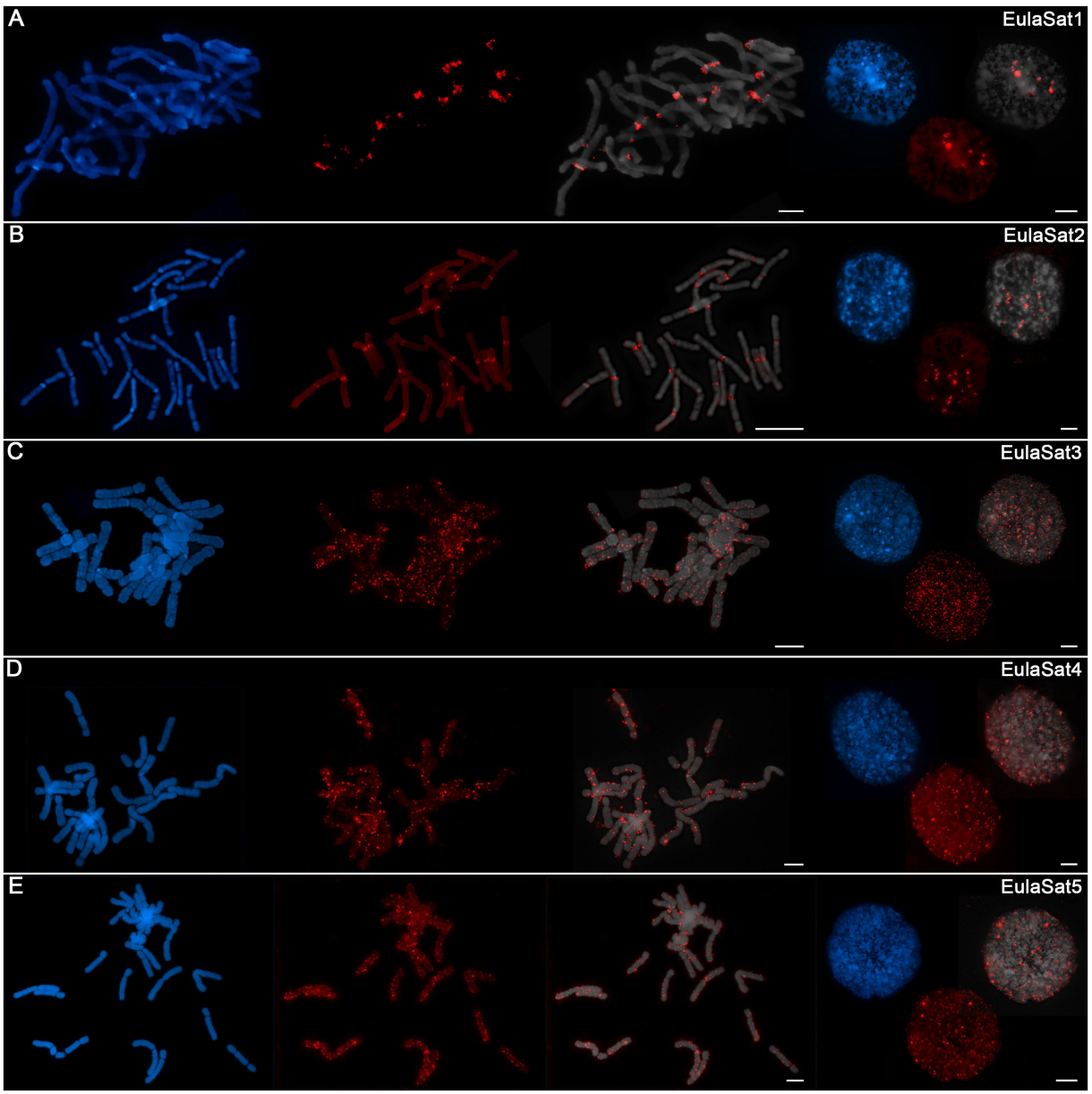

FIGURE 4 | Chromosomal localization of the EulaSat families along *Larix decidua* chromosomes. Chromosomes have been counterstained with DAPI, indicated in blue and gray. Fluorescent *in situ* hybridizations of EulaSat1 **(A)**, EulaSat2 **(B)**, EulaSat3 **(C)**, EulaSat4 **(D)**, and EulaSat5 **(E)** to *L. decidua* meta- and interphases are shown in red.

Distribution, Abundance, and Genomic Organization in Related Conifer Genomes

We used bioinformatics and experimental approaches to investigate the abundance and genomic organization of the EulaSat repeats in related species. Using a read mapping approach, we screened whole genome shotgun Illumina reads of twelve Pinaceae species (**Figure 5**), including four larches, three pines, three spruces, a fir, and a Douglas fir. As outgroups, we also analyzed DNA of more distantly related yew (*Taxus baccata*) and juniper (*Juniperus communis*) trees.

As read mapping may misrepresent the factual genome representation of repeats due to inherent G/C biases (Benjamini and Speed, 2012; Chen et al., 2013), we complemented our bioinformatics approach with an experimental verification. For this, we comparatively hybridized the satDNA probes onto restricted genomic DNA and quantified the repeat abundance in eleven species (**Figure 6**). Our species sampling includes *L. decidua*, *L. kaempferi*, *L. gmelinii*, *L. sibirica* (lanes 1–4), and a single representative of additional gymnosperm genera: *Pseudotsuga menziesii* (lane 5), *Pinus sylvestris* (lane 6), *Picea abies* (lane 7), *Abies sibirica* (lane 8), *Taxus baccata* (lane 9),

Juniperus communis (lane 10), and *Ginkgo biloba* (lane 11). Please note that *L. sibirica* is absent from **Figure 6A**.

Both approaches show that EulaSat1, EulaSat2, EulaSat4, and EulaSat5 are present in all *Larix* accessions analyzed, indicating their wide-spread occurrence throughout the genus (**Figures 5, 6**).

1. EulaSat1 is highly abundant in all *Larix* species examined, but without occurrence outside of the genus (**Figure 5**). Supporting this, EulaSat1 hybridization revealed clear ladder signals in the genus *Larix* for all three of the tested species, already after 25 min of exposition (**Figure 6A**). We observed similar patterns and signal strengths in all *Larix* species tested, indicating similar EulaSat1 monomer sizes with organization in long arrays across the genus. The remaining genomes did not produce any signal, pointing to EulaSat1 absence. Longer exposition time of 3 h revealed no further information.
2. A similar high abundance in *Larix* sp. was detected for EulaSat2. EulaSat2 was also present in *P. menziesii*, but in lower quantity (**Figure 5**). After EulaSat2 hybridization, clear ladder-like pattern is visible for all larch species tested (**Figure 6B**), supporting the organization of similar-sized monomers in a tandem arrangement. In addition, for *P. menziesii* (lane 5), very weak signals corresponding to the dimer and trimer are distinguishable, becoming more prominent after longer exposure (not shown), without additional signals in any other lanes. Hybridization to *L. gmelinii* DNA (lane 3) does only produce faint monomeric and dimeric bands, and instead leads to many signals in the higher, multimeric region. As the *L. gmelinii* DNA was restricted completely, this indicates a less conserved *Alu*I restriction site in the EulaSat2 satDNA.
3. Computationally, the three EulaSat3 subfamilies have been analyzed individually, indicating considerable genomic

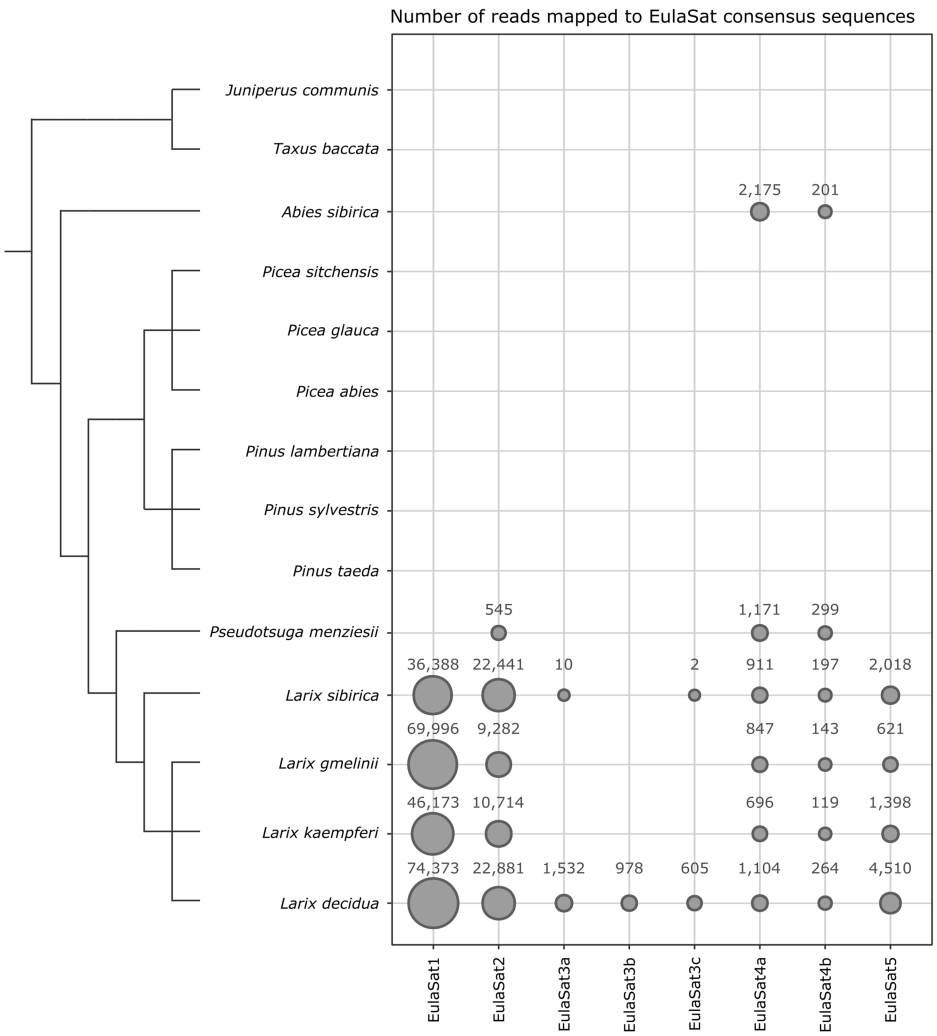

FIGURE 5 | Distribution of the EulaSat tandem repeats in fourteen gymnosperm genomes surveyed by read mapping. The area of each bubble represents the amount of whole genome shotgun Illumina reads mapping to the EulaSat consensus sequences. A total of three million paired reads have been used as input for the mapping analysis. The dendrogram indicates the evolutionary relationship between the species according to Wei and Wang (2003) for the genus *Larix* and Lu et al. (2014) for the overall phylogeny. The branch lengths are not to scale.

impact only in *L. decidua*. We did not detect the presence in *L. kaempferi* and *L. gmelinii*. For subfamilies EulaSat3a and EulaSat3c, only few *L. sibirica* hits mapped to the consensus, suggesting a reduced abundance in this genome. The other gymnosperm sequences tested did not contain any similarity to the EulaSat3 subfamilies (**Figure 5**). The patchy distribution across the *Larix* genus was also apparent experimentally (**Figure 6C**), with hybridization revealing exclusive signals in *L. decidua* and *L. sibirica*. In both species, the monomeric band constituted the strongest signal, suggesting the high conservation of the *Mbo*I restriction site within EulaSat3. In *L. decidua*, the satDNA-typical ladder pattern was formed, whereas in *L. sibirica*, the multimeric bands were absent. As the signals were still faint after 17 days of exposure, we conclude a relatively low abundance in both genomes.

4. Out of all satDNAs analyzed, the EulaSat4a and EulaSat4b subfamilies had the broadest distribution. Apart from their presence in the *Larix* genomes, they also populate *P. menziesii* and *A. sibirica* genomes. In all six EulaSat4-containing genomes, EulaSat4a has been more abundant than EulaSat4b (**Figure 5**). Corroborating this, the corresponding autoradiograph showed signals in species of the *Larix*, *Pseudotsuga*, and *Abies* genera (**Figure 6D**; lanes 1–5, 8). The remaining Pinaceae species (*Pinus sylvestris* and *Picea abies*) did not carry any signals, with longer exposition time (7 days) not changing this result. Hybridization to the larches produced very similar patterns, pointing to similar genomic organization. In *A. sibirica* (lane 8), the lowest band represents a double signal, presumably generated by conserved *Mbo*I restriction sites in the two EulaSat4 subrepeats (169 and 203 bp). However, hybridization to *P. menziesii* (lane 5) produced a stronger ladder with bands slightly shifted toward lower molecular weights, suggesting a small deletion within the EulaSat4 monomers in this species.
5. Read mappings indicate EulaSat5 restriction to *Larix* genomes, with highest abundance in *L. decidua* (**Figure 5**). However, the corresponding probe hybridized to the species of the *Larix* and the *Pseudotsuga* genera (**Figure 6E**; lanes 1–5). Signal patterns of the larch species tested resemble each other, with a relatively strong monomeric band, a fainter dimeric band, and a smear at a higher molecular weight. In *P. menziesii* (lane 5), the smear was overlaid by a very faint band at approximately 480 bp, indicating low abundance. Longer exposition (6 days) of the autoradiograph did not reveal EulaSat5 in further species.

Experimental and computational approaches revealed that two satDNA families also occurred outside of the *Larix* genus, i.e., in *Pseudotsuga* (EulaSat2 and EulaSat4) and *Abies* (EulaSat4). For both genera, genome assemblies were made publicly available,

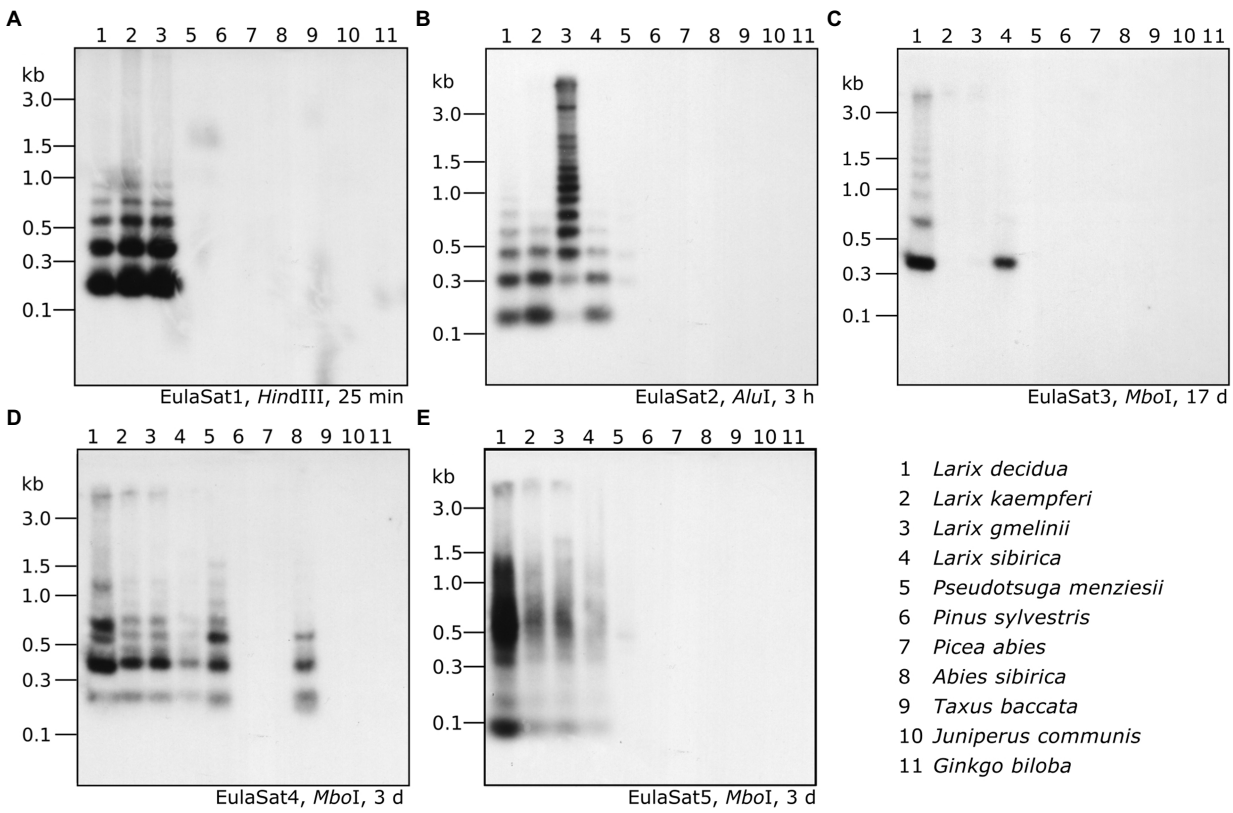

FIGURE 6 | Organization and abundance of EulaSat repeats in related gymnosperm genomes. Genomic DNA of eleven gymnosperms has been restricted as indicated in each panel and was analyzed by comparative Southern hybridization of EulaSat1 **(A)**, EulaSat2 **(B)**, EulaSat3 **(C)**, EulaSat4 **(D)**, and EulaSat5 **(E)**. Exposure times ranged between 25 min and seventeen days, as indicated below the autoradiographs.

namely, assemblies of *P. menziesii* and of *A. alba* (Neale et al., 2017; Mosca et al., 2019). Hence, we have broadened our species panel to include these datasets into our analysis: We queried both assemblies with all satDNA consensuses and deeply inspected the five scaffolds with the most satDNA hits in the genome assemblies of *P. menziesii* and *A. alba*:

In *P. menziesii*, we extracted long EulaSat2 arrays spanning scaffolds over a megabase, with and without higher order arrangements (**Supplementary Figure S4A**). This indicates that the EulaSat2 family, though less abundant (**Figures 5, 6B**), still plays a major role in this genome.

For EulaSat4, in *P. menziesii*, we detected some arrays over 20 kb, often interrupted by other repeats (**Supplementary Figure S4B**). The arrays included variable monomers and different homogenization with or without higher order. In *A. alba*, longer arrays have been detected more frequently. Strikingly, we noticed less monomer variation, with stronger homogenization and a higher abundance of EulaSat4a than EulaSat4b (**Supplementary Figure S4C**).

Taken together, our three approaches (read mapping, analyses of genome assemblies, and experimental quantification) corroborate the different abundances of the five satDNA repeats in the gymnosperms. We confirmed the presence of EulaSat1, EulaSat2, EulaSat4, and EulaSat5 in all *Larix* genomes tested. EulaSat2 and EulaSat4 reside also in more distantly related Pinaceae genomes. In contrast, experimental and bioinformatic evidence supports the young age of EulaSat3 that is restricted to Siberian and European larches.

Only Very Few Differences Distinguish the Chromosomes of *L. decidua* From Those of *L. kaempferi*

We aimed to combine the information gained from *in situ* hybridization to *L. decidua* chromosomes (**Figure 4**) as well as from the quantitative comparisons of conifer genomes (**Figures 5, 6**). We now asked how *L. decidua* and *L. kaempferi* genomes differ on a chromosomal scale and if this information can be used to determine the parentage of individual chromosomes in hybrids.

Therefore, we have comparatively hybridized the most promising tandem repeat landmark probes onto metaphases of both larch species (**Figures 7A–D**), including also the 5S and 35S rDNA probes and the satDNAs EulaSat1 and EulaSat2.

To check how the rDNA tandem repeat loci compare, we investigated the localization of the 5S and 35S rDNAs (**Figures 7A,B**). Both species harbor two 5S rDNA sites (magenta), located distally at the chromosome arms. For the 35S rDNA, we observed hybridization on three chromosome pairs for *L. decidua* and two pairs for *L. kaempferi* (green), all localized at the secondary constrictions of the chromosomes (**Figures 7A,B**).

Regarding the EulaSat1 and EulaSat2 satDNA families, a comparative hybridization onto *L. decidua* and *L. kaempferi* metaphases showed that the satDNA arrays bordered for both species, but with limited co-localization (**Figures 7C,D**). Overall, the comparison between the major satDNAs EulaSat1 and EulaSat2 yielded only very few differences between both species.

We then shifted attention to the genome-specific, but dispersed EulaSat3 satDNA family that may be used to discern the parentage of individual chromosomes in hybrids. For this, we have prepared metaphases from *Larix* × *eurolepis*, a hybrid between *L. decidua* and *L. kaempferi* (**Figure 7E**). Hybridization of the 5S (magenta) and 35S rDNAs (green) have yielded two and five signals, respectively, with the uneven 35S rDNA site number being a testimony to the hybrid status of the individual. The EulaSat3 hybridization yields chromosomes with dispersed EulaSat3 hybridization, indicating *L. decidua* heritage, as well as chromosomes without signals, pointing to descendance from *L. kaempferi*. Nevertheless, due to the dispersed pattern, theEulaSat3 satellite can only give clear parental information for few chromosomes and should be complemented by additional markers, if any become available.

Summarizing, genomes and chromosomes of European and Japanese larches are very similar, with only very few hallmark differences. These include the number in rDNA sites and the genome-specific satDNA family EulaSat3.

DISCUSSION

Similar Repeat Profiles in European and Japanese Larch Genomes Likely Result From Repeat Accumulation and Reduced Turnover

Large conifer genomes evolve only slowly and keep many of their genomic repeats buried within the genomes. With only limited downsizing, we hypothesized that two closely related conifer genomes (such as those from European and Japanese larches) may not accumulate many changes in their overall repeat landscapes. To test this, we have investigated the repeat profiles of these related larch genomes, starting with a broad repeat comparison and then focusing on the repeat class with the fastest sequence turnover, the satDNAs. As a result, our study provides a first comparative overview of the repeat content in two larch species (*L. decidua* and *L. kaempferi*). It also surveys the satDNAs abundances in genomes of related conifer taxa to draw first evolutionary conclusions. Nevertheless, as only a single individual has been analyzed for all examined species, our results cannot account for intra-species variation and any evolutionary satDNA dynamics that may take place on a population genomics scale. Instead, our data allow drawing more general conclusions about satDNA evolution between related species, especially in regard to presence/absence patterns.

We have applied short read sequencing followed by read clustering to efficiently gain insights into both genomes' satDNA contents (as laid out by Weiss-Schneeweiss et al., 2015; Novák et al., 2017). This approach has been successfully used to characterize the repeat landscapes of many non-model plant species as for example beans, various grasses, camellias, crocuses, quinoa, and ferns (Cai et al., 2014; Heitkam et al., 2015; Ávila Robledillo et al., 2018; Kirov et al., 2018; Liu et al., 2019; Schmidt et al., 2019; Heitkam et al., 2020; Ribeiro et al., 2020),

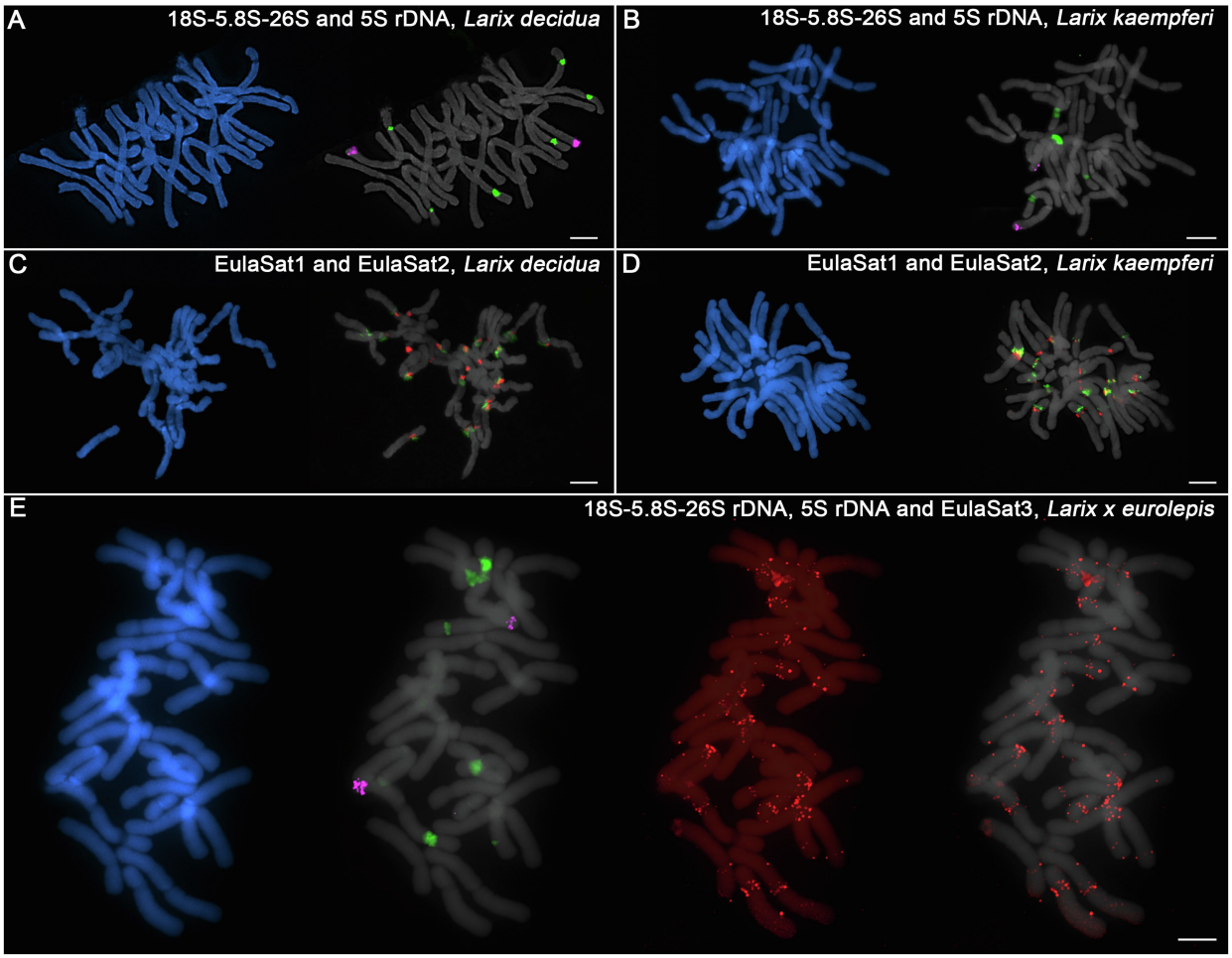

FIGURE 7 | Chromosomal location of rDNAs and the EulaSat families for comparison of *L. kaempferi*, *L. decidua*, and *L.* × *eurolepis*. The chromosomes have been counterstained with DAPI, indicated in blue and gray. Reproduced are fluorescent *in situ* hybridizations of the 5S (magenta) and 18S-5.8S-26S (35S) rDNAs (green) to metaphases of *L. decidua* **(A)** and *L. kaempferi* **(B)**. EulaSat1 (green) and EulaSat2 (red) were comparatively hybridized along metaphase chromosomes of *L. decidua* **(C)** and *L. kaempferi* **(D)**. The genome-specific EulaSat3 family (red) was hybridized along chromosomes of the interspecific *L.* × *eurolepis* hybrid, along with probes for the 5S (magenta) and 35S rDNAs (green; **E**).

and also of non-model animals, such as locusts, grasshoppers, or fishes (Ruiz-Ruano et al., 2016; Ferretti et al., 2020; Boštjančić et al., 2021). For larch genomes, we provided evidence that LTR retrotransposons and derived fragments are their main components, well in line with reports for the related pines and spruces (Kamm et al., 1996; Kossack and Kinlaw, 1999; Nystedt et al., 2013; Stevens et al., 2016; Voronova et al., 2017; Perera et al., 2018).

As only highly repetitive sequences ≥90% are considered in the *RepeatExplorer* cluster analysis, the size estimations of *Larix* repeat fractions (approximately 68% of the analyzed genomes) are bound to be vast underrepresentations, excluding the more fragmented repeats. Especially in large genomes, such as those of the conifers analyzed here, fragmentation and slow repeat divergence lead to barely recognizable transposable elements (TEs), often termed "dark matter" (Maumus and Quesneville, 2016). With increasing genome sizes, these dark matter repeats accumulate, leading to the observed and potentially misleading low repeat fraction estimates, as also recently highlighted by Novák et al. (2020).

For *L. decidua* and *L. kaempferi*, we find overall strikingly similar repeat profiles, especially regarding the TE content, without major differences between European and Japanese larches. The similarities include both, repeat family and abundance. In line with evidence from other conifers (Prunier et al., 2016), these results also suggest the limited TE elimination, usually carried out by recombination, reshuffling, fragmentation, or removal as aftermath to genomic rearrangements (Ma et al., 2004; Ren et al., 2018; Kögler et al., 2020; Maiwald et al., 2021; Schmidt et al., 2021). Along the same lines, we did not observe any transpositional bursts of amplification during the speciation of the larches. Thus, only limited TE-induced genomic novelty has likely occurred in the larches' accumulative genome landscapes.

As a side note, apart from transposable elements and tandem repeats, a small repeat fraction corresponds to organellar DNA. Whether these sequences originated from nuclear integrations

or from plastids and mitochondria cannot be assessed. Nevertheless, as this fraction is low in copy number (2%) and relatively similar to both genomes, we have decided to keep these sequences listed.

Young and Old satDNAs Contribute Both to Genomic Novelty in Larch Genomes

We asked whether the conifer's genomic background of transposable element accumulation and fragmentation has impacted the evolution of satDNAs. These usually evolve by continued rounds of mutation and fixation leading to relatively fast sequence turnovers, even at structurally important chromosomal locations, such as the centromeric regions (Dawe, 2005; Plohl et al., 2012). Indeed, the satDNA abundances in the analyzed larch genomes differed much more than the respective TE portions: We estimated the total satDNA content of *L. decidua* to 3.2% and of *L. kaempferi* to 2.0%, corresponding to satDNA amounts of approximately 416 and 220 Mb, respectively. Overall, the larch satDNA proportions are of the same order of magnitude as already estimated from BACs and fosmids for the related pines (1%; Wegrzyn et al., 2013; Neale and Wheeler, 2019b). However, compared to the large values the satDNA genome fraction can occupy in angiosperms (up to 36%; Ambrozová et al., 2011), the relative amount for larches is rather low, though also not uncommon for angiosperm genomes (Garrido-Ramos, 2017).

To better understand the contribution of satDNA to the genomic differences in larches, we investigated five satDNA families in detail. Here, we will discuss their evolutionary trajectories ranging from the evolutionarily oldest families occurring in several conifer genera (EulaSat2 and EulaSat4), over to those distributed only in the genus *Larix* (EulaSat1 and EulaSat5), to the species-specific family EulaSat3.

The most widely distributed satDNA identified is EulaSat4, with occurrences in larches, Douglas fir and Siberian fir. Interestingly, the comparative read mappings, Southern hybridizations, and analyses of available genome assemblies point to longer and more homogenized EulaSat4 arrays in common and Douglas firs than those observed in larches. We therefore think that EulaSat4 is an evolutionarily old repeat, probably playing a larger role in the *Abies* and *Pseudotsuga* genomes. EulaSat4's patchy distribution across the conifers is an example of a satDNA family's occurrence that is incongruent with the species phylogeny. The satellite library hypothesis may explain this pattern, by assuming that a common set of satDNAs resides in genomes in low copy numbers (Fry and Salser, 1977; Utsunomia et al., 2017; Palacios-Gimenez et al., 2020). Different satDNA amplification would then lead to the observed patchy abundance pattern of EulaSat4. These low copy satDNAs may reside within transposable elements, possibly using these for their conservation and amplification (McGurk and Barbash, 2018; Belyayev et al., 2020; Vondrak et al., 2020). EulaSat4's dispersed localization along all *L. decidua* chromosomes as well as the complex *RepeatExplorer* cluster graphs may indicate such a retrotransposon association. As retrotransposons are strongly conserved across conifer species (Zuccolo et al., 2015; this report), it is likely that EulaSat4 has been retained within a transposable element, followed by patchy amplification in *Larix*, *Pseudotsuga*, and *Abies* species.

The EulaSat2 family co-localizes with the primary constriction of the *L. decidua* and *L. kaempferi* chromosomes, indicating a possible role in centromere formation. Although some plants have centromeres that differ fundamentally from each other (Gong et al., 2012), that does not seem to be the case for larches. The centromeres of all chromosomes harbor EulaSat2, indicating similar sequences and structures. The 148 bp monomers of EulaSat2 are well in line with lengths observed for other centromeric satDNAs, such as that of rice (Zhang et al., 2013), and a bit shorter than the canonical ~170 bp monomers of the mammalian alpha satellite (Willard and Waye, 1987). EulaSat2 is more abundant in *L. decidua* than in *L. kaempferi*, indicating recent array size fluctuations. Nevertheless, EulaSat2 is evolutionarily older with presence in the related Douglas fir, but absence from the more distantly related pine, spruce, and fir species tested. In fact, centromeric satDNAs of related spruces have already been characterized and differ strongly from Eulasat2 in sequence and monomer length (305 bp; Sarri et al., 2008, 2011).

In all larches analyzed, the most abundant satDNA is EulaSat1, also known as LPD (Hizume et al., 2002). Its canonical monomer length of 173 bp is similar in all *Larix* species analyzed, but was not detected outside the genus. It is generally assumed that the most abundant satDNA localizes at the centromeres (Melters et al., 2013); however, some exceptions have been already reported, e.g., for camellias (Heitkam et al., 2015). Instead of the expected centromeric locations, EulaSat1 constitutes the highly heterochromatic, DAPI-positive band present on most of its 24 chromosomes. Regarding EulaSat1's evolution, our data indicate strong EulaSat1 amplification after the split from *Pseudotsuga*. Interestingly, the different species set tested by Hizume et al. (2002) indicates also a patchy abundance in some *Picea*, *Pinus*, *Abies*, and *Tsuga* species – claims that we cannot verify with our data. Nevertheless, we can convincingly show that differences in abundance between *L. decidua* and *L. kaempferi* point to EulaSat1 array expansions and reductions during the more recent evolutionary events.

In contrast to EulaSat1 and EulaSat2, only short arrays were detected for EulaSat5; the second satDNA family restricted to the larches. *In situ* hybridization marked a scattered localization along all chromosomes, typical for short satDNA arrays. As with EulaSat4, an explanation for the short arrays may be an association with transposable elements. We have observed partially mixed *RepeatExplorer* clusters that may point toward an embedment within retrotransposons and also often observed for short satDNA arrays (Meštrović et al., 2015; Satović et al., 2016; Belyayev et al., 2020; Sultana et al., 2020). Similarly, concatenated TEs or part from TEs may have satDNA-like properties, but tend to occur dispersedly along chromosomes (Vondrak et al., 2020; Maiwald et al., 2021). Here, our data are not sufficient to conclusively resolve the large-scale organization of EulaSat5 in larch genomes.

In contrast to all other families investigated, EulaSat3 has experienced a very recent birth, indicating an evolutionarily

young age. EulaSat3 is clearly absent from *L. kaempferi*, but occurs in *L. decidua* with three subfamilies, all containing distinct 345 bp monomers. Their arrangement in higher order is detectable by close inspection of the monomer consensuses and the autoradiograms after Southern hybridization, indicating still ongoing homogenization. FISH and hybridization to *Hpa*II/*Msp*I-restricted genomic DNA have indicated that at least some EulaSat3 monomers are embedded in euchromatic regions. We speculate that these genomic regions are still actively restructured and recombined, processes that potentially restrict the EulaSat3 array size. Taken together, EulaSat3 has developed recently and presumably marks the evolutionarily young regions of *L. decidua* genomes.

To investigate whether EulaSat3 can be applied as a chromosome-specific marker of *L. decidua* parentage in hybrid offspring, we have tested, if chromosome regions from *L. decidua* can be identified in *L.* × *eurolepis* hybrids between *L. kaempferi* and *L. decidua*. Although the *in situ* hybridization clearly marks some chromosome regions as derived from *L. decidua*, this method is not as useful as hoped for the clear differentiation of parentally derived regions along larch hybrid chromosomes.

Nevertheless, EulaSat3's genome specificity within the *Larix* genus as well as the differences in abundance for many of the remaining satDNAs indicates that even large, highly repetitive genomes with slow sequence turnovers can yield new, evolutionarily young repeats and generate sequence innovation to further genome evolution. Whether these repeats also carry a phylogenetic signal and may be used for taxonomic means (e.g., as suggested by Dodsworth et al., 2014) is still open. Along the same lines, the analysis of more individuals and genotypes may advance our understanding of intra-species variation and the evolutionary satDNA dynamics that may take place within a population.

Conclusion

As conifers largely accumulate transposable elements with only reduced active removal processes, their genomes become huge, loaded with many fragmented, barely recognizable repeat copies. As a result, we believe that closely related conifers harbor very similar repeat landscapes. We have tested this hypothesis for two larch species and detected highly similar TE profiles as well as very few differences in their tandem repeat compositions. Nevertheless, despite the high overall repeat similarity, we detected EulaSat3, a satDNA family present in European larches, but absent from their Japanese counterparts. This illustrates that repeat-driven genome innovation still plays a role, even in the huge, repetitive, and fragmented conifer genomes.

DATA AVAILABILITY STATEMENT

Whole genome shotgun Illumina sequences have been deposited at the European Bioinformatics Institute (EBI)'s short read archive in the project PRJEB42507 (http://www.ebi.ac.uk/ena/data/view/PRJEB42507). Cloned sequences used as Southern and FISH probes have been deposited at EBI with the accession numbers LR994496 to LR994500. Consensus sequences of the satDNA families analyzed are included in the **Supplementary Materials** (**Supplementary Data S1**).

AUTHOR CONTRIBUTIONS

TH and TS coordinated the project and interpreted the results. TH, LS, and BW wrote the manuscript, and all authors contributed, proof-read, and edited. TH and AK coordinated the genome sequencing process. TH performed the bioinformatics analysis and contributed to probe preparation. LS and BW performed Southern experiments. LS, SL, and SB prepared metaphase spreads and performed FISH. KM, MB, UT, HW, and DK selected and provided plant material, and guided the project intellectually. All authors contributed to the article and approved the submitted version.

FUNDING

We gratefully acknowledge funding by the German Federal Ministry of Food and Agriculture (Fachagentur Nachwachsende Rohstoffe e. V.; grants 22031714 and 22002216) to TS, DK, and HW.

ACKNOWLEDGMENTS

We remember the late Thomas Schmidt, who has initiated the comparative analyses of larch genomes and has made the ongoing cooperation between three of the partner institutes listed here possible. We also keep in mind the late Ines Walter who has established cytogenetics for larches in our laboratory. As a tribute, we have included her very first, serendipitous FISH on larches (**Figure 7A**).

We thank the Forest Botanical Garden of Tharandt, the Staatsbetrieb Sachsenforst, and Madlen Walther of the Humboldt Universität zu Berlin for the contribution of plant material. We thank Ulrike Herzschuh (Alfred Wegener Institute Potsdam) for providing pre-publication access to sequencing data of *Larix gmelinii*, and we acknowledge Stefan Wanke for supporting our research initiative. Furthermore, we acknowledge the TU Dresden Center for Information Services and High Performance Computing (ZIH) for computer time allocations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at https://www.frontiersin.org/articles/10.3389/fgene.2021.683668/full#supplementary-material

REFERENCES

Ahuja, M. R., and Neale, D. B. (2005). Evolution of genome size in conifers. *Silvae Genet.* 54, 126–137. doi: 10.1515/sg-2005-0020

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

Ambrozová, K., Mandáková, T., Bures, P., Neumann, P., Leitch, I. J., Koblízková, A., et al. (2011). Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. *Ann. Bot.* 107, 255–268. doi: 10.1093/aob/mcq235

Ávila Robledillo, L., Koblížková, A., Novák, P., Böttinger, K., Vrbová, I., Neumann, P., et al. (2018). Satellite DNA in *Vicia faba* is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. *Sci. Rep.* 8:5838. doi: 10.1038/s41598-018-24196-3

Belyayev, A., Josefiová, J., Jandová, M., Mahelka, V., Krak, K., and Mandák, B. (2020). Transposons and satellite DNA: on the origin of the major satellite DNA family in the *Chenopodium* genome. *Mob. DNA* 11:20. doi: 10.1186/s13100-020-00219-7

Benjamini, Y., and Speed, T. P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 40:e72. doi: 10.1093/nar/gks001

Bennett, M. D., and Leitch, I. J. (2019). Plant DNA C-values database (release 7.1, Apr 2019). Available at: https://cvalues.science.kew.org/ (Accessed March 2021).

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

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

Boštjančić, L. L., Bonassin, L., Anušić, L., Lovrenčić, L., Besendorfer, V., Maguire, I., et al. (2021). The *Pontastacus leptodactylus* (Astacidae) repeatome provides insigh0t into genome evolution and reveals remarkable diversity of satellite DNA. *Front. Genet.* 11:1820. doi: 10.3389/fgene.2020.611745

Cai, Z., Liu, H., He, Q., Pu, M., Chen, J., Lai, J., et al. (2014). Differential genome evolution and speciation of *Coix lacryma-jobi* L. and *Coix aquatica* Roxb. Hybrid Guangxi revealed by repetitive sequence analysis and fine karyotyping. *BMC Genomics* 15:1025. doi: 10.1186/1471-2164-15-1025

Chen, Y. C., Liu, T., Yu, C. H., Chiang, T. Y., and Hwang, C. C. (2013). Effects of GC bias in next-generation-sequencing data on *de novo* genome assembly. *PLoS One* 8:e62856. doi: 10.1371/journal.pone.0084522

Dawe, R. K. (2005). Centromere renewal and replacement in the plant kingdom. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11573–11574. doi: 10.1073/pnas.0505100102

Dodsworth, S., Chase, M. W., Kelly, L. J., Leitch, I. J., Macas, J., Novák, P., et al. (2014). Genomic repeat abundances contain phylogenetic signal. *Syst. Biol.* 64, 112–126. doi: 10.1093/sysbio/syu080

Ferretti, A. B. S. M., Milani, D., Palacios-Gimenez, O. M., Ruiz-Ruano, F. J., and Cabral-de-Mello, D. C. (2020). High dynamism for neo-sex chromosomes: satellite DNAs reveal complex evolution in a grasshopper. *Heredity* 125, 124–137. doi: 10.1038/s41437-020-0327-7

Fry, K., and Salser, W. (1977). Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* 12, 1069–1084. doi: 10.1016/0092-8674(77)90170-2

Garcia, S., and Kovařík, A. (2013). Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. *Heredity* 111, 23–33. doi: 10.1038/hdy.2013.11

Garrido-Ramos, A. M. (2017). Satellite DNA: an evolving topic. *Genes* 8:230. doi: 10.3390/genes8090230

Gong, Z., Wu, Y., Koblizkova, A., Torres, G. A., Wang, K., Iovene, M., et al. (2012). Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell* 24, 3559–3574. doi: 10.1105/tpc.112.100511

Heitkam, T., Petrasch, S., Zakrzewski, F., Kögler, A., Wenke, T., Wanke, S., et al. (2015). Next-generation sequencing reveals differentially amplified tandem repeats as a major genome component of northern Europe's oldest *Camellia japonica*. *Chromosom. Res.* 23, 791–806. doi: 10.1007/s10577-015-9500-x

Heitkam, T., Weber, B., Walter, I., Liedtke, S., Ost, C., and Schmidt, T. (2020). Satellite DNA landscapes after allotetraploidisation of quinoa (*Chenopodium quinoa*) reveal unique A and B subgenomes. *Plant J.* 103, 32–52. doi: 10.1111/tpj.14705

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., Schwarzacher, T., Anamthawat-Jónsson, K., Leitch, A. R., Shi, M., and Leitch, I. J. (1991). *In-situ* hybridization with automated chromosome denaturation. *Technique* 3, 109–116.

Hidalgo, O., Pellicer, J., Christenhusz, M., Schneider, H., Leitch, A. R., and Leitch, I. J. (2017). Is there an upper limit to genome size? *Trends Plant Sci.* 22, 567–573. doi: 10.1016/j.tplants.2017.04.005

Hizume, M., Shibata, F., Matsumoto, A., Maruyama, Y., Hayashi, E., Kondo, T., et al. (2002). Tandem repeat DNA localizing on the proximal DAPI bands of chromosomes in *Larix*, Pinaceae. *Genome* 45, 777–783. doi: 10.1139/g02-041

Hizume, M., Tominaga, H. H., Kondo, K., Gu, Z., and Yue, Z. (1993). Fluorescent chromosome banding in six taxa of Eurasian *Larix*, Pinaceae. *La Kromosomo II* 69, 2342–2354.

Jagannathan, M., Cummings, R., and Yamashita, Y. M. (2018). A conserved function for pericentromeric satellite DNA. *elife* 7:e34122. doi: 10.7554/eLife.34122

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467. doi: 10.1159/000084979

Kamm, A., Doudrick, R. L., Heslop-Harrison, J. S., and Schmidt, T. (1996). The genomic and physical organization of Ty1-*copia*-like sequences as a component of large genomes in *Pinus elliottii* var. *elliottii* and other gymnosperms. *Proc. Natl. Acad. Sci U. S. A.* 93, 2708–2713. doi: 10.1073/pnas.93.7.2708

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

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

Kirov, I. V., Gilyok, M., Knyazev, A., and Fesenko, I. (2018). Pilot satellitome analysis of the model plant, *Physcomitrella patens*, revealed a transcribed and high-copy IGS related tandem repeat. *Comp. Cytogenet.* 12, 493–513. doi: 10.3897/CompCytogen.v12i4.31015

Kögler, A., Seibt, K. M., Heitkam, T., Morgenstern, K., Reiche, B., Brückner, M., et al. (2020). Divergence of 3' ends as driver of short interspersed nuclear element (SINE) evolution in the Salicaceae. *Plant J.* 103, 443–458. doi: 10.1111/tpj.14721

Kossack, D. S., and Kinlaw, C. S. (1999). IFG, a *gypsy*-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. *Plant Mol. Biol.* 39, 417–426. doi: 10.1023/A:1006115732620

Kuzmin, D. A., Feranchuk, S. I., Sharov, V. V., Cybin, A. N., Makolov, S. V., Putintseva, Y. A., et al. (2019). Stepwise large genome assembly approach: a case of Siberian larch (*Larix sibirica* Ledeb). *BMC Bioinf.* 20:37. doi: 10.1186/s12859-018-2570-y

LePage, B. A., and Basinger, J. F. (1995). "The evolutionary history of the genus Larix (Pinaceae)" in *Ecology and Management of Larix Forests: A Look Ahead: Proceedings of an International Symposium*. eds. W. C. Schmidt and K. J. Mcdonald (Whitefish, Montana, USA: United States Department of Agriculture, Forest Service, Intermountain Research Station), 19–29.

Liu, B., Zhang, S. G., Zhang, Y., Lan, T. Y., Qi, L. W., and Song, W. Q. (2006). Molecular cytogenetic analysis of four *Larix* species by bicolor fluorescence *in situ* hybridization and DAPI banding. *Int. J. Plant Sci.* 167, 367–372. doi: 10.1086/499317

Liu, Q., Li, X., Zhou, X., Li, M., Zhang, F., Schwarzacher, T., et al. (2019). The repetitive DNA landscape in *Avena* (Poaceae): chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. *BMC Plant Biol.* 19:226. doi: 10.1186/s12870-019-2136-9

Lu, Y., Ran, J.-H., Guo, D.-M., Yang, Z.-Y., and Wang, X.-Q. (2014). Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. *PLoS One* 9:e107679. doi: 10.1371/journal.pone.0115776

Lubaretz, O., Fuchs, J., Ahne, R., Meister, A., and Schubert, I. (1996). Karyotyping of three Pinaceae species via fluorescent *in situ* hybridization and computer-aided chromosome analysis. *Theor. Appl. Genet.* 92, 411–416.

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

Maiwald, S., Weber, B., Seibt, K. M., Schmidt, T., and Heitkam, T. (2021). The Cassandra retrotransposon landscape in sugar beet (*Beta vulgaris*) and related Amaranthaceae: recombination and re-shuffling lead to a high structural variability. *Ann. Bot.* 127, 91–109. doi: 10.1093/aob/mcaa176

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2011). CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229. doi: 10.1093/nar/gkq1189

Maumus, F., and Quesneville, H. (2016). Impact and insights from ancient repetitive elements in plant genomes. *Curr. Opin. Plant Biol.* 30, 41–46. doi: 10.1016/j.pbi.2016.01.003

McGurk, M. P., and Barbash, D. A. (2018). Double insertion of transposable elements provides a substrate for the evolution of satellite DNA. *Genome Res.* 28, 714–725. doi: 10.1101/gr.231472.117

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

Meštrović, N., Mravinac, B., Pavlek, M., Vojvoda-Zeljko, T., Šatović, E., and Plohl, M. (2015). Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosom. Res.* 23, 583–596. doi: 10.1007/s10577-015-9483-7

Mosca, E., Cruz, F., Gómez-Garrido, J., Bianco, L., Rellstab, C., Brodbeck, S., et al. (2019). A reference genome sequence for the European silver fir (*Abies alba* mill.): a community-generated genomic resource. *G3: Genes Genom. Genet.* 9:2039. doi: 10.1534/g3.119.400083

Neale, D. B., McGuire, P. E., Wheeler, N. C., Stevens, K. A., Crepeau, M. W., Cardeno, C., et al. (2017). The Douglas-fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. *G3: Genes Genom. Genet* 7, 3157–3167. doi: 10.1534/g3.117.300078

Neale, D. B., and Wheeler, N. C. (2019a). *The Conifers: Genomes. Variation and Evolution*. Cham: Springer International Publishing.

Neale, D. B., and Wheeler, N. C. (2019b). "Noncoding and repetitive DNA" in *The Conifers: Genomes, Variation And Evolution*. eds. D. B. Neale and N. C. Wheeler (Cham: Springer International Publishing), 61–74.

Neumann, P., Novák, P., Hoštáková, 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. *Mob. DNA* 10:1. doi: 10.1186/s13100-018-0144-1

Novák, P., Guignard, M. S., Neumann, P., Kelly, L. J., Mlinarec, J., Koblížková, A., et al. (2020). Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nat. Plants* 6, 1325–1329. doi: 10.1038/s41477-020-00785-x

Novák, 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

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., Robledillo, L. A., Koblızková, 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

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211

Oliveira, L. C., and Torres, G. A. (2018). Plant centromeres: genetics, epigenetics and evolution. *Mol. Biol. Rep.* 45, 1491–1497. doi: 10.1007/s11033-018-4284-7

Paesold, S., Borchardt, D., Schmidt, T., and Dechyeva, D. (2012). A sugar beet (*Beta vulgaris* L.) reference FISH karyotype for chromosome and chromosome-arm identification, integration of genetic linkage groups and analysis of major repeat family distribution. *Plant J.* 72, 600–611. doi: 10.1111/j.1365-313X.2012.05102.x

Palacios-Gimenez, O. M., Milani, D., Song, H., Marti, D. A., Lopez-Leon, M. D., Ruiz-Ruano, F. J., et al. (2020). Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. *Genome Biol. Evol.* 12, 88–102. doi: 10.1093/gbe/evaa018

Pâques, L. E., Foffova, E., Heinze, B., Lelu-Walter, M.-A., Liesebach, M., and Philippe, G. (2013). "Larches (Larix sp.)" in *Forest Tree Breeding in Europe*. ed. L. E. Pâques (Netherlands: Springer), 13–106.

Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes* 9:88. doi: 10.3390/genes9020088

Perera, D., Magbanua, Z. V., Thummasuwan, S., Mukherjee, D., Arick, M. 2nd, Chouvarine, P., et al. (2018). Exploring the loblolly pine (*Pinus taeda* L.) genome by BAC sequencing and *c0t* analysis. *Gene* 663, 165–177. doi: 10.1016/j.gene.2018.04.024

Plohl, M., Mestrovic, N., and Mravinac, B. (2012). Satellite DNA evolution. *Genome Dyn.* 7, 126–152. doi: 10.1159/000337122

Prunier, J., Verta, J.-P., and MacKay, J. J. (2016). Conifer genomics and adaptation: at the crossroads of genetic diversity and genome function. *New Phytol.* 209, 44–62. doi: 10.1111/nph.13565

Ren, L., Huang, W., Cannon, E. K. S., Bertioli, D. J., and Cannon, S. B. (2018). A mechanism for genome size reduction following genomic rearrangements. *Front. Genet.* 9:454. doi: 10.3389/fgene.2018.00454

Ribeiro, T., Vasconcelos, E., dos Santos, K. G. B., Vaio, M., Brasileiro-Vidal, A. C., and Pedrosa-Harand, A. (2020). Diversity of repetitive sequences within compact genomes of *Phaseolus* L. beans and allied genera *Cajanus* L. and *Vigna* Savi. *Chromosom. Res.* 28, 139–153. doi: 10.1007/s10577-019-09618-w

Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., and Camacho, J. P. M. (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* 6:28333. doi: 10.1038/srep28333

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. New York, USA: Cold Spring Harbor Laboratory Press.

Sarri, V., Ceccarelli, M., and Cionini, P. G. (2011). Quantitative evolution of transposable and satellite DNA sequences in *Picea* species. *Genome* 54, 431–435. doi: 10.1139/g11-007

Sarri, V., Minelli, S., Panara, F., Morgante, M., Jurman, I., Zuccolo, A., et al. (2008). Characterization and chromosomal organization of satellite DNA sequences in *Picea abies*. *Genome* 51, 705–713. doi: 10.1139/G08-048

Satović, E., Vojvoda Zeljko, T., Luchetti, A., Mantovani, B., and Plohl, M. (2016). Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements. *BMC Genomics* 17:997. doi: 10.1186/s12864-016-3347-1

Schmidt, N., Seibt, K. M., Weber, B., Schwarzacher, T., Schmidt, T., and Heitkam, T. (2021). Broken, silent, and in hiding: Tamed endogenous pararetroviruses escape elimination from the genome of sugar beet (*Beta vulgaris*). *Ann. Bot.* [Preprint]. doi: 10.1093/aob/mcab042

Schmidt, T., Heitkam, T., Liedtke, S., Schubert, V., and Menzel, G. (2019). Adding color to a century-old enigma: multi-color chromosome identification unravels the autotriploid nature of saffron (*Crocus sativus*) as a hybrid of wild *Crocus cartwrightianus* cytotypes. *New Phytol.* 222, 1965–1980. doi: 10.1111/nph.15715

Schmidt, T., and Heslop-Harrison, J. S. (1998). Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends Plant Sci.* 3, 195–199. doi: 10.1016/S1360-1385(98)01223-0

Schmidt, T., Schwarzacher, T., and Heslop-Harrison, J. S. (1994). Physical mapping of rRNA genes by fluorescent *in-situ* hybridization and structural analysis of 5S rRNA genes and intergenic spacer sequences in sugar beet (*Beta vulgaris*). *Theor. Appl. Genet.* 88, 629–636. doi: 10.1007/BF01253964

Seibt, K. M., Schmidt, T., and Heitkam, T. (2018). FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. *Bioinformatics* 34, 3575–3577. doi: 10.1093/bioinformatics/bty395

Stevens, K. A., Wegrzyn, J., Zimin, A., Puiu, D., Crepeau, M., Cardeno, C., et al. (2016). Sequence of the sugar pine megagenome. *Genetics* 206, 1613–1626. doi: 10.1534/genetics.116.193227

Sultana, N., Menzel, G., Heitkam, T., Schmidt, T., Kojima, K. K., Bao, W., et al. (2020). Bioinformatics and molecular analysis of satellite repeat diversity in *Vaccinium* genomes. *Genes* 11:527. doi: 10.3390/genes11050527

Utsunomia, R., Ruiz-Ruano, F. J., Silva, D. M. Z. A., Serrano, É. A., Rosa, I. F., Scudeler, P. E. S., et al. (2017). A glimpse into the satellite DNA library in Characidae fish (Teleostei, Characiformes). *Front. Genet.* 8:103. doi: 10.3389/fgene.2017.00103

Vondrak, T., Ávila Robledillo, L., Novák, P., Koblížková, A., Neumann, P., and Macas, J. (2020). Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. *Plant J.* 101, 484–500. doi: 10.1111/tpj.14546

Voronova, A., Belevich, V., Korica, A., and Rungis, D. (2017). Retrotransposon distribution and copy number variation in gymnosperm genomes. *Tree Genet. Genomes* 13, 1–23. doi: 10.1007/s11295-017-1165-5

Wegrzyn, J. L., Lee, J. M., Tearse, B. R., and Neale, D. B. (2008). TreeGenes: a forest tree genome database. *Int. J. Plant Genomics* 2008:412875. doi: 10.1155/2008/412875

Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L.-S., Loopstra, C. A., Vasquez-Gross, H. A., et al. (2014). Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196, 891–909. doi: 10.1534/genetics.113.159996

Wegrzyn, J. L., Lin, B. Y., Zieve, J. J., Dougherty, W. M., Martínez-García, P. J., Koriabine, M., et al. (2013). Insights into the loblolly pine genome: characterization of BAC and fosmid sequences. *PLoS One* 8:e72439. doi: 10.1371/journal.pone.0072439

Wei, X. X., and Wang, X. Q. (2003). Phylogenetic split of *Larix*: evidence from paternally inherited cpDNA trnT-trnF region. *Plant Syst. Evol.* 239, 67–77. doi: 10.1007/s00606-002-0264-3

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. Koeltz Botanical Books.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Willard, H. F., and Waye, J. S. (1987). Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet.* 3, 192–198. doi: 10.1016/0168-9525(87)90232-0

Zhang, S. G., Yang, W. H., Han, S. Y., Han, B. T., Li, M. X., and Qi, L. W. (2010). Cytogenetic analysis of reciprocal hybrids and their parents between *Larix leptolepis* and *Larix gmelinii*: implications for identifying hybrids. *Tree Genet. Genomes* 6, 405–412. doi: 10.1007/s11295-009-0258-1

Zhang, T., Talbert, P. B., Zhang, W., Wu, Y., Yang, Z., Henikoff, J. G., et al. (2013). The CentO satellite confers translational and rotational phasing on CenH3 nucleosomes in rice centromeres. *Proc. Natl. Acad. Sci. U. S. A.* 110, E4875–E4883. doi: 10.1073/pnas.1319548110

Zimmermann, H. H., Harms, L., Epp, L. S., Mewes, N., Bernhardt, N., Kruse, S., et al. (2019). Chloroplast and mitochondrial genetic variation of larches at the Siberian tundra-taiga ecotone revealed by *de novo* assembly. *PLoS One* 14:e0216966. doi: 10.1371/journal.pone.0216966

Zonneveld, B. J. M. (2012). Conifer genome sizes of 172 species, covering 64 of 67 genera, range from 8 to 72 picogram. *Nord. J. Bot.* 30, 490–502. doi: 10.1111/j.1756-1051.2012.01516.x

Zuccolo, A., Scofield, D. G., De Paoli, E., and Morgante, M. (2015). The Ty1-*copia* LTR retroelement family PARTC is highly conserved in conifers over 200MY of evolution. *Gene* 568, 89–99. doi: 10.1016/j.gene.2015.05.028

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 Genetics

BRIEF RESEARCH REPORT
published: 24 August 2021
doi: 10.3389/fgene.2021.694866

First Description of a Satellite DNA in Manatees' Centromeric Regions

*Mirela Pelizaro Valeri[1], Guilherme Borges Dias[2], Alice Alves do Espírito Santo[1], Camila Nascimento Moreira[3], Yatiyo Yonenaga-Yassuda[3], Iara Braga Sommer[4], Gustavo C. S. Kuhn[1] and Marta Svartman[1]**

[1]Laboratório de Citogenômica Evolutiva, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [2]Department of Genetics and Institute of Bioinformatics, University of Georgia, Athens, GA, United States, [3]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil, [4]Centro Nacional de Pesquisa e Conservação da Biodiversidade Marinha do Nordeste, Instituto Chico Mendes de Conservação da Biodiversidade, Brasília, Brazil

Trichechus manatus and *Trichechus inunguis* are the two Sirenia species that occur in the Americas. Despite their increasing extinction risk, many aspects of their biology remain understudied, including the repetitive DNA fraction of their genomes. Here we used the sequenced genome of *T. manatus* and TAREAN to identify satellite DNAs (satDNAs) in this species. We report the first description of TMAsat, a satDNA comprising ~0.87% of the genome, with ~684 bp monomers and centromeric localization. In *T. inunguis*, TMAsat showed similar monomer length, chromosome localization and conserved CENP-B box-like motifs as in *T. manatus*. We also detected this satDNA in the *Dugong dugon* and in the now extinct *Hydrodamalis gigas* genomes. The neighbor-joining tree shows that TMAsat sequences from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* lack species-specific clusters, which disagrees with the predictions of concerted evolution. We detected a divergent TMAsat-like homologous sequence in elephants and hyraxes, but not in other mammals, suggesting this sequence was already present in the common ancestor of Paenungulata, and later became a satDNA in the Sirenians. This is the first description of a centromeric satDNA in manatees and will facilitate the inclusion of Sirenia in future studies of centromeres and satDNA biology.

Keywords: tandem repeats, *Trichechus manatus*, *Trichechus inunguis*, chromosome mapping, fluorescent *in situ* hybridization, TAREAN

OPEN ACCESS

Edited by:
Francisco J. Ruiz-Ruano, University of East Anglia, United Kingdom

Reviewed by:
Nevenka Mestrovic, Rudjer Boskovic Institute, Croatia
Isidoro Feliciello, University of Naples Federico II, Italy

****Correspondence:***
Marta Svartman
svartmanm@ufmg.br

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *13 April 2021*
Accepted: *30 July 2021*
Published: *24 August 2021*

Citation:
Valeri MP, Dias GB, Espírito Santo AA, Moreira CN, Yonenaga-Yassuda Y, Sommer IB, Kuhn GCS and Svartman M (2021) First Description of a Satellite DNA in Manatees' Centromeric Regions. Front. Genet. 12:694866. doi: 10.3389/fgene.2021.694866

INTRODUCTION

The order Sirenia encompasses four extant herbivorous aquatic mammals. The Dugongidae family includes the *Dugong dugon* and the Steller's sea cow *Hydrodamalis gigas*, the latter now extinct due to overhunting, and the Trichechidae family includes three manatee species: *Trichechus manatus*, *Trichechus inunguis*, and *Trichechus senegalensis* (Domning, 2018). *Dugong dugon* occurs across coastal waters in the Indo-West Pacific Ocean, and *T. senegalensis* is restricted to the west coast of Africa, making *T. manatus*, the West Indian manatee, and *T. inunguis*, the Amazonian manatee, the only sirenians to occur in the Americas. The West Indian manatee occurs in Caribbean waters and the Atlantic coast ranging from Florida to the northeast of Brazil, and *T. inunguis* is found along the Amazon River basin (Bonvicino et al., 2020).

All extant sirenians are considered as vulnerable by the International Union for Conservation of Nature and Natural Resources (IUCN; Deutsch et al., 2008; Keith, 2015; Marmontel et al., 2016; Marsh and Sobtzick, 2019).

The West Indian manatee has two recognized subspecies: *Trichechus manatus latirostris* (Florida Manatee), found in the United Estates and Gulf of Mexico coasts, and *T. m. manatus* (Antillean manatee), found in the Caribbean, Central and South America. Recent morphological and genetic analyses suggest the need for a revision in the *T. manatus* taxonomy considering the influence of the Amazon River as a barrier to gene flow. These studies showed that the *T. m. manatus* populations from the Caribbean and up to the Amazon River mouth are phylogenetically closer to the populations of *T. m. latirostris* from the United States than to the Brazilian *T. m. manatus* populations south of the Amazon River mouth (Vianna et al., 2006; Barros et al., 2017; Lima et al., 2019, 2021). Hybrids between *T. manatus* and *T. inunguis* have also been reported on the sympatric area at the Amazon River mouth (Vianna et al., 2006; Lima et al., 2019; Luna et al., 2021).

Satellite DNAs (satDNAs) are a type of repetitive DNA found in most eukaryotic genomes. They are arranged as long arrays of tandem repeats with variable unit length, number of copies and chromosome organization. SatDNAs are usually associated with chromosome landmarks such as centromeres, telomeres, and heterochromatic regions. Despite the fact that satDNAs do not encode proteins, they are associated with important biological functions such as formation and maintenance of heterochromatin at telomeres and centromeres, and maintenance of chromosome integrity and genome stability (reviewed in Shapiro and von Sternberg, 2005; Biscotti et al., 2015; Shatskikh et al., 2020). SatDNAs can form higher-order repeat (HOR) units made of multimers with a number of diverged monomers that are tandemly repeated as a set (reviewed in Plohl et al., 2012; Vlahović et al., 2016). HOR organization has been found in several satDNAs, including the alfa centromeric satDNA in humans, and may be relevant to the centromeric function (Sujiwattanarat et al., 2015; Sullivan et al., 2017). In addition, satDNAs monomer sequences can present internal repetitions, which may be related with secondary structures relevant to centromeric function (Kasinathan and Henikoff, 2018). Centromeric satDNAs in mammals usually present the CENP-B box, a conserved 17 bp region known to be the DNA-binding domain for the centromeric protein B (CENPB), with nine nucleotides (nTTCGnnnnAnnCGGGn) composing the most evolutionarily conserved domain (ECD; Muro et al., 1992; Masumoto et al., 2004; Alkan et al., 2011; Kasinathan and Henikoff, 2018). Most satDNAs are under concerted evolution, a process by which new mutations within monomers are quickly homogenized across the repeat family and fixed in reproductively isolated populations, resulting in intraspecific repeat homogeneity but interspecific divergence (Dover, 1982; Plohl et al., 2012; Smalec et al., 2019). Moreover, according to the library model, related species may share a collection of satDNAs sequences with mostly quantitative interspecies differences due to expansion or contraction (even elimination) during the evolution (Fry and Salser, 1977; Meštrović et al., 1998). Another aspect of satDNAs evolution is their relationship with mobile elements, since there are several examples of satDNAs derived from transposons and retrotransposons in plants and animals (reviewed in Meštrović et al., 2015).

The repetitive DNA fraction of manatees' genomes has been poorly studied, especially in the case of satDNAs. We used the sequenced genome of *T. manatus* and the TAREAN (Novák et al., 2017, 2020) pipeline to explore the satDNAs present in this genome. Herein, we describe for the first time the centromeric satDNA of the West Indian manatee, which we found to be also present in the Amazonian manatee, the dugong, and in the extinct Steller's sea cow. We characterized this sequence *in silico* and mapped it in *T. manatus* and *T. inunguis* chromosomes. In addition, we investigated the presence of the TMAsat sequence in mammals outside the order Sirenia, which allowed us to establish a rough timeline for its origin.

MATERIALS AND METHODS

De novo Identification of Satellite DNAs

In order to identify satDNAs in manatees, we used whole-genome sequencing data from *T. m. latirostris* (accession number SRR328416) available in the *National Center for Biotechnology Information* – NCBI and the TAREAN pipeline (Novák et al., 2017). The first step of this pipeline is a graph-based clustering, which performs all to all similarity comparisons of DNA sequencing reads, resulting in clusters of those reads derived from repetitive elements. Then, it examines the presence of circular or globula-like graph structures to identify potential tandem repeats, classified as putative high or low confidence satellites. The raw Illumina reads (~100 bp long) used in this analysis were randomly sampled by TAREAN, comprising ~2.4% (870,965 reads) of the 3.67 pg estimated genome size (Kasai et al., 2013). The reads that make up each cluster are partially assembled into contigs that were used for repeat annotation with the CENSOR web server (Kohany et al., 2006) that contains a collection of Mammalia repeats from RepBase, updated in 08-24-2020 (Bao et al., 2015). The single potential tandem repeat cluster (13) with globula/ring-like structure was analyzed in detail through similarity searches against the *T. manatus* reference genome (accession GCA_000243295.1) using the BLASTn tool with default parameters (Altschul et al., 1990) to verify if the sequence is a tandem repeat. In addition to the annotation using the CENSOR web server, this cluster was annotated through BLASTn similarity searches against the whole nucleotide collection (nr/nt).

The identified satDNA sequence was characterized regarding its genome proportion, monomer length, AT content, and presence of internal direct or inverted duplications. The satDNA genome proportion was estimated by TAREAN. TAREAN tries to improve the assembly process by applying a k-mer-based approach to obtain a less fragmented monomer consensus, but it restricts itself to the 50% most prevalent k-mers in a cluster. For this reason, we chose the whole-genome assembly resource as a more representative sample of TMAsat diversity. The most common sequence (MCS) of TMAsat was generated using Geneious (prime version 2020.2.4) with a 25% threshold and 66 monomeric sequences retrieved from the reference genome, previously aligned with the muscle aligner

implemented in MEGA X. The MCS was used to estimate monomer length, AT content, and presence of internal repetitions. The last feature was also conducted in the Geneious software using the diagonal plot method in high sensitive mode, with window size of 50 bp and identity threshold of 60%.

We searched for the presence of TMAsat in the two other Sirenia species with a sequenced genome available in NCBI, *D. dugon* (under accession numbers of assembled genome GCA_015147995.1 and raw Illumina reads DRR251525) and the extinct *H. gigas* (under accession numbers of assembled genome GCA_013391785.1 and raw Illumina reads SRR12067498). First, we used TMAsat sequence as query in BLASTn similarity searches against these assembled genomes. In addition, we also used the raw Illumina reads (~150 bp long) and TAREAN to identify TMAsat in these genomes. The analyzed reads were randomly sampled by TAREAN totalizing 1,038,927 in *D. dugon* and 570,097 in *H. gigas*. The MCS of TMAsat in *D. dugon* and *H. gigas* were generated using monomeric sequences retrieved from the reference genome after BLAST searches, totalizing 50 sequences from *D. dugon* and 40 from *H. gigas*. The TMAsat MCS in *T. inunguis* was obtained using the five cloned sequences obtained by PCR. All MCS were generated as described previously for *T. manatus*.

Biological Samples

We used biological samples of *T. manatus* and *T. inunguis* to determine TMAsat chromosomal distribution and investigate its presence in *T. inunguis*, whose genome has not been sequenced. Skin sample from a male *T. manatus* captured at Porto de Pedras/AL, Brazil (−9.164167 and −35.294444) in 2019 was provided by CEPENE/ICMBio (SISBIO 60829-2) and used for fibroblast culture. A fibroblast cell line from a male *T. inunguis* established in 1998 was provided by Dr. Yatiyo Yonenaga-Yassuda, from the University of São Paulo, and was previously analyzed by Assis et al. (1998). Chromosome spreads from fibroblast cultures were obtained according to Stanyon and Galleni (1991) and genomic DNAs were extracted with the Wizard Genomic DNA Purification Kit (Promega).

PCR Amplification, Cloning, and Sequencing of Satellite DNAs

TMAsat was amplified by PCR from the *T. inunguis* genomic DNA using primers designed from the satDNA consensus sequence (estimated by TAREAN) as follow: TMAsat-F CTCCTTCAAGCTGCTTAACT and TMAsat-R GGGAACTTACACTTGCTGCT. The PCR cycling conditions were as follows: 95°C – 3 min, 35 cycles: 95°C – 30 s; 55°C for 30 s; 72°C – 1 min; and 72°C – 3 min for final elongation. The PCR product corresponding to monomer size was excised from the agarose gel and purified using the Illustra GFX PCR DNA and Gel Band Purification Kit. The purified products were ligated into the pGEM-T Easy vector (Promega) and used in the transformation of *Escherichia coli* XL1-BLUE strain electrocompetent cells (Phoneutria). Five recombinant colonies of TMAsat were sequenced (access numbers MW272776–MW272780) with the ABI3130 platform (Applied Biosystems).

Fluorescence *in situ* Hybridization

Fluorescent *in situ* hybridization (FISH) was performed using the TMAsat cloned (MW272776) sequence as probe on metaphase spreads of *T. manatus* and *T. inunguis*. FISH was performed with 200 ng of biotin-labelled probes, following (Valeri et al., 2020). The analyses and image acquisition were performed under a Zeiss Axioimager 2 epifluorescence microscope equipped with a CCD camera and with the AxioVision software (Carl Zeiss MicroImaging, Jena, Germany), respectively.

In silico Characterization of satDNAs

DNA polymorphisms and nucleotide diversity along the satDNA sequences were analyzed using the software DnaSP 6.12.03 (Rozas et al., 2017) with the same monomer sequences used to generate the MCS from *T. manatus*, *D. dugon*, and *H. gigas*. In this analysis, the monomer sequences were previously aligned with the muscle method (Edgar, 2004) implemented in MEGA X and the window length and step size were set for 10 and 1 bp, respectively. Windows were classified as conserved or variable if they exhibited more than two SDs bellow or above the nucleotide average variability, respectively.

Monomer sequences of TMAsat from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* were aligned with the muscle method implemented in MEGAX and used for the construction of a neighbor-joining tree. These sequences were the same used to obtain the MCS, totalizing 161 sequences, including 66 from *T. manatus*, five from *T. inunguis*, 50 from *D. dugon* and 40 from *H. gigas*. The neighbor-joining tree was obtained using MEGA X with 500 bootstrap replicates and the final tree was visualized in iTOL v4.3.3[1] (Letunic and Bork, 2019). We also used the same set of sequences to estimate the inter- and intra-specific nucleotide divergence (number of base substitutions per site), as well as the average nucleotide divergence over all pairwise sequence comparisons using MEGA X.

We searched for any putative CENP-B box in TMAsat MCS from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* using the 17 bp sequence containing the ECD (nTTCGnnnnAnnCGGGn; Masumoto et al., 2004) and CENP-B box sequences of *Loxodonta africana*/*Dasypus novemcintus* (CTTTGCCGAGAACGGAG; Alkan et al., 2011). This search was conducted in the Geneious software in global pairwise alignment mode and 51% similarity cost matrix.

To investigate the presence of TMAsat in other mammals, we utilized the MCS from *T. manatus* as query in BLASTn similarity searches against Mammalia (NCBI:txid40674) wgs database excluding Sirenia (NCBI:txid9774; search date 06-07-2021). The flanking regions of TMAsat similarity hits were analyzed with the CENSOR web server (Kohany et al., 2006) containing the Mammalia RepBase library (updated in 06-14-2021; Bao et al., 2015). To better analyze these hits with the TMAsat consensus sequence, we compared their sequences using dotplots and pairwise alignments in the Geneious software.

[1]https://itol.embl.de/

RESULTS

In silico Identification and satDNA Analyses

The only potential satDNA identified (with low confidence) by TAREAN in the *T. manatus* genome was represented by the cluster 13. This sequence was analyzed in detail through similarity searches against the *T. manatus* reference genome (accession GCA_000243295.1) using the BLASTn tool with default parameters (Altschul et al., 1990). Despite being classified by TAREAN with low confidence, we verified this sequence tandemly repeated at least 25 times on assembled contigs of *T. manatus*. These repeats comprise 0.87% of the genome of *T. manatus* with monomer length of ~684 bp estimated by

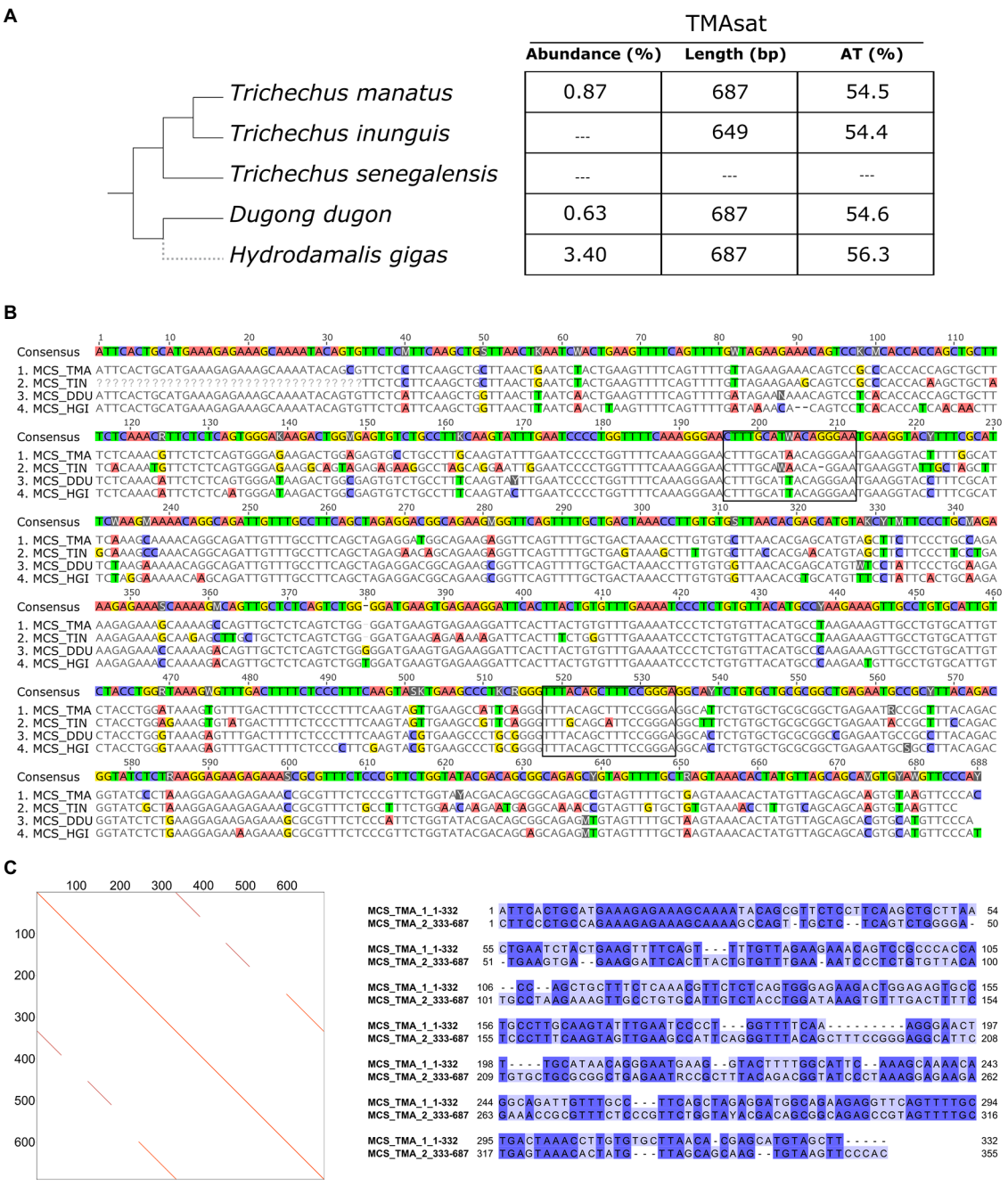

FIGURE 1 | (A) TMAsat abundance, monomer length and AT content of the monomers for the analyzed species. **(B)** Alignment of TMAsat MCS from *Trichechus manatus*, *Trichechus inunguis*, *Dugong dugon*, and *Hydrodamalis gigas* showing the two putative CENP-B box like motifs. **(C)** Dot plot comparison of the TMAsat MCS sequence from *T. manatus* against itself and pairwise alignment of TMAsat position 1–332 against 333–687, with 55.8% of DNA sequence identity.

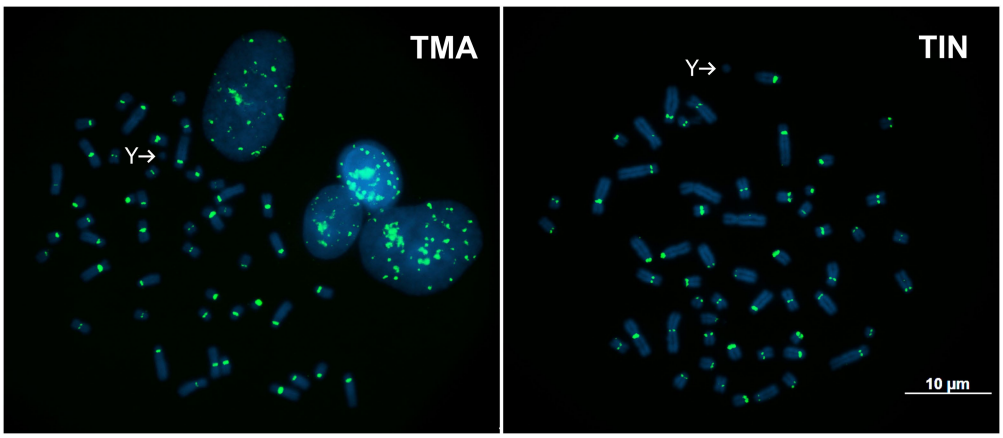

FIGURE 2 | Metaphases of *T. manatus* (TMA) and *T. inunguis* (TIN) after FISH with the TMAsat probe. Y chromosomes without signals of TMAsat are indicated.

TAREAN. The consensus sequence generated by TAREAN (**Supplementary Figure 1**) did not show similarity with any known repetitive DNA from the mammalian RepBase collection (Bao et al., 2015). We named this new satDNA as TMAsat (for *T. manatus* satellite).

The TMAsat MCS from *T. manatus* was generated from an alignment of 66 monomers manually isolated from the assembled reference genome (**Supplementary Figure 2**; **Supplementary File 1**). It showed monomer length of 687 bp and 54.5% of AT content (**Figures 1A,B**). The dotplot of TMAsat against itself revealed a segment repeated twice inside TMAsat, from position 1 to 332 and 333 to 687 (**Figure 1C**), A pairwise alignment of the two segments of TMAsat, 1–332 and 333–687 bp, showed that they are related but quite divergent, with only 55.8% identity (**Figure 1C**). A detailed investigation in the assembled contigs showed that the TMAsat unit of ~687 bp is organized in higher-order structure, mostly alternating the segments TMAsat1 (1–332) and TMAsat2 (333–687). However, we found one case of TMAsat1 dimer (accession NW_004443969.1 position 56,989–75,023 bp), few cases of TMAsat2 dimer (accessions NW_004444053.1; NW_004444936.1; and NW_004444425.1) and in one contig (accession NW_004443969.1) three, six and 10 tandemly repeated units of TMAsat2.

Genomic Distribution of TMAsat in the Genus *Trichechus*

TMAsat was amplified by PCR from *T. inunguis* genomic DNA, and the resulting PCR products showed a similar monomer length of ~647 bp (**Supplementary Figure 3**). The PCR product was cloned and sequenced in order to confirm that it was indeed homologous to TMAsat. The MCS based on the cloned sequences showed similar AT content and 89.6% of identity (**Figures 1A,B**; **Supplementary Figure 4**). A selected TMAsat cloned sequence was labeled and used as probe in FISH on chromosomes of both *T. manatus* and *T. inunguis*. TMAsat showed centromeric localization in *T. manatus* ($2n = 48$) and *T. inunguis* ($2n = 56$), mapping to the centromeres of all chromosomes, except the Y (**Figure 2**; **Supplementary Figure 5**). TMAsat localization is compatible with the CBG-banding pattern in both species, which reveals centromeric heterochromatin in all chromosomes (Assis et al., 1998; Gray et al., 2002), with the exception of the Y.

TMAsat in Other Sirenia

Besides *T. manatus*, there are two additional Sirenia species with sequenced genomes available: *D. dugon* and the extinct *H. gigas*, both belonging to the Dugongidae family. A search for TMAsat sequences on the assembled contigs of these species revealed the presence of tandemly repeated TMAsat sequences. TAREAN returned with high confidence one cluster of a putative satDNA with 685 bp length in both species, cluster 8 in *D. dugon* and cluster 3 in *H. gigas*, which contained homologous sequences to TMAsat (**Figures 1A,B**). In *D. dugon*, cluster 8 represents 0.63% of the genome and the MCS generated from the 50 monomers retrieved from the assembled genome is 687 bp long with 54.6% of AT content (**Supplementary Figure 6**; **Supplementary File 2**). In addition, we found evidence of other HOR configurations rather than alternating TMAsat1 and TMAsat2 in *D. dugon*: a dimer of TMAsat1 (BMBL01107524.1 and BMBL01079760.1), a dimer of TMAsat2 (BMBL01112453.1 and BMBL01093845.1), four (BMBL01013125.1), five (BMBL01107524.1) and six (BMBL01055248.1) tandemly repeated copies of TMAsat2.

In *H. gigas*, cluster 3 comprises 3.4% of the genome and the MCS based on 40 monomers from the reference genome is 687 bp long and has 56.3% of AT content (**Supplementary Figure 7**; **Supplementary File 3**). In this species, the most frequent TMAsat organization is the alternating segments of TMAsat1 and TMAsat2, and we only found one dimer of TMAsat2 (JACANZ010402190.1).

The sliding window analysis of nucleotide variability of this satDNA in *T. manatus*, *D. dugon*, and *H. gigas* revealed the presence of conserved and variable regions within the monomers (**Figures 3A–C**). However, we did not have access to biological samples of *D. dugon* or *H. gigas* to map TMAsat on their chromosomes. The monomeric TMAsat sequences from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* were aligned and

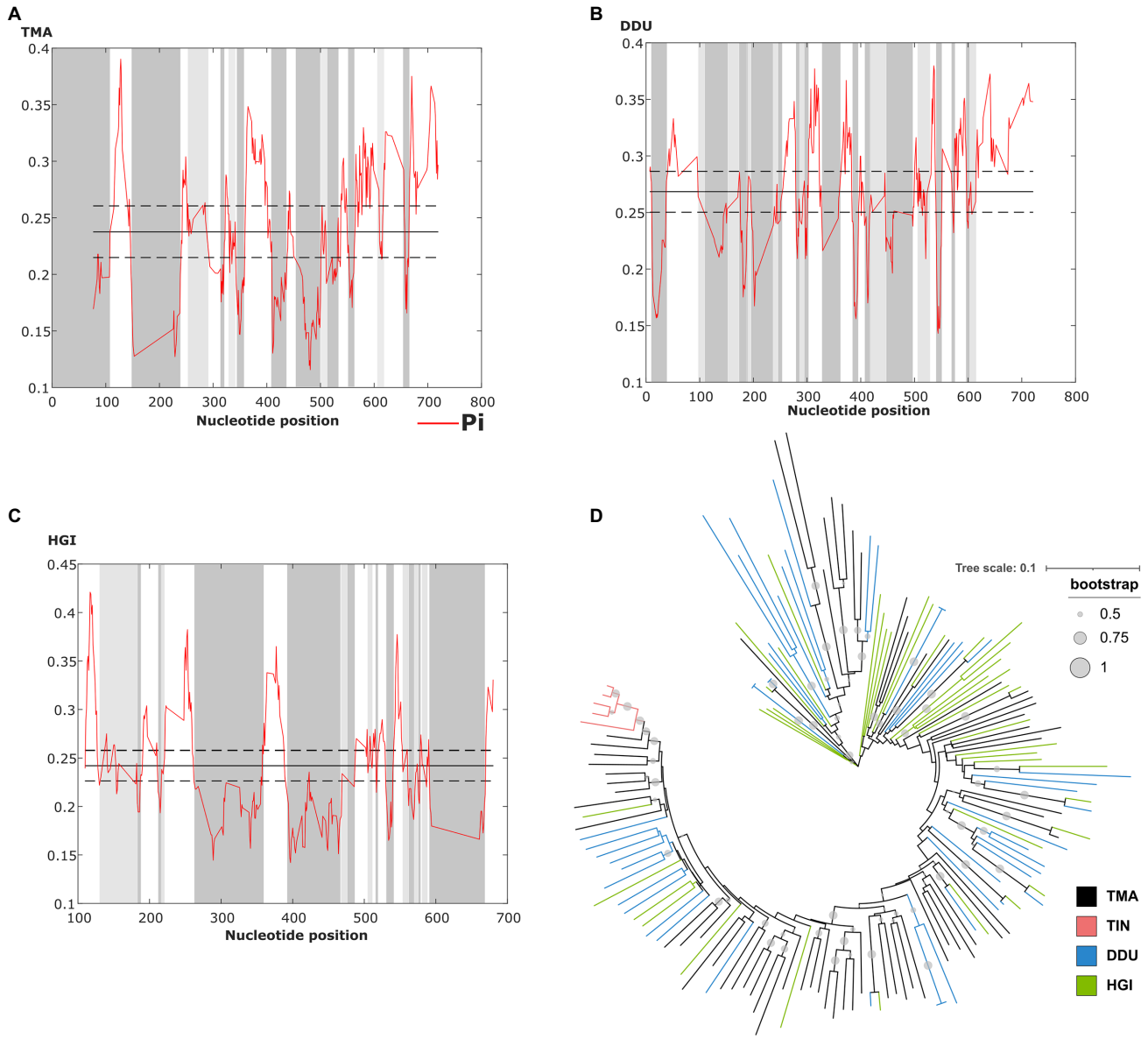

FIGURE 3 | Identification of conserved and variable TMAsat segments of **(A)** *T. manatus*, **(B)** *D. dugon* and **(C)** *H. gigas* by sliding window analysis (sliding window of 10 bp and step size of 1 bp). Average nucleotide diversity (Pi) is indicated by the red line, while average nucleotide diversity is indicated by the solid line, and average diversity ± 2 SD is indicated by the dashed line. **(D)** Neighbor-joining tree with TMAsat sequences of *T. manatus* (TMA), *T. inunguis* (TIN), *D. dugon* (DDU), and *H. gigas* (HGI). Bootstrap values generated from 500 replicates.

TABLE 1 | The two putative CENPB box-like motifs identified in the MCS of TMAsat from *T. manatus* (TMA), *T. inunguis* (TIN), *D. dugong* (DDU), and *H. gigas* (HGI).

	Position 196–212 bp	**Position 518–534 bp**
ECD	NTTCGNNNNANNCGGGN	NTTCGNNNNANNCGGGN
L. africana and *D. novemcinctus*	CTTTGCCGAGAACGGAG	CTTTGCCGAGAACGGAG
T. manatus	CTTTGCATAACAGGGAA	TTTACAGCTTTCCGGGA
T. inunguis	CTTTGCAWAACAGGAAT	TTTGCAGCATTCCGGGA
D. dugon	CTTTGCATTACAGGGAA	TTTACAGCTTTCCGGGA
H. gigas	CTTTGCATTACAGGGAA	TTTACAGCTTTCCGGGA

Conserved nucleotides in the evolutionarily conserved domain (ECD) are shown in red/highlighted, and conserved nucleotides compared with L. africana and D. novemcinctus motif (Alkan et al., 2007) other than the ECD domain are shown in blue.

used to construct a neighbor-joining tree, which did not reveal any species-specific clustering (**Figure 3D**). We also estimated the inter and intraspecific nucleotide divergence (**Supplementary Table 1**), as well as the average divergence over all sequence pairs ($d = 0.34$). As expected from the Neighbor Joining results, TMAsat intraspecific diversity was not lower than interspecific diversity, except in *T. inunguis* ($d = 0.06$). The low diversity of TMAsat sequences in *T. inunguis* may be due to the low number of sequences used in the analysis and the use of PCR.

CENP-B Box Is Present in TMAsat

The CENP-B box is a 17 bp region conserved among mammalian centromeric satDNAs and known to be the DNA-binding domain for the centromeric protein CENPB. We searched for putative CENP-B box-like motifs within TMAsat MCS from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas*, and found two putative motifs MCS (**Figure 1B**). The first is located in position 196 to 212 bp, matching best with the CENP-B box sequence found in *L. africana* and *D. novemcinctus*. The second putative motif was found in position 518–534 bp. Both motifs display 5–6 identical nucleotides to ECD out of nine in all Sirenia species (**Table 1**). The two putative CENP-B box motifs were present in a conserved segment as indicated by the sliding window analysis of nucleotide variability among satDNA monomers from *T. manatus* and *D. dugon* (**Figures 3A,B**). In *H. gigas*, these motifs spanned both conserved and variable regions of the TMAsat monomer (**Figure 3C**).

The CENP-B box-like motifs found in positions 196–212 bp of TMAsat from *T. manatus* (CTTTGCATAACAGGGAA) and *T. inunguis* (CTTTGCAWAACA-GGAAT) shared 14 out of the 17 nucleotides with each other. In *D. dugon* and in *H. gigas* the CENPB-box-like motif was the same (CTTTGCATTACAGGGAA) and shared 15 out of 17 nucleotides with *T. manatus*. Six bases in *T. manatus*, *D. dugon* and *H. gigas* and five in *T. inunguis*, out of the nine from the ECD were conserved. The second putative motif (position 518–534 bp) showed six out of nine identical bases to the ECD in the four analyzed species. *T. manatus*, *D. dugon*, and *H. gigas* shared an identical second motif (TTTACAGCTTTCCGGGA), whereas *T. inunguis* differed in two nucleotides (TTTGCAGCATTCCGGGA).

TMAsat in Other Mammals

We investigated the presence of TMAsat in other mammals using the MCS from *T. manatus* as query in similarity searches against Mammalia sequences in the wgs database from NCBI excluding Sirenia. The total number of returned hits was 13 distributed in four species (**Supplementary Table 2**). With a cut off for query cover equal or greater than 30%, we found four hits in the African elephant (*L. africana*) and four hits in the Asian elephant (*Elephas maximus*). In addition to African and Asian elephants, *Procavia capensis* and *Heterohyrax brucei* appeared in the hits with query covers smaller than 30%. Looking closer into these contigs from *L. africana*, *E. maximus*, *Procavia capensis*, and *Heterohyrax brucei*, we verified few sequences in tandem (maximum of 18) with the repetition unit comprising roughly one TMAsat HOR monomer. The small number of hits found suggests that this sequence is not a typical satDNA in these taxa, but is instead a repetition related to a transposable element. Indeed, nine out of 10 TMAsat arrays were flanked by LINE-1 in *L. africana*, and 15 out of 16 in *E. maximus*.

DISCUSSION

The TMAsat, reported herein for the first time, was the only putative satDNA found in our analysis, comprising less than 1% of the *T. manatus* genome and mapping to the centromeric regions of all chromosomes, except the Y. The TMAsat could be absent or undetectable by FISH due to low copy number or sequence divergence on the Y chromosome. In *T. inunguis*, we confirmed the presence of TMAsat by PCR and FISH and despite the two species having different karyotypes ($2n = 48$ and $2n = 56$, respectively), TMAsat displayed the same chromosome localization (**Figure 2**). This could be related to the recent ~1.34 million years ago (Mya) divergence time between the species (de Souza et al., 2021).

We also detected the TMAsat in *D. dugon* and *H. gigas* with similar monomer length, comprising 0.63 and 3.4% of the genomes, respectively. The different genome proportion found in *T. manatus* (Illumina HiSeq; 150x genome coverage), *D. dugon* (Illumina Novaseq6000; 64x genome coverage) and *H. gigas* (Illumina NovaSeq; 11x genome coverage) could be due to different genome coverage and/or sequencing platforms used for each species, and may not reflect real interspecific variation. This is especially true in the case of the extinct *H. gigas*, whose DNA source for genome sequencing is a petrous bone from a specimen who probably died during the 1760s (Sharko et al., 2021), and thus the abundance estimates need to be taken with caution.

Although there are slight differences within the MCS from each species, the Neighbor Joining analysis does not indicate intraspecific homogeneous monomers. Only the monomers from *T. inunguis* were grouped together, probably due to the low number of sequences used in the analysis or biased PCR amplification with the selected primers. Nevertheless, we cannot discard a species-specific TMAsat sequence in *T. inunguis* since some mutations are present in all or almost all five sequences and are absent or present in just few monomers outside the species. West Indian and Amazonian manatees present a recent divergence time (de Souza et al., 2021) and an incomplete reproductive isolation (Vianna et al., 2006; Lima et al., 2019), which could contribute to the TMAsat high interspecific homogeneity observed. Overall, the species-specific mutations of the group are probably not yet fixed, despite the ~46.83 Mya estimated split of Trichechidae and Dugongidae, thus lacking species-specific sequences as reflected in the neighbor joining tree (**Figure 3D**), which disagrees with the predictions of concerted evolution. This process, which has been described for many satDNAs, promotes fast sequence homogenization within a species or population, resulting in much higher interspecific than intraspecific differences (Plohl et al., 2012). Although interspecific satDNA sequence conservation is unexpected according to the concerted evolution model, interspecific homogeneity of centromeric satDNAs was observed in other mammalian groups, like in rodents from the *Peromyscus* genus

(Smalec et al., 2019), in four squirrel monkeys (*Saimiri* genus; Valeri et al., 2020) and in two species of two-toed sloths from the genus *Choloepus* (Sena et al., 2020). In all these cases, a possible centromeric function was hypothesized. Moreover, the library model of satDNA evolution relies on the preexistence of a satDNA collection in related species, with the differences observed among the species mostly due to amplification-contraction events of these sequences pool, and does not imply in rapid sequence changes (Plohl et al., 2009). These could be the case of TMAsat evolution if considering the monomer variants as independent amplification-contraction units.

In addition to the centromeric localization in *T. manatus* and *T. inunguis*, we detected the CENPB-box like motif, another centromeric feature, twice in the TMAsat sequences of all four Sirenia species. In *T. manatus* and in *D. dugon*, both putative CENPB-boxes were located in conserved segments of TMAsat. Even though the CENPB-box-like motif found in TMAsat does not present all the nine nucleotides of the ECD, we cannot exclude its functional activity. Among *Peromyscus* species, the CENPB-box-like motifs found in the centromeric satDNA had between four and six conserved bases out of nine ECD nucleotides. It has been suggested that a divergent motif sequence may be required for functional activity in this group (Smalec et al., 2019), which could also be the case for manatees and the dugong. Divergent motif sequences have also been observed in the centromeric satDNAs of the African elephant (*L. africana*), nine-banded armadillo (*D. novemcintus*; Alkan et al., 2011) and in the two-toed sloths of the genus *Choloepus* (Sena et al., 2020).

The only genomes outside Sirenia in which the TMAsat sequence was found were those of the Order Proboscidea (elephants) and Hyracoidea (hyraxes), that together with Sirenia are reunited in Paenungulata, a subgroup of the Superorder Afrotheria (Foley et al., 2016). With only a few hits (with the short arrays mostly flanked by the transposable element L1), the TMAsat sequence is probably not a typical satDNA in these species. TMAsat in Sirenian probably evolved from these ancestral sequences still found in elephants and hyraxes, which could be the basis for both TMAsat1 and TMAsat2.

In the tree sirenians with sequenced genome, the most frequent organization of TMAsat arrays was the alternating TMAsat1 and TMAsat2 form. In the few exceptions, we found more consecutive TMAsat2 units than TMAsat1. Other satDNAs were found organized as a composite of two related units, mostly in the alternating form as TMAsat. This is the case of S1a-S1b in European brown frogs (Feliciello et al., 2006) and Tcast1a-Tcast1b in the red flour beetle *Tribolium castaneum* (Feliciello et al., 2011, 2014), in which the rolling circle amplification followed by substitutions by homologous recombination were proposed to explain the origin of the composite a-b arrays.

The sequenced genomes we used were generated from short reads (average 100–150 bp) that do not cover the total length of the monomeric unit of TMAsat, resulting in an assembly that may not represent well the long satDNA arrays. Further analyses with Southern blot and dot blot experiments as well as long-reads sequencing may help clarify the overall organization of repeats in the genome and within the long satDNA arrays. As an example, Vondrak et al. (2020) using ultra-long nanopore reads found nine out of 11 putative satDNA sequences derived from short tandem arrays located within LTR-retrotransposons that occasionally expanded in length, and just two organized in long arrays typical of satDNA. In addition, the long-reads sequencing approach proved a valuable contribution in determining the origin of the satDNAs. Several satDNAs from plants and animals derived from tandem amplification of internal segments of TEs (Dias et al., 2015; Meštrović et al., 2015; Vondrak et al., 2020), as was the case of TMAsat described herein, that could be L1 related.

In conclusion, we reported for the first time the centromeric satDNA in the West Indian manatee, which seems to be present across Sirenia, a group with all extant species under threat of extinction. TMAsat monomers from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* lack species-specific sequences, contradicting the predictions of concerted evolution. The TMAsat-like ancestral sequence is present in other Paenungulata, such as elephants and hyraxes, suggesting that TMAsat suffered an expansion within Sirenia less than ~69 Mya (Foley et al., 2016; de Souza et al., 2021), after the divergence of Sirenia from Proboscidea and Hyracoidea.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/genbank/, MW272776–MW272780.

ETHICS STATEMENT

The animal study was reviewed and approved by SISBio/ICMBio permit 60829-2.

AUTHOR CONTRIBUTIONS

MV and GD conceived and designed the experiments, analyzed the data, and contributed to writing – original draft preparation. MV and AE performed the experiments. CM, YY-Y, and IS obtained the materials for molecular and cytological analyses. MV, GD, AE, GK, and MS contributed to writing – review and editing. GK and MS contributed to supervision and project administration. MS contributed to funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

MV and AE received fellowships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). MS and GK are supported by fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Processes 310433/2018-5 and 308386/2018-3, respectively). This work was funded with grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

ACKNOWLEDGMENTS

We thank the Programa de Pós-Graduação em Genética, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, for their support in the execution and publishing of this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.694866/full#supplementary-material

REFERENCES

Alkan, C., Cardone, M. F., Catacchio, C. R., Antonacci, F., O'Brien, S. J., Ryder, O. A., et al. (2011). Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome Res.* 21, 137–145. doi: 10.1101/gr.111278.110

Alkan, C., Ventura, M., Archidiacono, N., Rocchi, M., Sahinalp, S. C., and Eichler, E. E. (2007). Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Comput. Biol.* 3, 1807–1818. doi: 10.1371/journal.pcbi.0030181

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

Assis, M. F. L., Best, R. C., Barros, R. M. S., and Yonenaga-Yassuda, Y. (1998). Cytogenetic study of *Trichechus inunguis* (Amazonian manatee). *Braz. J. Genet.* 11, 41–50.

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:11. doi: 10.1186/s13100-015-0041-9

Barros, H. M. D. R., Meirelles, A. C. O., Luna, F. O., Marmontel, M., Cordeiro-Estrela, P., Santos, N., et al. (2017). Cranial and chromosomal geographic variation in manatees (Mammalia: Sirenia: Trichechidae) with the description of the Antillean manatee karyotype in Brazil. *J. Zool. Syst. Evol. Res.* 55, 73–87. doi: 10.1111/jzs.12153

Biscotti, M. A., Olmo, E., and Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosom. Res.* 23, 415–420. doi: 10.1007/s10577-015-9499-z

Bonvicino, C. R., Viana, M. C., de Oliveira, E. H. C., Emin, R., Silva Junior, J. D. S., de Sousa, M. E. M., et al. (2020). Distribution of south American manatees, *Trichechus manatus* Linnaeus, 1758 and *T. inunguis* (Natterer, 1883) (Sirenia: Trichechidae). *Bol. Mus. Para. Emílio Goeldi, Ciênc. Nat.* 15, 573–599. doi: 10.46357/bcnaturais.v15i3.246

de Souza, É. M. S., Freitas, L., da Silva Ramos, E. K., Selleghin-Veiga, G., Rachid-Ribeiro, M. C., Silva, F. A., et al. (2021). The evolutionary history of manatees told by their mitogenomes. *Sci. Rep.* 11:3564. doi: 10.1038/s41598-021-82390-2

Deutsch, C. J., Self-Sulivan, C., and Mignucci-Giannoni, A. (2008). Data from: Trichechus manatus, West Indian Manatee. *IUCN Red List Threat. Species*, e.T22103A9356917. doi: 10.2305/IUCN.UK.2008.RLTS.T22103A9356917.en

Dias, G. B., Heringer, P., Svartman, M., and Kuhn, G. C. S. (2015). Helitrons shaping the genomic architecture of drosophila: enrichment of DINE-TR1 in α- and β-heterochromatin, satellite DNA emergence, and piRNA expression. *Chromosom. Res.* 23, 597–613. doi: 10.1007/s10577-015-9480-x

Domning, D. P. (2018). "Sirenian evolution," in *Encyclopedia of Marine Mammals*. eds. B. Würsig, J. G. M. Thewissen and K. M. Kovacs (London, United Kingdom: Academic Press), 856–859.

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299, 111–117. doi: 10.1038/299111a0

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

Feliciello, I., Akrap, I., Brajkovi, J., Zlatar, I., and Ugarkovic, D. (2014). Satellite DNA as a driver of population divergence in the red flour beetle Tribolium castaneum. *Genome Biol. Evol.* 7, 228–239. doi: 10.1093/gbe/evu280

Feliciello, I., Chinali, G., and Ugarković, D. (2011). Structure and population dynamics of the major satellite DNA in the red flour beetle Tribolium castaneum. *Genetica* 139, 999–1008. doi: 10.1007/s10709-011-9601-1

Feliciello, I., Picariello, O., and Chinali, G. (2006). Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from Rana dalmatina: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. *Gene* 383, 81–92. doi: 10.1016/j.gene.2006.07.016

Foley, N. M., Springer, M. S., and Teeling, E. C. (2016). Mammal madness: is the mammal tree of life not yet resolved? *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150140. doi: 10.1098/rstb.2015.0140

Fry, K., and Salser, W. (1977). Nucleotide sequences of HS-α satellite DNA from kangaroo rat dipodomys ordii and characterization of similar sequences in other rodents. *Cell* 12, 1069–1084. doi: 10.1016/0092-8674(77)90170-2

Gray, B. A., Zori, R. T., McGuire, P. M., and Bonde, R. K. (2002). A first generation cytogenetic ideogram for the Florida manatee (Trichechus manatus latirostris) based on multiple chromosome banding techniques. *Hereditas* 137, 215–223. doi: 10.1034/j.1601-5223.2002.01657.x

Kasai, F., O'Brien, P. C. M., and Ferguson-Smith, M. A. (2013). Afrotheria genome; overestimation of genome size and distinct chromosome GC content revealed by flow karyotyping. *Genomics* 102, 468–471. doi: 10.1016/j.ygeno.2013.09.002

Kasinathan, S., and Henikoff, S. (2018). Non-B-form DNA is enriched at centromeres. *Mol. Biol. Evol.* 35, 949–962. doi: 10.1093/molbev/msy010

Keith, L. D. (2015). Data from: Trichechus senegalensis (African Manatee). *IUCN Red List Threat. Species*, e.T22104A97168578. doi: 10.2305/IUCN.UK.2015-4.RLTS.T22104A81904980.en

Kohany, O., Gentles, A. J., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinform.* 7:474. doi: 10.1186/1471-2105-7-474

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Lima, C. S., Magalhães, R. F., Marmontel, M., Meirelles, A. C., Carvalho, V. L., Lavergne, A., et al. (2019). A hybrid swarm of manatees along the guianas coastline, a peculiar environment under the influence of the Amazon River plume. *An. Acad. Bras. Cienc.* 91:e20190325. doi: 10.1590/0001-3765201920190325

Lima, C. S., Magalhães, R. F., and Santos, F. R. (2021). Conservation issues using discordant taxonomic and evolutionary units: a case study of the American manatee (Trichechus manatus, Sirenia). *Wildl. Res.* 48:385. doi: 10.1071/WR20197

Luna, F. O., Beaver, C. E., Nourisson, C., Bonde, R. K., Attademo, F. L. N., Miranda, A. V., et al. (2021). Genetic connectivity of the west Indian manatee in the southern range and limited evidence of hybridization with Amazonian manatees. *Front. Mar. Sci.* 7, 1–15. doi: 10.3389/fmars.2020.574455

Marmontel, M., de Souza, D., and Kendall, S. (2016). Data from: Trichechus inunguis. *IUCN Red List Threat. Species*, e.T22102A43793736. doi: 10.2305/IUCN.UK.2016-2.RLTS.T22102A43793736.en

Marsh, H., and Sobtzick, S. (2019). Data from: Dugong Dugon (amended version of 2015 assessment). *IUCN Red List Threat. Species* 2019, e.T22104A97168578. doi: 10.2305/IUCN.UK.2015-4.RLTS.T22104A81904980.en

Masumoto, H., Nakano, M., and Ohzeki, J. I. (2004). The role of CENP-B and α-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. *Chromosom. Res.* 12, 543–556. doi: 10.1023/B:CHRO.0000036593.72788.99

Meštrović, N., Mravinac, B., Pavlek, M., Vojvoda-Zeljko, T., Šatović, E., and Plohl, M. (2015). Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosom. Res.* 23, 583–596. doi: 10.1007/s10577-015-9483-7

Meštrović, N., Plohl, M., Mravinac, B., and Ugarković, D. (1998). Evolution of satellite DNAs from the genus palorus-experimental evidence for the "library" hypothesis. *Mol. Biol. Evol.* 15, 1062–1068. doi: 10.1093/oxfordjournals.molbev.a026005

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., and Okazaki, T. (1992). Centromere protein B assembles human centromeric α-satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell Biol.* 116, 585–596. doi: 10.1083/jcb.116.3.585

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

Novák, P., Robledillo, L. Á., 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

Plohl, M., Meštrović, N., and Mravinac, B. (2012). "Satellite DNA evolution," in *Repetitive DNA. Genome Dyn. Vol. 7*. ed. M. A. Garrido-Ramos (Basel, Karger), 126–152.

Plohl, M., Petrović, V., Luchetti, A., Ricci, A., Šatović, E., Passamonti, M., et al. (2009). Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Hereditas* 104, 543–551. doi: 10.1038/hdy.2009.141

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Sena, R. S., Heringer, P., Valeri, M. P., Pereira, V. S., Kuhn, G. C. S., and Svartman, M. (2020). Identification and characterization of satellite DNAs in two-toed sloths of the genus Choloepus (Megalonychidae, Xenarthra). *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-76199-8

Shapiro, J. A., and von Sternberg, R. (2005). Why repetitive DNA is essential to genome function. *Biol. Rev.* 80, 227–250. doi: 10.1017/S1464793104006657

Sharko, F. S., Boulygina, E. S., Tsygankova, S. V., Slobodova, N. V., Alekseev, D. A., Krasivskaya, A. A., et al. (2021). Steller's sea cow genome suggests this species began going extinct before the arrival of Paleolithic humans. *Nat. Commun.* 12, 1–8. doi: 10.1038/s41467-021-22567-5

Shatskikh, A. S., Kotov, A. A., Adashev, V. E., Bazylev, S. S., and Olenina, L. V. (2020). Functional significance of satellite DNAs: insights from drosophila. *Front. Cell Dev. Biol.* 8:312. doi: 10.3389/fcell.2020.00312

Smalec, B. M., Heider, T. N., Flynn, B. L., and O'Neill, R. J. (2019). A centromere satellite concomitant with extensive karyotypic diversity across the Peromyscus genus defies predictions of molecular drive. *Chromosom. Res.* 27, 237–252. doi: 10.1007/s10577-019-09605-1

Stanyon, R., and Galleni, L. (1991). A rapid fibroblast culture technique for high resolution karyotypes. *Bolletino di Zool.* 58, 81–83. doi: 10.1080/11250009109355732

Sujiwattanarat, P., Thapana, W., Srikulnath, K., Hirai, Y., Hirai, H., and Koga, A. (2015). Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids. *Sci. Rep.* 5, 1–10. doi: 10.1038/srep10315

Sullivan, L. L., Chew, K., and Sullivan, B. A. (2017). α satellite DNA variation and function of the human centromere. *Nucleus* 8, 331–339. doi: 10.1080/19491034.2017.1308989

Valeri, M. P., Dias, G. B., Moreira, C. N., Yonenaga-Yassuda, Y., Stanyon, R., Silva Kuhn, G. C., et al. (2020). Characterization of satellite DNAs in squirrel monkeys genus Saimiri (Cebidae, Platyrrhini). *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-64620-1

Vianna, J. A., Bonde, R. K., Caballero, S., Giraldo, J. P., Lima, R. P., Clark, A., et al. (2006). Phylogeography, phylogeny and hybridization in trichechid sirenians: implications for manatee conservation. *Mol. Ecol.* 15, 433–447. doi: 10.1111/j.1365-294X.2005.02771.x

Vlahović, I., Glunčić, M., Rosandić, M., Ugarković, Đ., and Paar, V. (2016). Regular higher order repeat structures in beetle *Tribolium castaneum* genomex. *Hereditas* 104, 543–551. doi: 10.1093/gbe/evw174

Vondrak, T., Ávila Robledillo, L., Novák, P., Koblížková, A., Neumann, P., and Macas, J. (2020). Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. *Plant J.* 101, 484–500. doi: 10.1111/tpj.14546

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 September 2021
doi: 10.3389/fgene.2021.728664

Chromosomal Differentiation of *Deschampsia* (Poaceae) Based on Four Satellite DNA Families

María Laura González[1], *Jorge Oscar Chiapella*[2] *and Juan Domingo Urdampilleta*[1]*

[1]*Instituto Multidisciplinario de Biología Vegetal (Consejo Nacional de Investigaciones Científicas y Técnicas - Universidad Nacional de Córdoba), Córdoba, Argentina,* [2]*Instituto de Investigaciones en Biodiversidad y Medioambiente (Consejo Nacional de Investigaciones Científicas y Técnicas - Universidad Nacional Del Comahue), Bariloche, Argentina*

Diverse families of satellite DNA (satDNA) were detected in heterochromatin regions of *Deschampsia*. This kind of repetitive DNA consists of tandem repeat sequences forming big arrays in genomes, and can contribute to lineages differentiation. The differentiation between types of satDNA is related to their sequence identity, the size and number of monomers forming the array, and their chromosomal location. In this work, four families of satDNA (D2, D3, D12, D13), previously isolated by genomic analysis, were studied on chromosomal preparations of 12 species of *Deschampsia* (*D. airiformis*, *D. antarctica*, *D. cespitosa*, *D. cordillerarum*, *D. elongata*, *D. kingii*, *D. laxa*, *D. mendocina*, *D. parvula*, *D. patula*, *D. venustula,* and *Deschampsia sp*) and one of *Deyeuxia* (*D. eminens*). Despite the number of satDNA loci showing interspecific variation, the general distribution pattern of each satDNA family is maintained. The four satDNA families are AT-rich and associated with DAPI + heterochromatin regions. D2, D3, and D12 have mainly subterminal distribution, while D13 is distributed in intercalary regions. Such conservation of satDNA patterns suggests a not random distribution in genomes, where the variation between species is mainly associated with the array size and the loci number. The presence of satDNA in all species studied suggests a low genetic differentiation of sequences. On the other hand, the variation of the distribution pattern of satDNA has no clear association with phylogeny. This may be related to high differential amplification and contraction of sequences between lineages, as explained by the library model.

Keywords: Deschampsia, cytogenetics, repetitive DNA, satellite DNA, FISH

OPEN ACCESS

Edited by:
Magdalena Vaio,
Universidad de la República, Uruguay

Reviewed by:
Sonia Garcia,
Consejo Superior de Investigaciones Científicas (CSIC), Spain
Natalia Borowska-Zuchowska,
University of Silesia in Katowice, Poland

***Correspondence:**
Juan Domingo Urdampilleta
jurdampilleta@imbiv.unc.edu.ar

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *21 June 2021*
Accepted: *06 September 2021*
Published: *21 September 2021*

Citation:
González ML, Chiapella JO and Urdampilleta JD (2021) Chromosomal Differentiation of Deschampsia *(Poaceae) Based on Four Satellite DNA Families. Front. Genet. 12:728664. doi: 10.3389/fgene.2021.728664*

INTRODUCTION

Repetitive DNA is one of the main components of plant genomes, reaching about 50–90% of genomic abundance (Bennett et al., 1982; Kubis et al., 1998; Heslop-Harrison, 2000), and is composed by dispersed (e.g., transposable elements) and tandem (e.g., satellite DNA) sequences, with variable abundance. Since satellite DNA (satDNA) usually form large arrays on chromosomes (frequently related to heterochromatin regions), it is detectable by fluorescence *in situ* hybridization (FISH) (Schwarzacher and Heslop-Harrison, 2000), allowing their study at the cytogenetical level. The main characteristics defining satDNA are monomer size over 100–150 bp and arrays up to 100 Mb with tandem disposition (Sharma and Raina, 2005; Hemleben et al., 2007; Mehrotra and Goyal, 2014). Although they are considered non-coding sequences, their monomeric size frequently varies between 150–180 and 320–360 bp, which corresponds with the structural motifs of mono- and

dinucleosomes (Kubis et al., 1998; Macas et al., 2002). The current genomic sequencing methods (e.g., NGS), together with the advance of bioinformatics, constantly provide new data about the structural diversity of satDNA (Ruiz-Ruano et al., 2016; Garrido-Ramos, 2017; Lower et al., 2018), however many aspects remain to be understood.

The evolution of satDNA in genomes implies that both sequences diverge as changes in copy numbers. The abundance of satDNA in eukaryote genomes can vary widely and rapidly between generations, leading to high polymorphism in the satellite arrays' length (Plohl et al., 2012). The sequence identity inside an array evolves according to the process called concerted evolution, which causes more similar monomers than expected due to random changes (Dover, 1986; Plohl, 2010). The monomers homogenization is much faster in species with sexual reproduction, given meiotic recombination (Mantovani et al., 1997). The chromosome organization has a fundamental influence on processes such as chromosome pairing, segregation, gene organization, and expression, and repetitive DNA such as satDNA have an important role in DNA packaging and chromatin condensation (Heslop-Harrison, 2000). Since satDNA distribution may facilitate the recognition of homologous chromosome pairs, changes between lineages have been precursors of speciation (Hemleben et al., 2007). The similarities and differences in genomic satDNA between species can be explained by the "library model", which suggests differential amplification of satDNA between independent lineages (Garrido-Ramos, 2015). However, the patchy distribution of some satDNA types across eukaryotes (animal and plants) suggests that a scenario of multiple horizontal transfers during evolution may be considered (Yang et al., 2020).

The grass genus *Deschampsia* P. Beauv. is a cosmopolitan genus which includes about 30 species, 15 of them growing in South America, including *D. antarctica* which also occurs in Antarctica (Parodi, 1949; Chiapella and Zuloaga, 2010). Some species have difficult circumscription, forming species complexes (Chiapella et al., 2011; Tzvelev et al., 2015). Regarding *Deschampsia* cytogenetics, the species show a basic chromosome number of x = 13, with a few exceptions reported in the northern hemisphere. The most common chromosome number is 2n = 26, followed by 2n = 52, and the chromosomal complement has metacentric, submetacentric, and acrocentric chromosomes, in similar proportions between species (Albers, 1980; Winterfeld and Röser, 2007b; Cardone et al., 2009; González et al., 2021). In contrast with the conserved chromosome morphology, the *18-5.8-26S* and *5S* rDNA patterns have high intra and inter-specific variability of loci number and position, which allow the determination of chromosome markers specific to some phylogenetically related species groups (Chiapella, 2007; Saarela et al., 2017; González et al., 2021).

Several satDNA families were isolated from tribe Poeae using restriction enzymes (Grebenstein et al., 1995, 1996), and some of them are present in various genera, including *Deschampsia*, suggesting an ancient origin (Röser et al., 2014). Likewise, the occurrence of several specific satDNA types were reported in different groups of grasses (Anamthawat-Jónsson and Heslop-Harrison, 1993; Vershinin et al., 1995; Nakajima et al., 1996). The analysis of repetitive DNA from WGS in *D. antarctica* and *D. cespitosa* allowed researchers to recognize 34 satDNA families (González et al., 2020). This is a high satDNA diversity compared to other grasses, such as *Eragrostis tef* (one satDNA family), *Agropyron cristatum* (fourteen satDNA families), *Festuca pratensis* Huds. (eight satDNA families), and *Poa* (four satDNA families) (Gebre et al., 2016; Křivánková et al., 2017; Said et al., 2018; Wei et al., 2020). *Deschampsia antarctica* and *D. cespitosa* showed a slightly differentiated satDNA genomic pattern when considering abundance and diversity. Two types of chromosome distribution patterns were observed in the analyzed satDNA families, a C-type pattern (clustered) and M-type pattern (mixed combination of both clustered and dispersed) (González et al., 2020).

The use of C-type satDNA in *Deschampsia* would allow us to reveal if there exists interspecific variation of the distribution of such sequences in chromosomes, which may be related to the speciation processes. In this way, D2 (359 pb), D3 (377 pb), D12 (366 pb), and D13 (563 pb) families previously analyzed in *D. antarctica* and *D. cespitosa* (González et al., 2020), could be good markers to analyze the interspecific variation. The phylogenetic relationships between the South American species were inferred with molecular data and suggest a recent common ancestor between *Deschampsia* and *Deyeuxia* sec. *Stylagrostis*, which could have had chromosomal characteristics currently shared by both taxa (Saarela et al., 2017; González et al., 2021). Due to phylogenetic results, Saarela et al. (2017) suggested transferring seven South American species of *Deyeuxia* sect. *Stylagrostis* to *Deschampsia*, hence more studies of new sources of variation will be useful. With the aim of evaluating karyotypic variations related to phylogenetic hypotheses we analyzed chromosomal distribution of some satDNA families in twelve different *Deschampsia* species and one of *Deyeuxia eminens*.

MATERIAL AND METHODS

Plant Material

We collected twelve species of *Deschampsia* (one species, vouchers MLG 79 and MLG 81, does not match any described species), *Deyeuxia eminens* (which belongs to sect. *Stylagrostis*), and *Avenella flexuosa* from Antarctica and Argentinian Patagonia (**Table 1**). Living plants were transported to the laboratory and kept in pots in culture chambers at 14°C, to obtain root tips for cytogenetic techniques. Leaves were kept in silica gel for DNA extractions. The vouchers were included in the collection of the herbarium of the Botanical Museum of Córdoba (CORD).

Cytogenetic Techniques

The mitotic chromosomes preparations were obtained from root meristems pretreated with 2 mM 8-hydroxyquinoline for 4–6 h at 14°C and fixed in ethanol/acetic acid (3:1, v:v). The tissues were digested with Pectinex enzyme solution (Novozimes) and squashed in 45% acetic acid. Preparations were frozen in liquid nitrogen to remove the coverslip.

TABLE 1 | Studied species and results from FISH with satDNA. CORD: herbarium number (MLG: Maria Laura González; JC: Jorge Chiapella; JDU: Juan Domingo Urdampilleta; MG: Melisa Giorgis). 2n: somatic chromosome number. N: number of hybridization signals. P: hybridization pattern (st: subterminal, i: intercalary, *m*: metacentric, *sm*: submetacentric, *a*: acrocentric). Arg: *Argentina*.

Species	CORD	Locality	2n	D2		D3		D12		D13	
				N	P	N	P	N	P	N	P
D. airiformis	MLG 41	Arg, Chubut, Tehuelches	26	6	4 st *m*; 2 st *a*	4	2 st *m*; 2 st *a*	–	–	–	–
D. antarctica	JC 2775	Antarctic, 25 de Mayo Island	26	8	6 st *m*; 2 i *a*	12	10 st *m*; 4 st *a*	4	2–6 st *m*; 0–2 st *a*	2	2 i *a*
D. cespitosa	MLG 35	Arg, Chubut, Río Senguer	26	6	6 st *m*	7	6 st *m*; 1 st *sm*	4	4 st *m*	2	2 i *a*
	JDU 850	Arg, Chubut, Rio Senger	26	6	4 st *m*; 2 i *a*	–	–	–	–	–	–
	JDU 823	Arg, Mendoza, Malargüe	26	8	5 st *m*; 2 st *sm*; 1 st *a*	7	7 st *m*	–	–	–	–
D. cordillerarum	MLG 69	Arg, Mendoza, Las Heras	26	11	3 st *m*; 2 st *sm*; 2 i *a*; 4 st *a*	12	6 st *m*; 2 st *sm*; 4 st *a*	–	–	–	2 i *a*
D. elongata	MLG 56	Arg, Chubut, Languiñeo	26	12	8 st *m*; 4 i *a*	8	8 st *m*	10	8 st *m*; 2 i *sm*	2	2 i *a*
D. kingii	JDU 842	Arg, Chubut, Languiñeo	52	10	8 st *m*; 2 st *a*	20	18 st *m*; 2 st *a*	6	4 st *m*; 2 st *a*	2	2 i *a*
D. laxa	MLG 114	Arg, Rio Negro, Bariloche	26	10	7 st *m*; 2 i *a*: 1 i/st *a*	7	7 st *m*	4	4 st *m*	2	2 i *a*
D. mendocina	MLG 91	Arg, Mendoza, Malargüe	26	12	6 st *m*; 2 st *sm*; 2 i *a*; 2 st *a*	12	8 st *m*; 2 st *sm*; 2 st *a*	–	–	–	–
	MLG 101	Arg, Mendoza, Malargüe	26	–	–	–	–	4	4 st *m*	2	2 i *a*
D. parvula	MLG 48	Arg, Chubut, Languiñeo	26	6	4 st *m*; 2 i *a*	8	4 st *m*; 2 st *sm*; 2 st *a*	6	4 st *m*; 2 st *a*	2	2 i *a*
D. patula	JDU 878	Arg, Santa Cruz, Güer Aike	26	6	4 st *m*; 2 i *a*	10	8 st *m*; 2 st *a*	–	–	–	–
D. venustula	MLG 62	Arg, Neuquén, Picunches	26	6	2 st *m*; 4 i *a*	4	2 st *m*; 2 st *sm*	4	4 st *m*	2	2 i *a*
Deschampsia sp	MLG 81	Arg, Mendoza, San Carlos	26	21	10 st *m*; 2 st *sm*; 4 i *a*; 5 st *a*	15	10 st *m*; 2 st *sm*; 3 st *a*	3	3 st *m*	2	2 i *a*
Deyeuxia eminens	MG 1835	Arg, Córdoba, Calamuchita	26	9	1 st *m*; 1 st *sm*; 4 i *a*; 1 st *a*; 1 i/st *a*	11	5 st *m*; 2 st *sm*; 4 st *a*	–	–	–	–
A. flexuosa	JDU 848	Arg, Chubut, Languiñeo	28	0	–	0	–	–	–	–	–

Fluorescence *in situ* hybridization (FISH) was carried out to detect satellite DNA patterns, following protocols by Schwarzacher and Heslop-Harrison (2000). To reach stringency above 76%, the hybridization mix had 2x SSC, 50% v/v Formamide, 20% v/v dextran sulfate, 0.1% v/v SDS, and 4–6 ng/μL probes. Post-hybridization washes consisted of 2x SSC, 0.1x SSC, 2× SSC, for 10 min at 42 °C each. We selected four families with a conspicuous hybridization pattern among the previously isolated and mapped satDNA families in *D. antarctica* and *D. cespitosa* to analyze chromosomal distribution in other *Deschampsia* species and related taxa: D2, D3, D12, and D13. Probes were obtained by PCR as described in González et al. (2020) and labelled with biotin (Bionick, Invitrogen) (D2 and D13) or digoxigenin (D3 and D12) (DIG Nick translation mix, Roche). Double-target FISH was performed by hybridizing D2 with D3, and D12 with D13. The satDNA D2 and D3 were mapped in 14 species: *D.* cfr. *airiformis*, *D. antarctica*, *D. cespitosa* (three localities were used), *D. cordillerarum*, *D. elongata*, *D. kingii*, *D.* cfr. *laxa*, *D.* cfr. *mendocina*, *D. parvula*, *D. patula*, *D. venustula*, *Deschampsia* sp, *A. flexuosa,* and *Deyeuxia eminens*. The satDNA D12 and D13 were mapped in nine species: *D. antarctica*, *D. cespitosa*, *D. elongata*, *D. kingii*, *D. laxa*, *D. mendocina*, *D. parvula*, *D. venustula*, and *Deschampsia sp*. The detection was made with Avidin-FITC (Sigma-Aldrich) and anti-DIG antibodies conjugated with rhodamine (Roche). Chromosome metaphases were photographed using an Olympus BX61 microscope coupled with a monochromatic camera and Cytovision software (Leica Biosystems), and the images were pseudo-colored. Idiograms were constructed for *Deschampsia* species and *Deyeuxia eminens*. The karyotype conformation of each species was taken from González et al. (2021).

Reconstruction of Ancestral States of Chromosomal Traits

For the reconstruction of ancestral states of satDNA chromosomal traits, the packages *ape* (Paradis et al., 2004), *geiger* (Harmon et al., 2008), and *phytools* (Revell, 2012) were used in R. The phylogenetic hypothesis used was the maximum likelihood tree reconstructed with *ITS*, *ETS,* and *trn*L-F markers by González et al. (2021), which includes the same plant vouchers used for the cytogenetic analysis of this study. The ER (equal transition rates) and ARD (all transition rates are different) models were tested for discrete traits. For continuous traits, the models BM (Brownian motion), OU (Ornstein-Uhlenbeck), and EB (Early-burst) were tested. The models were fitted using the package *geiger*, and the best models were selected according to AIC criterion for each trait. The reconstruction of ancestral states of discrete traits was carried out using maximum likelihood, with the Markov k-state (Mk1) model according to the better-adjusted transition model using the package *ape*. The reconstruction of ancestral states of continuous traits was carried out using the package *phytools*. The phylogenetic signal was tested using the package *geiger*.

RESULTS

All studied species showed hybridization to all studied probes, except for *A. flexuosa* which did not show hybridization for D2

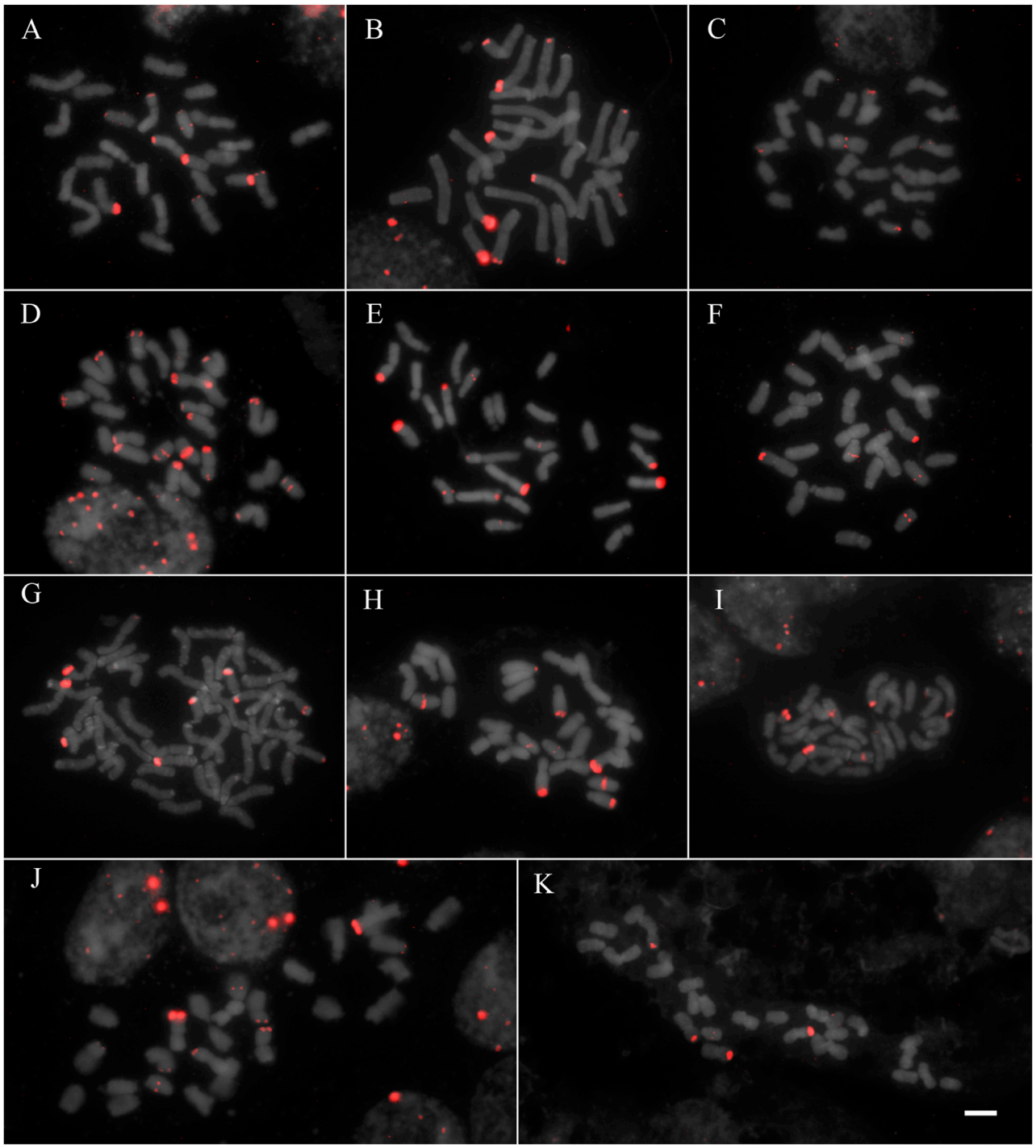

FIGURE 1 | FISH of satDNA D2 (red) of studied species **(A)** *D. cordillerarum* MLG 69 **(B)** *D. mendocina* MLG 91 **(C)** *D. airiformis* MLG 41 **(D)** *Deschampsia* sp MLG 81 **(E)** *D. elongata* MLG 56 **(F)** *D. venustula* MLG 62 **(G)** *D. kingii* JDU 842 **(H)** *Deyeuxia eminens* MG 1835 **(I)** *D. patula* 878 **(J)** *D. laxa* MLG 114 **(K)** *D. parvula* MLG 48. Scale: 5 μm.

and D3 satDNA probes (**Table 1**). The four satDNA families studied here are AT-rich and were frequently observed in association with DAPI + bands observed after FISH (**Supplementary Figure S1, S2, S3, and S4**). The species showed variation of the number and position of loci. The satDNA D2 showed variation of the loci number, ranging from 6 to 21 hybridization signals between species (which corresponded from 2.5 to 10.5 signals per basic complement

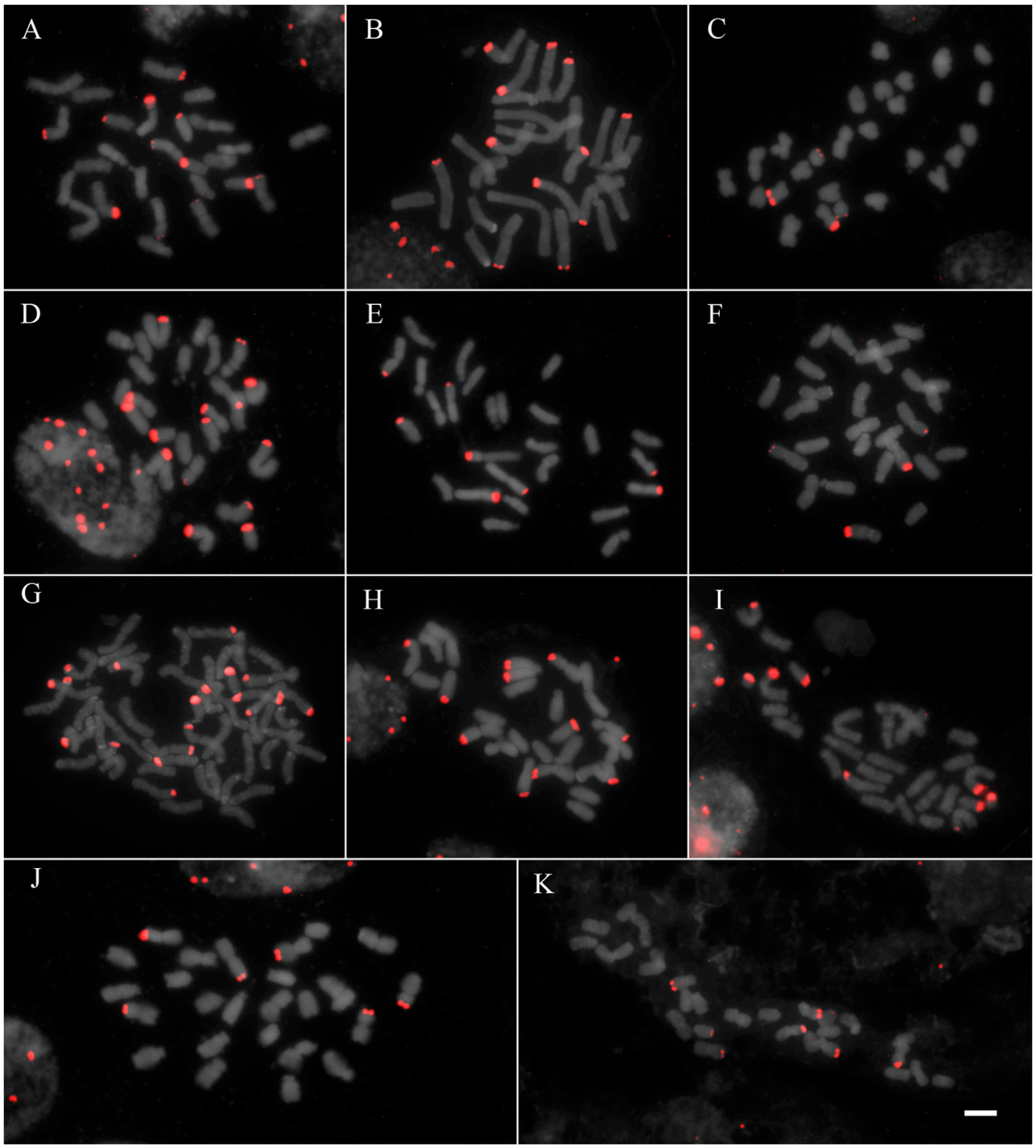

FIGURE 2 | FISH of satDNA D3 (red) of studied species **(A)** *D. cordillerarum* MLG 69 **(B)** *D. mendocina* MLG 91 **(C)** *D. airiformis* MLG 41 **(D)** *Deschampsia* sp MLG 81 **(E)** *D. elongata* MLG 56 **(F)** *D. venustula* MLG 62 **(G)** *D. kingii* JDU 842 **(H)** *Deyeuxia eminens* MG 1835 **(I)** *D. patula* 878 **(J)** *D. laxa* MLG 114 **(K)** *D. parvula* MLG 48. Scale: 5 μm.

x), noting that most species showed six hybridization signals (**Figure 1**; **Table 1**). The satDNA D3 showed from 4 to 20 hybridization signals (from 2 to 7.5 per basic complement x) (**Figure 2**; **Table 1**). The satDNA D12 showed from 3 to 10 hybridization signals between species (from 1.5 to five per basic complement x), noting that most species showed four hybridization signals (**Figure 3**; **Table 1**). The satDNA D13 was constant in all studied species, showing 2 hybridization

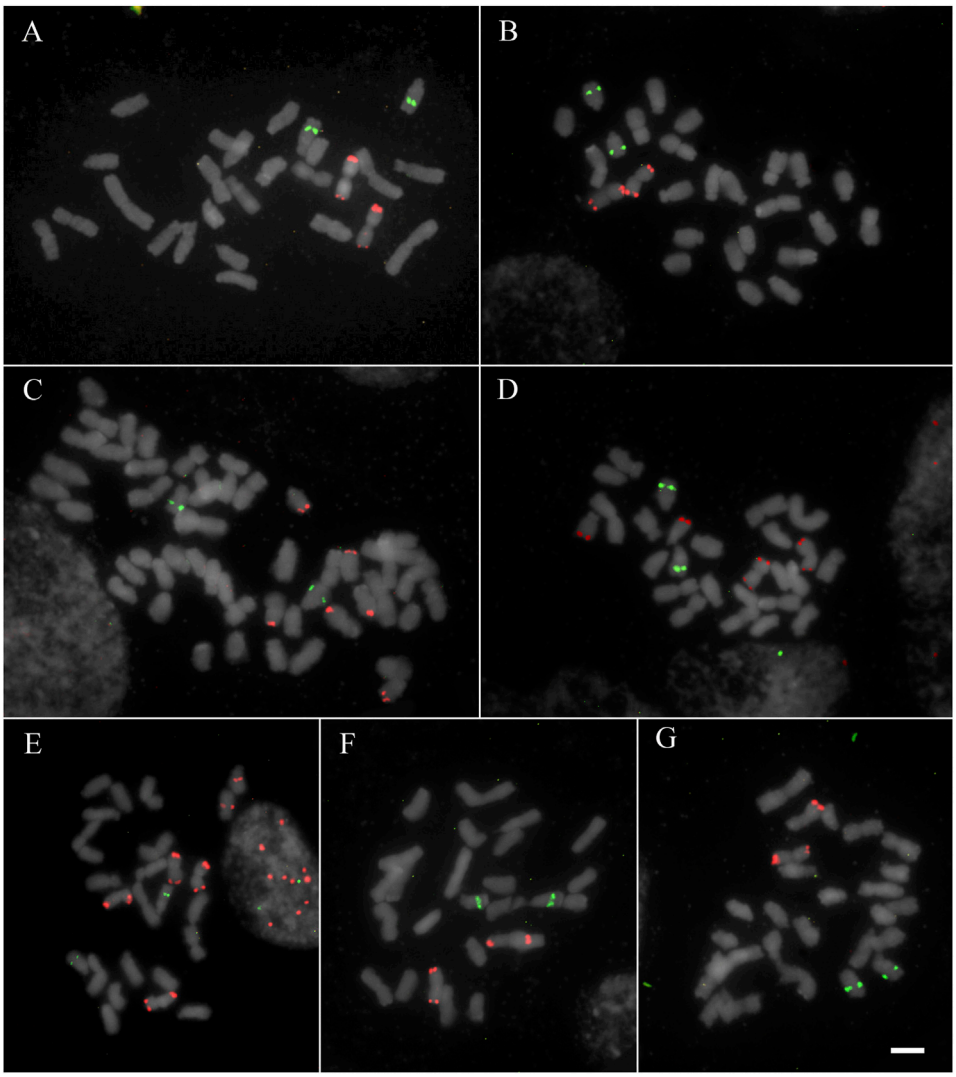

FIGURE 3 | FISH of satDNAs D12 (red) and D13 (green) of studied species **(A)** *D. venustula* MLG 62 **(B)** *D. laxa* MLG 114 **(C)** *D. kingii* JDU 842 **(D)** *D. parvula* MLG 48 **(E)** *D. elongata* MLG 56 **(F)** *D. mendocina* MLG 101 **(G)** *Deschampsia* sp MLG 81. Scale: 5 μm.

signals (corresponding with 0.5–1 locus per basic complement x, since there was one tetraploid species) (**Figure 3**; **Table 1**). Only considering D2 and D3, since they were hybridized in all species, the species with more satDNA loci per basic complement was *Deschampsia sp*, followed by *D. mendocina* and *D. cordillerarum* (**Table 1**). Likewise, *D. elongata* was the species with the highest number of signals observed for D12.

The satDNA family D2 was mostly located at subterminal regions of different chromosome types, forming blocks of variable size. All species showed subterminal loci of D2 on metacentric chromosomes, and the species group formed by *D. cordillerarum*, *D. mendocina*, *Deyeuxia eminens,* and *Deschampsia sp* showed subterminal loci on submetacentric chromosomes. Also, some loci were found in acrocentric chromosomes, at an intercalary position in most species (except for *D. kingii* and *D. airiformis*) and at a subterminal position for *D. cordillerarum*, *D. mendocina*, *Deyeuxia eminens*, *Deschampsia sp*, *D. airiformis,* and *D. kingii* (**Figure 1** and **Supplementary Figure S4**).

The satDNA family D3 was exclusively observed at subterminal regions. All species showed D3 loci on metacentric chromosomes, and most species also showed on the long arm of submetacentric chromosomes except for *D. laxa*, *D. airiformis*, *D. elongata*, *D. antarctica,* and *D. patula*. The satDNA D3 was also observed on acrocentric chromosomes at a subterminal position in most species, except for *D. cespitosa*, *D. laxa*, *D. venustula,* and *D. elongata*. This satDNA was frequently co-localized with D2 (**Figures 1, 2** and **Supplementary Figure S1, S2, S3, and S4**).

The satDNA family D12 was detected at a subterminal position on metacentric chromosomes in all studied species. Particularly for D12, most of the studied species showed metacentric chromosome pairs with hybridization signals in

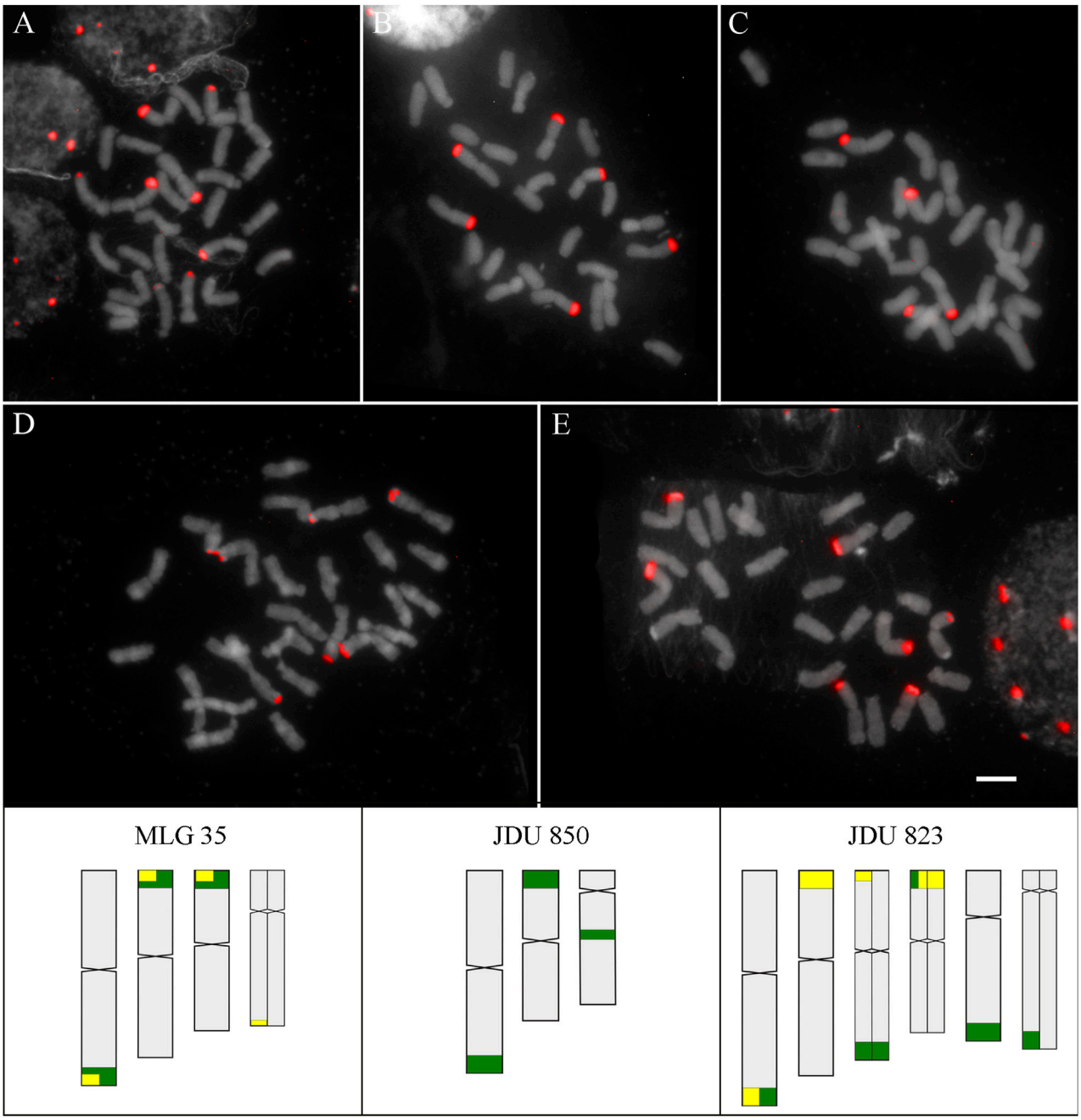

FIGURE 4 | SatDNA D2 and D3 in *D. cespitosa*. Above: FISH of satDNA D2 and D3 of *D. cespitosa* **(A)** D2 of JDU 823 **(B)** D2 of MLG 35 **(C)** D2 of JDU 850 **(D)** D3 of MLG 35 **(E)** D3 of JDU 823. Scale: 5 μm. Below: Idiograms of the three localities, showing the position of satDNA D2 (green) and D3 (yellow). Heteromorphic chromosome pairs are shown divided into two.

both arms (*symmetric chromosome*). In general, the species showed only one chromosome pair with satDNA D12 (in one or two chromosomal arms), except for *D. kingii*, *D. antarctica,* and *D. elongata*, which showed three or four pairs. This satDNA was also detected in acrocentric chromosomes at a subterminal position in *D. antarctica*, *D. parvula,* and *D. kingii*. *Deschampsia elongata* was the only species showing one locus of D12 on submetacentric chromosomes (**Figure 3** and **Supplementary Figure S4**).

The chromosomal distribution of satDNA family D13 was conserved in the analyzed species (**Figure 3** and **Supplementary Figure S4**). We observed only one locus at an intercalary position on an acrocentric chromosome pair for all the species, inclusive for the tetraploid *D. kingii*.

The hybridization pattern of D2 and D3 satDNA was studied in three and two localities of *D. cespitosa*, respectively, revealing intraspecific variation. The satDNA D2 showed six loci in two localities separated by approximately 26 km (JDU 850 and MLG 35), and eight loci for the most distant locality (JDU 823). The satDNA D3 showed three loci for both studied localities (JDU 823 and MLG 35). In addition to the numerical variation, the disposition of loci showed variation for both probes in all localities (**Table 1**; **Figure 4**).

For the reconstruction of ancestral states of satDNA, we used a total of six chromosomal traits, two discrete and four continuous (**Supplementary Table 1**). None of these showed a significant phylogenetic signal according to parameter λ (Pagel, 1999), despite some of them showing values of 1. For discrete traits, the analysis with ER (the better fitted model) did not reflect high probabilities on the ancestral state (presence/absence) in most of the nodes, except in common ancestors of small clades. The continuous traits fitted with a BM model, and does not suggested any trend in the number of satDNA loci in *Deschampsia*. However, increases or decreases of loci can be seen for small groups or species, as an increase of satDNA D3 in the *Deschampsia* sp/*D. cordillerarum*/*D. mendocina* clade and the *D. antarctica*/*D. patula* clade. Particularly, *Deschampsia* sp shows an increase in the number of signals for both satDNA D2 and D3, while *D. elongata* shows an increase in D12. The satDNA D13 was constant in terms of loci number per basic complement, with the exception of a decrease in *D. kingii*, a polyploid species (**Supplementary Figure S5**).

DISCUSSION

Constitutive heterochromatin is an important genomic component, essential in nuclear architecture, DNA repair, and genome stability (Strom et al., 2017). The first analyses of heterochromatin in *Deschampsia* were made on *D. cespitosa* from Europe, by the use of the C-band technique, which detected big heterochromatin blocks at subterminal regions of most chromosomes and at intercalary regions of a few chromosomes (Garcia-Suarez et al., 1997), a frequent feature of several genera of tribe Poeae. Heterochromatin is commonly composed of several kinds of tandem repetitive sequences (satDNA), usually with unknown origin and function, considered as fast-changing genomic components (Fuchs et al., 1994; Kuipers et al., 2002; Winterfeld and Röser, 2007a). The variation of such repetitive DNA between related species could provide information about its evolution, and clarify the phylogenetic relationships between lineages (Beckmann and Soller, 1986; King et al., 1995). Highly repetitive sequences were detected for the first time in *D. cespitosa* by membrane hybridization of four satDNA previously isolated from *Helictotrichon* (which belong to the same tribe, Poeae) and were reported in several other grasses (CON1, CON2, COM1, and COM2) (Grebenstein et al., 1995, 1996). Further studies of the satellitome corroborated the presence of these satDNAs in South American *D. cespitosa* and *D. antarctica* (except for COM2), as well as described another 31 satDNA families, with two kinds of hybridization patterns: "mainly clustered", and mixed "clustered and dispersed" (González et al., 2018, 2020). The satDNA here studied correspond with the pool of clustered satDNA, of which D2 and D3 are the most abundant satDNA of *D. antarctica* and *D. cespitosa* genomes, while D12 and D13 (homologous with CON1 and CON2, respectively) are commonly detected in the genome of several Poaceae (Alix et al., 1998; Röser et al., 2014; Zhao et al., 2017).

As originally described, the species of *Deschampsia* are characterized by their appearance in subterminal regions of AT-rich heterochromatin (as observed with DAPI staining post-FISH), and these regions also show several kinds of clustered satDNA as we found here for D2, D3, and D12 (Garcia-Suarez et al., 1997; Winterfeld and Röser, 2007a; Amosova et al., 2017; González et al., 2020, 2021). The presence of AT-rich repetitive sequences has been frequently detected at the terminal region of other grass genomes (Flavell, 1986; Kubis et al., 1998; Winterfeld and Röser, 2007a; Hemleben et al., 2007), and such repetitive DNA accumulation may be caused by the low recombination rate in the telomeric chromosome regions (Charlesworth et al., 1994). On the other hand, intercalary heterochromatic bands are less common in *Deschampsia* chromosomes, but they may also coincide with satDNA. Here we found intercalary distribution for D13 and D2 in acrocentric chromosomes only, however, previous studies showed that chromosome intercalary regions in *Deschampsia* are also rich in dispersed satDNA (González et al., 2020).

The variability of loci number and chromosome distribution of repetitive DNA between and within species is common for *Deschampsia*. Studies of rDNA distribution (*18-5.8-26S* and *5S*) have shown chromosomal variation related with phylogeny of species and geographic distances (González et al., 2016, 2021). Here we found high variation of the distribution pattern of satDNA D2, D3, and D12 between *Deschampsia* species. Particularly, D13 was the only satDNA which did not show any variation in its distribution between the studied species. Even *D. kingii*, which is tetraploid with 52 chromosomes, showed only one locus of D13 in the same position and chromosome type as the rest of the species. Previous observations between *D. antarctica* and *D. cespitosa* showed the same hybridization patterns only for two of the thirteen studied satDNA families, D13 and D14 (homologous to satDNA pSc200, widely spread in tribes Poeae and Triticeae), which indicates that this conservation of the distribution pattern is not the most frequently observed in satDNA families (Bedbrook et al., 1980; González et al., 2020). Despite the fact that satDNA could enhance the phylogenetic information of species (Plohl et al., 2012), we found a lack of association between phylogenetic relationships of *Deschampsia* and the chromosomal distribution and the loci number of satDNA. The genomic abundance of satDNA changes widely and rapidly between generations, leading to polymorphism of arrays (Plohl et al., 2012), and therefore some variation between and within species may be expected.

Here we report different hybridization patterns of D2 and D3 between localities of *D. cespitosa* from South America. The intraspecific variation of chromosomal distribution of repetitive DNA has been previously found for *Deschampsia*, by

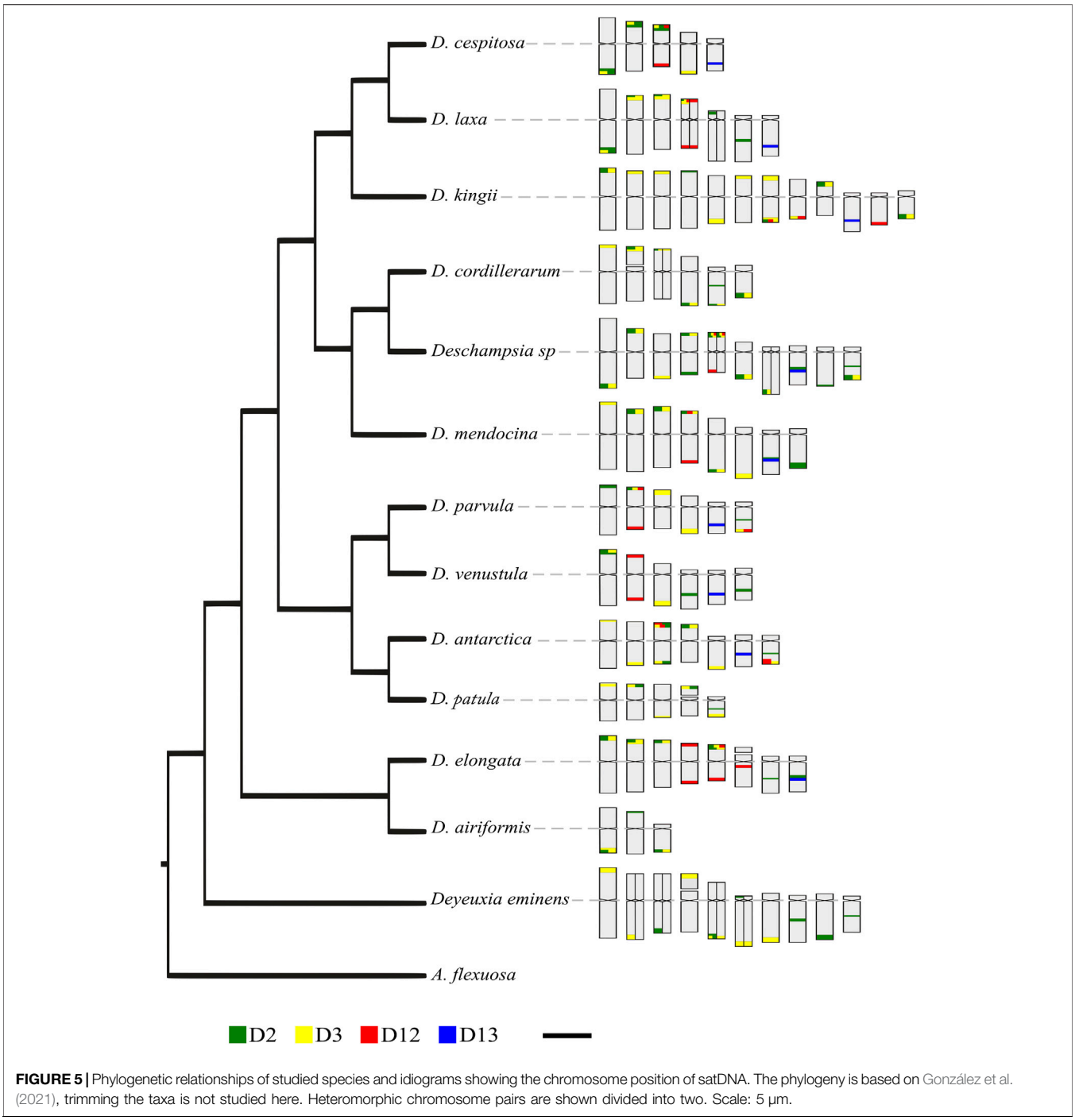

FIGURE 5 | Phylogenetic relationships of studied species and idiograms showing the chromosome position of satDNA. The phylogeny is based on González et al. (2021), trimming the taxa is not studied here. Heteromorphic chromosome pairs are shown divided into two. Scale: 5 µm.

studying the rDNA and satDNA in several localities of *D. antarctica,* and rDNA in localities of *D. cespitosa* (Winterfeld and Röser, 2007b; González et al., 2016, 2018, 2021; Navrotska et al., 2018). In both cases, this variability seems to be related with geographic distances since the nearest localities share more loci with each other, and satDNA showed more variation than rDNA. On the other hand, the Patagonian localities of other species, such as *D. elongata* and *D. parvula* also showed different hybridization patterns of rDNA than localities from other world regions (United States and Falkland Islands, respectively) (Amosova et al., 2017, 2021).

The intraspecific chromosomal variability of satDNA could explain the lack of association between chromosomal changes and phylogenetic relationships observed in *Deschampsia* species (**Figure 5**). Since the chromosome morphologies and karyotype composition remain highly conserved in *Deschampsia* (Kawano, 1963; Albers, 1980; Winterfeld and Röser, 2007b; Cardone et al., 2009; González et al., 2016, 2021), the high variability of

sequences distribution is likely a consequence of differential amplification and loss of repetitive DNA loci between lineages (Fry and Salser, 1977; Ruiz-Ruano et al., 2016). Also small rearrangements can be involved, without notable changes in chromosome numbers and morphologies (Raskina et al., 2008; Nani et al., 2016). The divergent evolution of satDNA may be involved in micro and macroevolutionary processes, facilitating the reproductive isolation of groups by the emergence of chromosomal barriers, eventually giving rise to speciation (Noor et al., 2001; Widmer et al., 2009; Plohl et al., 2012). At the same time, the eventual secondary contacts of lineages with a differentiated pattern of repetitive DNA could give rise to allopolyploids (Mallet, 2007; Soltis and Soltis, 2009; González et al., 2016).

Despite the fact that most repetitive DNA are considered as fast evolutionary elements, mainly in abundance but also in sequence, some satDNA may persist over long evolutionary periods (Plohl, 2010). This may be the case of several satDNAs found in *Deschampsia* which show homology with sequences previously described in Poaceae (Röser et al., 2014; González et al., 2020). The homology of D12 and D13, with the previously described CON1 and CON2 was of 74 and 78% (González et al., 2018). The repetitive DNA shared between distant Poaceae tribes or genus may be related with an ancestral repetitive DNA pool in grasses which evolved differentially between lineages (Hemleben et al., 2007). This is known as the "library model", and implies that the satDNA profile of a species is the result of differential amplification or contraction from a sequences set of an ancestral genome (Fry and Salser, 1977; Plohl et al., 2012). Some satDNAs may persist in genomes as low copy number sequences over long evolutionary periods and eventually amplify giving rise to high repetitive DNA (Ugarkovic, 2008; Röser et al., 2014). Other satDNA can be lost in some lineages as a consequence of an unequal exchange, which would eventually cause a single copy sequence and then be lost by drift (Charlesworth et al., 1986). One possible reason for low evolutionary rates is the preference of some monomers over others due to their potential function (Plohl, 2010; Plohl et al., 2012). Also, the heterochromatic environment can cause an extreme conservation of the sequence due to low recombination frequency, which together with the concerted evolution would cause a prolonged persistence of repetitive sequences throughout evolutionary time (Charlesworth et al., 1986).

The satDNA D12 (CON1) and D13 (CON2) could have had an early appearance in Poaceae; specifically, D13 seems to belong exclusively to tribe Poeae, while D12 is widespread in Poaceae including members of subfamilies Pooideae, Panicoideae, Chloridideae, and Oryzoideae (Grebenstein et al., 1996; Alix et al., 1998; Winterfeld and Röser, 2007a; Röser et al., 2014; Zhao et al., 2017). The other two satDNA here studied, D2 and D3, seem to be exclusively of *Deschampsia* lineage (González et al., 2020). Since we found D2 and D3 in all studied *Deschampsia* and *Deyeuxia eminens*, it is likely these satDNAs originated in a recent common ancestor between these two groups. As has been previously found, the sect. *Stylagrostis* from *Deyeuxia* (where *D. eminens* belong) seems to be highly related with *Deschampsia* (Saarela et al., 2017; González et al., 2021), and the presence of D2 and D3 in the *D. eminens* genome supports their monophyletic relationship.

Despite the chromosomal distribution variability of repetitive DNA (González et al., 2020, 2021), the presence of a clear hybridization for all probes in all *Deschampsia* species indicates likely conservation of the monomer sequence in the group. This can be a consequence of low evolutionary rates of the studied satDNA, or a recent species radiation. Either way, since satDNA are considered rapid evolution DNA, the presence of all studied satDNA in *Deschampsia* and *Deyeuxia eminens* genomes reinforce the idea of a low genetic differentiation between species (in terms of DNA sequences), as previously found in another studies of this group (Chiapella, 2007; Chiapella et al., 2011; Fasanella et al., 2017; González et al., 2020, 2021). In fact, previous studies suggest that karyotype differentiation between *Deschampsia* species is more associated with changes in chromosomal distribution than changes in the DNA sequence of satDNA (González et al., 2020; Amosova et al., 2021).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MG designed the study, made the experiments, and write the manuscript. JU designed, supervised the study, and review the manuscript. JC review the manuscript.

ACKNOWLEDGMENTS

We thank CONICET, ANPCyT-FONCyT, and SECyTUNC for financial support; and Dirección Nacional del Antártico for logistic support in Antarctica. This work was supported by ANPCyT-FONCyT under project PICT-2018–1691; the International Association for Plant Taxonomy (IAPT) by 2018 Research Grants; and CONICET under project PIP-11220150100392CO.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.728664/full#supplementary-material

REFERENCES

Albers, F. (1980). Vergleichende Karyologie der Gräser-SubtribenAristaveninae undAirinae (Poaceae-Aveneae). *Pl Syst. Evol.* 136, 137–167. doi:10.1007/BF01004624

Alix, K., Baurens, F.-C., Paulet, F., Glaszmann, J.-C., and D'Hont, A. (1998). Isolation and Characterization of a Satellite DNA Family in the *Saccharum* Complex. *Genome* 41, 854–864. doi:10.1139/g98-076

Amosova, A. V., Bolsheva, N. L., Zoshchuk, S. A., Twardovska, M. O., Yurkevich, O. Y., Andreev, I. O., et al. (2017). Comparative Molecular Cytogenetic Characterization of Seven *Deschampsia* (Poaceae) Species. *PLoS One* 12, e0175760. doi:10.1371/journal.pone.0175760

Amosova, A. V., Ghukasyan, L., Yurkevich, O. Y., Bolsheva, N. L., Samatadze, T. E., Zoshchuk, S. A., et al. (2021). Cytogenomics of *Deschampsia* P. Beauv. (Poaceae) Species Based on Sequence Analyses and Fish Mapping of CON/COM Satellite Dna Families. *Plants* 10, 1105. doi:10.3390/plants10061105

Anamthawat-Jónsson, K., and Heslop-Harrison, J. S. (1993). Isolation and Characterization of Genome-specific DNA Sequences in Triticeae Species. *Mol. Gen. Genet.* 240, 151–158. doi:10.1007/BF00277052

Beckmann, J. S., and Soller, M. (1986). Restriction Fragment Length Polymorphisms and Genetic Improvement of Agricultural Species. *Euphytica* 35, 111–124. doi:10.1007/BF00028548

Bedbrook, J. R., Jones, J., O'Dell, M., Thompson, R. D., and Flavell, R. B. (1980). A Molecular Description of Telomeric Heterochromatin in Secale Species. *Cell* 19, 545–560. doi:10.1016/0092-8674(80)90529-2

Bennett, M. D., Smith, J. B., and Smith, R. I. L. (1982). DNA Amounts of Angiosperms from the Antarctic and South Georgia. *Environ. Exp. Bot.* 22, 307–318. doi:10.1016/0098-8472(82)90023-5

Cardone, S., Sawatani, P., Rush, P., García, A. M., Poggio, L., and Schrauf, G. (2009). Karyological Studies in Deschampsia antarctica Desv. (Poaceae). *Polar Biol.* 32, 427–433. doi:10.1007/s00300-008-0535-8

Charlesworth, B., Langley, C. H., and Stephan, W. (1986). The Evolution of Restricted Recombination and the Accumulation of Repeated DNA Sequences. *Genetics* 112, 947–962. doi:10.1093/genetics/112.4.947

Charlesworth, B., Sniegowski, P., and Stephan, W. (1994). The Evolutionary Dynamics of Repetitive DNA in Eukaryotes. *Nature* 371, 215–220. doi:10.1038/371215a0

Chiapella, J., and Zuloaga, F. O. (2010). A Revision of Deschampsia, Avenella, and Vahlodea (Poaceae, Poeae, Airinae) in South America1. *Ann. Mo. Bot. Garden* 97, 141–162. doi:10.3417/2008115

Chiapella, J. O., DeBoer, V. L., Amico, G. C., and Kuhl, J. C. (2011). A Morphological and Molecular Study in theDeschampsia Cespitosacomplex (Poaceae; Poeae; Airinae) in Northern North America. *Am. J. Bot.* 98, 1366–1380. doi:10.3732/ajb.1000495

Chiapella, J. O. (2007). A Molecular Phylogenetic Study of *Deschampsia* (Poaceae: Aveneae) Inferred from Nuclear ITS and Plastid trnL Sequence Data: Support for the Recognition of *Avenella* and *Vahlodea*. *Taxon* 55, 55–64. doi:10.2307/25065735

Dover, G. A. (1986). Molecular Drive in Multigene Families: How Biological Novelties Arise, Spread and Are Assimilated. *Trends Genet.* 2, 159–165. doi:10.1016/0168-9525(86)90211-8

Fasanella, M., Premoli, A. C., Urdampilleta, J. D., González, M. L., and Chiapella, J. O. (2017). How Did a Grass Reach Antarctica? the Patagonian Connection of *Deschampsia antarctica* (Poaceae). *Bot. J. Linn. Soc.* 185, 511–524. doi:10.1093/botlinnean/box070

Flavell, R. B. (1986). Repetitive DNA and Chromosome Evolution in Plants. *Phil. Trans. R. Soc. Lond. B* 312, 227–242. doi:10.1098/rstb.1986.0004

Fry, K., and Salser, W. (1977). Nucleotide Sequences of HS-α Satellite DNA from Kangaroo Rat dipodomys Ordii and Characterization of Similar Sequences in Other Rodents. *Cell* 12, 1069–1084. doi:10.1016/0092-8674(77)90170-2

Fuchs, J., Pich, U., Meister, A., and Schubert, I. (1994). Differentiation of Field Bean Heterochromatin Byin Situ Hybridization with a repeatedFokI Sequence. *Chromosome Res.* 2, 25–28. doi:10.1007/BF01539450

Garcia-Suarez, R., Alonso-Blanco, C., Fernandez-Carvajal, M. C., Fernandez-Prieto, J. A., Roca, A., and Giraldez, R. (1997). Diversity and Systematics ofDeschampsia Sensu Lato (Poaceae), Inferred from Karyotypes, Protein Electrophoresis, Total Genomic DNA Hybridization and Chloroplast DNA Analysis. *Pl Syst. Evol.* 205, 99–110. doi:10.1007/BF00982800

Garrido-Ramos, M. A. (2015). Satellite DNA in Plants: More Than Just Rubbish. *Cytogenet. Genome Res.* 146, 153–170. doi:10.1159/000437008

Garrido-Ramos, M. (2017). Satellite DNA: An Evolving Topic. *Genes* 8, 230. doi:10.3390/genes8090230

Gebre, Y. G., Bertolini, E., Pè, M. E., and Zuccolo, A. (2016). Identification and Characterization of Abundant Repetitive Sequences in *Eragrostis Tef* Cv. *Enatite* Genome. *BMC Plant Biol.* 16, 1–13. doi:10.1186/s12870-016-0725-4

González, M. L., Urdampilleta, J. D., Fasanella, M., Premoli, A. C., and Chiapella, J. O. (2016). Distribution of rDNA and Polyploidy in *Deschampsia antarctica* E. Desv. In Antarctic and Patagonic Populations. *Polar Biol.* 39, 1663–1677. doi:10.1007/s00300-016-1890-5

González, M. L., Chiapella, J. O., and Urdampilleta, J. D. (2018). Characterization of Some Satellite DNA Families in *Deschampsia antarctica* (Poaceae). *Polar Biol.* 41, 457–468. doi:10.1007/s00300-017-2205-1

González, M. L., Chiapella, J., Topalian, J., and Urdampilleta, J. D. (2020). Genomic Differentiation of *Deschampsia antarctica* and *D. Cespitosa* (Poaceae) Based on Satellite DNA. *Bot. J. Linn. Soc.* 194, 326–341. doi:10.1093/botlinnean/boaa045

González, M. L., Chiapella, J. O., and Urdampilleta, J. D. (2021). The Antarctic and South American Species of *Deschampsia*: Phylogenetic Relationships and Cytogenetic Differentiation. *Syst. Biodivers.* 19, 453–470. doi:10.1080/14772000.2020.1860151

Grebenstein, B., Grebenstein, O., Sauer, W., and Hemleben, V. (1995). Characterization of a Highly Repeated DNA Component of Perennial Oats (*Helictotrichon*, Poaceae) with Sequence Similarity to a A-genome-specific Satellite DNA of rice (*Oryza*). *Theoret. Appl. Genet.* 90, 1101–1105. doi:10.1007/BF00222928

Grebenstein, B., Grebenstein, O., Sauer, W., and Hemleben, V. (1996). Distribution and Complex Organization of Satellite DNA Sequences in Aveneae Species. *Genome* 39, 1045–1050. doi:10.1139/g96-131

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

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. (2000). Comparative Genome Organization in Plants: from Sequence and Markers to Chromatin and Chromosomes. *Plant Cell* 12, 617–636. doi:10.1105/tpc.12.5.61710.2307/3870990

Kawano, S. (1963). Cytogeography and Evolution of the Desghampsia Caespitosa Complex. *Can. J. Bot.* 41, 719–742. doi:10.1139/b63-059

King, K., Jobst, J. r., and Hemleben, V. (1995). Differential Homogenization and Amplification of Two Satellite DNAs in the Genus *Cucurbita* (Cucurbitaceae). *J. Mol. Evol.* 41, 996–1005. doi:10.1007/BF00173181

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, 96–105. doi:10.1159/000462915

Kubis, S., Schmidt, T., and Heslop-Harrison, J. S. (1998). Repetitive DNA Elements as a Major Component of Plant Genomes. *Ann. Bot.* 82, 45–55. doi:10.1006/anbo.1998.0779

Kuipers, A. G. J., Kamstra, S. A., De Jeu, M. J., and Visser, R. G. F. (2002). Molecular Characterization and Physical Localization of Highly Repetitive DNA Sequences from Brazilian *Alstroemeria* Species. *Chromosom. Res.* 10, 389–398. doi:10.1023/A:1016801702777

Lower, S. S., McGurk, M. P., Clark, A. G., and Barbash, D. A. (2018). Satellite DNA Evolution: Old Ideas, New Approaches. *Curr. Opin. Genet. Dev.* 49, 70–78. doi:10.1016/j.gde.2018.03.003

Macas, J., Mészáros, T., and Nouzová, M. (2002). PlantSat: A Specialized Database for Plant Satellite Repeats. *Bioinformatics* 18, 28–35. doi:10.1093/bioinformatics/18.1.28

Mallet, J. (2007). Hybrid Speciation. *Nature* 446, 279–283. doi:10.1038/nature05706

Mantovani, B., Tinti, F., Bachmann, L., and Scali, V. (1997). The Bag320 Satellite DNA Family in Bacillus Stick Insects (Phasmatodea): Different Rates of Molecular Evolution of Highly Repetitive DNA in Bisexual and

Parthenogenic Taxa. *Mol. Biol. Evol.* 14, 1197–1205. doi:10.1093/oxfordjournals.molbev.a025729

Mehrotra, S., and Goyal, V. (2014). Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics, Proteomics Bioinf.* 12, 164–171. doi:10.1016/j.gpb.2014.07.003

Nakajima, R., Noma, K., Ohtsubo, H., and Ohtsubo, E. (1996). Identification and Characterization of Two Tandem Repeat Sequences (TrsB and TrsC) and a Retrotransposon (RIRE1) as Genome-General Sequences in rice. *Genes Genet. Syst.* 71, 373–382. doi:10.1266/ggs.71.373

Nani, T. F., Pereira, D. L., Souza Sobrinho, F., and Techio, V. H. (2016). Physical Map of Repetitive DNA Sites in *Brachiaria* spp.: Intravarietal and Interspecific Polymorphisms. *Crop Sci.* 56, 1769–1783. doi:10.2135/cropsci2015.12.0760

Navrotska, D., Andreev, I., Betekhtin, A., Rojek, M., Parnikoza, I., Myryuta, G., et al. (2018). Assessment of the Molecular Cytogenetic, Morphometric and Biochemical Parameters of *Deschampsia antarctica* from its Southern Range Limit in Maritime Antarctic. *Polish Polar Res.* 39, 525–548. doi:10.24425/118759

Noor, M. A. F., Grams, K. L., Bertucci, L. A., and Reiland, J. (2001). Chromosomal Inversions and the Reproductive Isolation of Species. *Proc. Natl. Acad. Sci.* 98, 12084–12088. doi:10.1073/pnas.221274498

Pagel, M. (1999). Inferring the Historical Patterns of Biological Evolution. *Nature*.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* 20, 289–290. doi:10.1093/bioinformatics/btg412

Parodi, L. R. (1949). Las gramíneas sudamericanas del género *Deschampsia*. *Darwiniana* 8, 415–475.

Plohl, M., Meštrovic, N., and Mravinac, B. (2012). "Satellite DNA Evolution," in *Repetitive DNA* (Granada: Karger Publishers), 126–152. doi:10.1159/000337122

Plohl, M. (2010). Those Mysterious Sequences of Satellite DNAs. *Period. Biol.* 112, 403–410.

Raskina, O., Barber, J. C., Nevo, E., and Belyayev, A. (2008). Repetitive DNA and Chromosomal Rearrangements: Speciation-Related Events in Plant Genomes. *Cytogenet. Genome Res.* 120, 351–357. doi:10.1159/000121084

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

Röser, M., Winterfeld, G., Döring, E., and Schneider, J. (2014). Chromosome Evolution in Grass Tribes Aveneae/Poeae (Poaceae): Insights from Karyotype Structure and Molecular Phylogeny. *Schlechtendalia* 28, 1–22.

Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., and Camacho, J. P. M. (2016). High-throughput Analysis of the Satellitome Illuminates Satellite DNA Evolution. *Sci. Rep.* 6, 1–14. doi:10.1038/srep28333

Saarela, J. M., Bull, R. D., Paradis, M. J., Ebata, S. N., Peterson, P. M., Soreng, R. J., et al. (2017). Molecular Phylogenetics of Cool-Season Grasses in the Subtribes Agrostidinae, Anthoxanthinae, Aveninae, Brizinae, Calothecinae, Koeleriinae and Phalaridinae (Poaceae, Pooideae, Poeae, Poeae Chloroplast Group 1). *Pk* 87, 1–139. doi:10.3897/phytokeys.87.12774

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, 2213–2227. doi:10.1007/s00122-018-3148-9

Schwarzacher, T., and Heslop-Harrison, P. (2000). *Practical* in Situ *Hybridization*. Norwich: BIOS Scientific Publishers Ltd.

Sharma, S., and Raina, S. N. (2005). Organization and Evolution of Highly Repeated Satellite DNA Sequences in Plant Chromosomes. *Cytogenet. Genome Res.* 109, 15–26. doi:10.1159/000082377

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

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., and Karpen, G. H. (2017). Phase Separation Drives Heterochromatin Domain Formation. *Nature* 547, 241–245. doi:10.1038/nature22989

Tzvelev, N. N., Probatova, N. S., Probatova, N. S., and Chiapella, J. (2015). New Taxa of *Deschampsia* P. Beauv. (Poaceae) from Russia. *Bot. Pac.* 4, 1–6. doi:10.17581/bp.2015.04105

Ugarkovic, D. (2008). Satellite DNA Libraries and Centromere Evolution. *Toevolj* 2, 1–6. doi:10.2174/1874404400802010001

Vershinin, A. V., Schwarzacher, T., and Heslop-Harrison, J. S. (1995). The Large-Scale Genomic Organization of Repetitive DNA Families at the Telomeres of rye Chromosomes. *Plant Cell* 7, 1823–1833. doi:10.1105/tpc.7.11.1823

Wei, L., Liu, B., Zhang, C., Yu, Y., Yang, X., Dou, Q., et al. (2020). Identification and Characterization of Satellite DNAs in *Poa* L. *Mol. Cytogenet.* 13, 1–8. doi:10.1186/s13039-020-00518-x

Widmer, A., Lexer, C., and Cozzolino, S. (2009). Evolution of Reproductive Isolation in Plants. *Heredity* 102, 31–38. doi:10.1038/hdy.2008.69

Winterfeld, G., and Röser, M. (2007a). Chromosomal Localization and Evolution of Satellite DNAs and Heterochromatin in Grasses (Poaceae), Especially Tribe Aveneae. *Plant Syst. Evol.* 264, 75–100. doi:10.1038/ejhg.2015.20410.1007/s00606-006-0482-1

Winterfeld, G., and Röser, M. (2007b). Disposition of Ribosomal DNAs in the Chromosomes of Perennial Oats (Poaceae: Aveneae). *Bot. J. Linn. Soc.* 155, 193–210. doi:10.1111/j.1095-8339.2007.00690.x

Yang, J., Yuan, B., Wu, Y., Li, M., Li, J., Xu, D., et al. (2020). The Wide Distribution and Horizontal Transfers of Beta Satellite DNA in Eukaryotes. *Genomics* 112, 5295–5304. doi:10.1016/j.ygeno.2020.10.006

Zhao, Y., Yu, F., Liu, R., and Dou, Q. (2017). Isolation and Characterization of Chromosomal Markers in *Poa Pratensis*. *Mol. Cytogenet.* 10, 5. doi:10.1186/s13039-017-0307-7

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 September 2021
doi: 10.3389/fgene.2021.727314

High-Throughput Genomic Data Reveal Complex Phylogenetic Relationships in *Stylosanthes* Sw (Leguminosae)

Maria Alice Silva Oliveira[1‡], *Tomáz Nunes*[1‡], *Maria Aparecida Dos Santos*[1‡], *Danyelle Ferreira Gomes*[1], *Iara Costa*[1], *Brena Van-Lume*[2], *Sarah S. Marques Da Silva*[1], *Ronaldo Simão Oliveira*[3], *Marcelo F. Simon*[4], *Gaus S. A. Lima*[5], *Danilo Soares Gissi*[6], *Cícero Carlos de Souza Almeida*[1]*, *Gustavo Souza*[2]* and *André Marques*[1,7*†]

[1]Laboratory of Genetic Resources, Federal University of Alagoas, Arapiraca, Brazil, [2]Laboratory of Plant Cytogenetics and Evolution, Federal University of Pernambuco, Recife, Brazil, [3]Campus Xique Xique, Federal Institute of Education, Science and Technology of Bahia, Xique-Xique, Brazil, [4]Embrapa Cenargen, Brasília, Brazil, [5]Center of Agronomic Sciences, Federal University of Alagoas, Rio Largo, Brazil, [6]Department of Biostatistics, Institute of Biosciences–IBB, Plant Biology, Parasitology and Zoology, São Paulo State University–UNESP, Botucatu, Brazil, [7]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany

Allopolyploidy is widely present across plant lineages. Though estimating the correct phylogenetic relationships and origin of allopolyploids may sometimes become a hard task. In the genus *Stylosanthes* Sw. (Leguminosae), an important legume crop, allopolyploidy is a key speciation force. This makes difficult adequate species recognition and breeding efforts on the genus. Based on comparative analysis of nine high-throughput sequencing (HTS) samples, including three allopolyploids (*S. capitata* Vogel cv. "Campo Grande," *S. capitata* "RS024" and *S. scabra* Vogel) and six diploids (*S. hamata* Taub, *S. viscosa* (L.) Sw., *S. macrocephala* M. B. Ferreira and Sousa Costa, *S. guianensis* (Aubl.) Sw., *S. pilosa* M. B. Ferreira and Sousa Costa and *S. seabrana* B. L. Maass & 't Mannetje) we provide a working pipeline to identify organelle and nuclear genome signatures that allowed us to trace the origin and parental genome recognition of allopolyploids. First, organelle genomes were *de novo* assembled and used to identify maternal genome donors by alignment-based phylogenies and synteny analysis. Second, nuclear-derived reads were subjected to repetitive DNA identification with RepeatExplorer2. Identified repeats were compared based on abundance and presence on diploids in relation to allopolyploids by comparative repeat analysis. Third, reads were extracted and grouped based on the following groups: chloroplast, mitochondrial, satellite DNA, ribosomal DNA, repeat clustered- and total genomic reads. These sets of reads were then subjected to alignment and assembly free phylogenetic analyses and were compared to classical alignment-based phylogenetic methods. Comparative analysis of shared and unique satellite repeats also allowed the tracing of allopolyploid origin in *Stylosanthes*, especially those with high abundance such as the StyloSat1 in the Scabra complex. This satellite was *in situ* mapped in the proximal region of the chromosomes and made it possible to identify its previously proposed parents. Hence, with simple genome skimming data we were able to provide evidence for

OPEN ACCESS

Edited by:
Francisco J. Ruiz-Ruano, University of East Anglia, United Kingdom

Reviewed by:
Sergio Sebastián Samoluk, Instituto de Botánica del Nordeste, Argentina
Makenzie Mabry, Florida Museum of Natural History, United States

***Correspondence:**
Cícero Carlos de Souza Almeida
cicerocarlos@hotmail.com
Gustavo Souza
luiz.rodriguessouza@ufpe.br
André Marques
amarques@mpipz.mpg.de

†ORCID:
André Marques
orcid.org/0000-0002-9567-2576

‡These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *18 June 2021*
Accepted: *08 September 2021*
Published: *23 September 2021*

Citation:
Oliveira MAS, Nunes T, Dos Santos MA, Ferreira Gomes D, Costa I, Van-Lume B, Marques Da Silva SS, Oliveira RS, Simon MF, Lima GSA, Gissi DS, Almeida CCdS, Souza G and Marques A (2021) High-Throughput Genomic Data Reveal Complex Phylogenetic Relationships in Stylosanthes *Sw (Leguminosae). Front. Genet. 12:727314. doi: 10.3389/fgene.2021.727314*

the recognition of parental genomes and understand genome evolution of two *Stylosanthes* allopolyploids.

Keywords: sytlosanthes, allopolyploidy, repetitive DNA, organelle genome, chloroplast, mitochondrion, alignment and assembly free

INTRODUCTION

High-throughput sequencing (HTS) technologies have recently emerged as a versatile source of sequencing data allowing researchers to rapidly access different aspects of biodiversity based on four main approaches: genome skimming, RAD-Seq, RNA-Seq, and Hyb-Seq (Dodsworth et al., 2019). Of these, the skimming genome stands out for being the sequencing (usually in low coverage) of small random genome fragments (reads) through Next Generation Sequencing (NGS) technologies. The genome skimming analysis allowed the development of several new bioinformatics tools for genomic analysis of non-model organisms, for instance, RepeatExplorer (Novak et al., 2010; Novak et al., 2013). This pipeline has been used to characterize repetitive fractions of genomes, discover new repetitive elements, perform genomic comparative studies, develop probes for fluorescent *in situ* hybridization (Marques et al., 2015; McCann et al., 2020) or characterize distinct subgenomes in allopolyploids (Han et al., 2005; Hemleben et al., 2007).

Recent studies have shown the potential of genome skimming data for phylogenomic studies. Dodsworth et al. (2015) have demonstrated the potential to build phylogenetic topologies based on repeats abundance. This approach has been improved, incorporating the similarities of repeats to the construction of phylogenetic trees (Vitales et al., 2020). Recently developed tools which perform phylogenetic inferences from entire HTS data without the need of alignment or assembly, i.e., alignment and assembly free approaches (Fan et al., 2015; Sarmashghi et al., 2019), just by counting shared and unique k-mers, may allow the use of repeat-derived reads in phylogenetic inferences. On the other hand, genome skimming data also allows the assembly of complete organelle genomes (plastomes and mitogenomes), as well as large tandem repeats as the ribosomal DNA (rDNA) units (Dodsworth et al., 2019). The use of massive alignments of whole chloroplast genomes is frequently the method of choice for establishing phylogenetic relationships in plants, based on usual phylogenetic approaches as Bayesian Inference, Maximum Likelihood, etc. (Guo et al., 2017; Kim et al., 2019; Xie et al., 2020). Organelle inheritance mostly maternal for most plant species (Reboud and Zeyl, 1994; Greiner et al., 2014), makes the sequence of organelle genomes ideal for identifying patterns of maternal genome inheritance in hybrid species (Gastony and Yatskievych, 1992; Jankowiak et al., 2005). Thus, these different genomic and phylogenomic approaches could be important to characterize the origin and evolution of allopolyploid complexes.

The genus *Stylosanthes* Sw. (Leguminosae) belongs to the subfamily Papilionoideae and has a complex systematics, mainly related to the occurrence of natural allopolyploidy (Stace and Edye, 1984; Vanni, 2017). The taxonomy of the genus remains unsettled and controversial, with various authors favoring between 25 and 42 species, with at least 40 additional synonyms (Cameron and Chakraborty, 2004). At least 16 taxa are thought to have been originated by allopolyploidy, which seems to be directly related to the unresolved taxonomy of the genus (Liu and Musial, 2001). *Stylosanthes* shows close evolutionary proximity with the peanut genus *Arachis* L., forming sister lineages in the clade *Pterocarpus* (tribe Dalbergieae) (Cardoso et al., 2013). The genus is the most economically important forage legumes, with species grown worldwide as a pasture crop with grasses, as well as for land reclamation and restoration, soil stabilization, and regeneration, particularly in regions with low precipitation (Cameron and Chakraborty, 2004).

Stylosanthes is highly diversified and morphologically polymorphic, having cultivated pantropical species, mostly described for the American continent with two centers of diversification: Mexico and Brazil. Being the latter the main center of origin and diversification for the genus, with more than 30 species, of which 12 are endemic (Stace and Edye, 1984; da Costa and Valls, 2010; Santos-Garcia et al., 2012; Vanni, 2017). Species circumscription and identification are complex in *Stylosanthes* since many different species have overlapping morphological characters, many of them dubious e/or homoplastic (Costa and Ferreira, 1984; Mannetje, 1984; Vanni, 2017). This makes it necessary to use additional data to taxonomy, such as molecular markers (Liu et al., 1999; Liu and Musial, 2001), molecular phylogeny (Vander Stappen et al., 1999; Vander Stappen et al., 2002), genomics or cytogenetics (Marques et al., 2018; Franco et al., 2020).

The basic chromosome number of *Stylosanthes* is $x = 10$, with occurrence of diploids ($2n = 20$), tetraploids ($2n = 40$) and hexaploids ($2n = 60$) species (Stace and Edye, 1984; Vieira et al., 1993). Studies based on restriction fragment length polymorphisms (RFLP) and sequence-tagged site (STS) analyses identified ten genome compositions in the genus, named A to J (Liu et al., 1999; Ma et al., 2004). However, the origin and evolution of most *Stylosanthes* allopolyploid complexes remain largely unresolved (Maass and Sawkins, 2004). One of the few allopolyploids well characterized from a genomic point of view is *Stylosanthes scabra* Vogel (AABB), a hybrid between species of A and B genomes, i.e. *S. hamata* (L.) Taub. or *S. seabrana* B.L.Maass & 't Mannetje (A genomes) and *S. viscosa* (L.) Sw. (B genome) (Marques et al., 2018). This origin was further demonstrated by the whole chloroplast genome versus rDNA phylogeny and genomic *in situ* hybridization (GISH) (Marques et al., 2018). However, the origin of other allopolyploids of agronomic interest such as *S. capitata* Vogel (supposed to be a hybrid between species with D and E genomes) is still unknown (Liu et al., 1999; Liu and Musial, 2001; Vander Stappen et al., 2002).

TABLE 1 | List of accessions, ploidy level, reference number, SRA, plastome and mitogenome accession numbers and genome composition.

Specie`s name	Ploidy level/ Chromosome number	Code/ Accession number	Genbank accession	Plastome accession	Mitogenome accession	Genome composition[a]
S. hamata	Diploid (2n = 20)	SHA 2701 LC 7666	SRX3517479	NC_039159	MZ747306	A
S. viscosa	Diploid (2n = 20)	SVI 2702 A-01	SRX3517481	NC_039161	MZ747307	B
S. scabra	Tetraploid (2n = 40)	SSC 2703 CPAC-5234	SRX3517480	NC_039160	MZ747308	AB
S. capitata	Tetraploid (2n = 40)	SCA 2705 cv. Campo Grande	SRX5395139	MZ747315	MZ747309	AB
S. pilosa	Diploid (2n = 20)	SPI 2706 LC 7833	SRX5395138	MZ747316	MZ747310	E
S. macrocephala	Diploid (2n = 20)	SMA 2707 cv. Campo Grande	SRX5395140	MZ747317	MZ747311	D
S. capitata	Tetraploid (2n = 40)	SCA 2708 RS024	SRR13855961	MZ747318	MZ747312	DE
S. seabrana	Diploid (2n = 20)	SSE 2709 LC 6261	SRR13855960	MZ747319	MZ747313	A
S. guianensis	Diploid (2n = 20)	SGU 2710 EPAMIG 906	SRR13855959	MZ747320	MZ747314	G
Arachis hypogaea	Tetraploid (2n = 40)		DRR056349	NC_037358	-	

[a]*Genome compositions following the classification of Liu and Musial. (2001).*

In order to clarify the origin of *Stylosanthes* allopolyploids, we tested the efficiency of different bioinformatic analysis using HTS data from three allopolyploid accessions and six diploids, including five different genome compositions. We focused on the characterization of two allopolyploid complexes: *S. scabra* (*S. hamata*/*S. seabrana* + *S. viscosa*) and *S. capitata* (*S. macrocephala* M. B. Ferreira and Sousa Costa + *S. pilosa* M. B. Ferreira and Sousa Costa). For this, we performed comparative genomic analysis, anchored by phylogenomic inferences based on whole organelle (plastome and mitogenome), rDNA, satellite DNAs, and total reads. The whole plastome and mitogenome of all these were assembled and characterized comparatively. *In situ* hybridization based on species-specific satDNA repeats has further confirmed the origin of *S. scabra* and opens a field for further cytogenetic research on *Stylosanthes* allopolyploids. Finally, based on different phylogenetic approaches we discuss the phylogenetic complexity of the genus and the utility of HTS data to help the characterization of *Stylosanthes* allopolyploids.

MATERIAL AND METHODS

Plant Material

Samples of nine *Stylosanthes* species analyzed here, including two complexes (*S. scabra* and *S. capitata*) are listed in **Table 1**. For comparative studies, available data from *Arachis hypogaea* L. and our previous data from *S. scabra*, *S. hamata*, and *S. viscosa* (Marques et al., 2018) were used. Plant tissue (young leaves, fresh 5–20 g each) of all nine *Stylosanthes* accessions were collected from plants growing in the greenhouse of the Laboratory of Genetic Resources at Federal University of Alagoas.

High Throughput Sequencing

The genomic paired-end short reads for *Stylosanthes hamata*, *S. viscosa*, and *S. scabra* samples, which belong to the *S. scabra* complex, were the same obtained by Marques et al. (2018), and downloaded from the available accession numbers on NCBI (**Table 1**). Similarly, as outgroup in our analyses, we have downloaded genomic paired-end short reads for *Arachis hypogaea*.

For *S. pilosa* (LC 7833), *S. macrocephala* (cv. Campo Grande), *S. capitata* (cv. Campo Grande), *S. capitata* (RS024), *S. seabrana* (LC 6261), and *S. guianensis* (Aubl.) Sw. (EPAMIG 906), we have collected young fresh leaves (±1–5 g) for DNA isolation with the kit NUCLEOSPIN PLANT II (Macherey-Nagel). The isolated DNA was checked in 1% (*p*/v) agarose gel and the concentration measured with a NanoDrop 2000 (Thermo Scientific).

The HTS was done with GenOne Soluções em Biotecnologia, Rio de Janeiro, Brasil, where 5 μg of gDNA were used for each sample for the library preparation. The sequencing library was

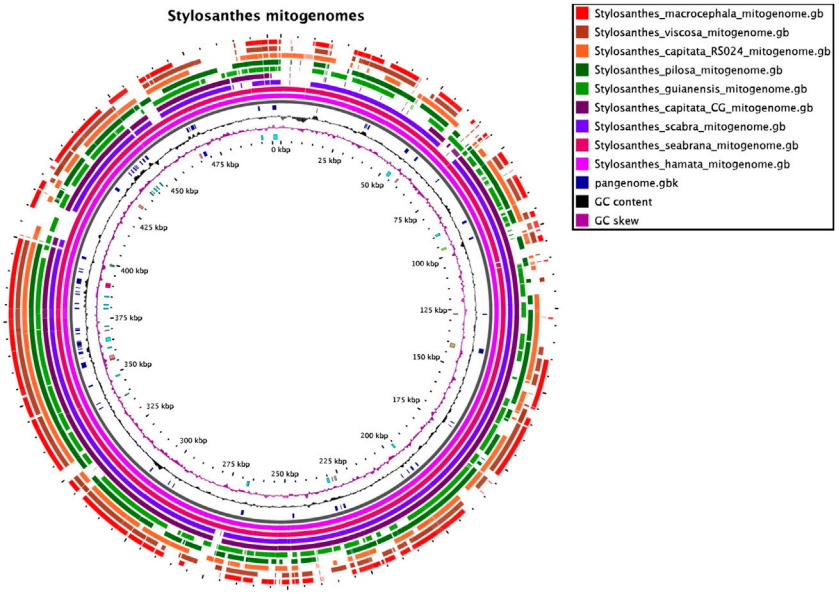

FIGURE 1 | The mitochondrial pangenome of *Stylosanthes*. Annotated sequences and features are shown in the inner circle. For details see **Supplementary Figure S1**, **Supplementary Tables S1, S2**.

generated using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, United Kingdom) following the manufactures recommendations. The genomic DNA was randomly sheared to a final fragmented size of 350 bp by Bioruptor and further selected and ligated with adapters. The library was analyzed for the size distribution of fragments by Agilent2100 Bioanalyzer and quantified by real-time PCR. The library was then paired-end (2 × 150 bp) sequenced with Illumina NovaSeq 6000 sequencer generating 3 Gb per sample.

De Novo Organelle Genome and rDNA Assembly

Plastomes of *S. hamata* (SHA 2701), *S. viscosa* (SVI 2702), and *S. scabra* (SSC 2703) were already assembled before (Marques et al., 2018). For plastome assembly the total number of unprocessed paired-end reads obtained for *S. capitata* (SCA 2705), *S. pilosa* (SPI 2706), *S. macrocephala* (SMA 2707), *S. capitata* (SCA 2708), *S. seabrana* (SSE 2709), and *S. guianensis* (SGU 2710) were used (**Table 1**). *De novo* plastome assemblies of reads were performed by NOVOPlasty v3.8.3 (Dierckxsens et al., 2017) using default parameters. As NOVOPlasty does not need quality trimming of the reads, all reads for each species were used. NOVOPlasty was able to assemble a single circularized contig for each species, representing the whole plastome including all regions: Long Single Copy (LSC), Short Single Copy (SSC), and both Inverted Repeats (IRs) regions.

NOVOPlasty v3.8.3 was also used to assemble the mitogenomes from the nine *Stylosanthes* samples. From the nine samples, NOVOPlasty was able to retrieve circular mitogenomes for six species: *S. hamata* (SHA 2701), *S. viscosa* (SVI 2702), *S. scabra* (SSC 2703), *S. capitata* (SCA 2705), *S. seabrana* (SSE 2709) and *S. guianensis* (SGU 2710). For *S. capitata* (SCA 2708), *S. macrocephala* (SSE 2709), and *S. pilosa* (SPI 2706) a single linear contig was obtained. All plastomes and mitogenomes contigs obtained were imported into Geneious v. 9.1.8 and the assembly was checked by mapping the raw reads to the contigs using the Geneious mapper with low sensitivity. Plastomes and mitogenomes were annotated using the Geneious annotation tool, guided by the available Leguminosae plastomes and mitogenomes on NCBI. Annotations were manually checked to correct misannotated regions. Plastome and mitogenome maps were generated using OrganellarGenomeDraw (OGDraw v1.2) (Lohse et al., 2013). Repeats (>95% similarity and >500 bp) in each mitochondrial genome were identified with the "Find Repeats" tool available on Geneious v. 9.1.8. All organelle genome accession numbers are provided in **Table 1**.

To obtain the complete sequence of 5S and 18S-5.8S-28S (35S) rDNA units including the NTS and ITS spacers, respectively, rDNA contigs from the output of RepeatExplorer2 were identified and used as a reference to map the reads from the entire HTS dataset from each sample. Consensus sequences for both 5S and 35S rDNA units were annotated based on comparison with other rRNA genes in Genbank and used for alignment-based analysis. Alternatively, the reads mapped to each unit were used for the alignment and assembly free approach (see below).

Synteny Comparison of Mitochondrial Genomes

The nine mitochondrial genomes developed here were compared with each other. The software SyMAP (Soderlund et al., 2011) was

used to find syntenic regions in a pairwise-based comparison shown on **Figure 1**. For this syntenic blocks were calculated based on the annotation and order of genic regions.

Characterization of DNA Repeats

One million quality filtered paired-end reads of each sample, including *A. hypogaea*, were uploaded to the RepeatExplorer2 Galaxy web server (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/) for *de novo* repeat identification and characterization. All samples were subjected to individual and comparative analysis, including *A. hypogaea*, using the RepeatExplorer2 (RE2) version with the long run parameters. In comparative mode reads of each species were sampled according to their ploidy levels and genome sizes based on the reference diploid *S. hamata* and the polyploid *S. scabra* genome sizes (Marques et al., 2018). A custom repeat database was built from the first RE2 run including characterized satDNA repeats from *Stylosanthes* and the repeat library from *A. hypogaea* genome (available for download at https://peanutbase.org/) and used for a second run to facilitate repeat annotation and comparison among the samples. Additionally, all samples were also analyzed with TAREAN (Novak et al., 2017), a tandem repeat identification tool available in RE2, which allows quick characterization of sequence composition and diversity of satDNA repeats. Finally, the main repeat clusters were classified into the main repeat families and compared by abundance among the samples.

Graph-Based Clustering of Satellitome, Mitogenome, and rDNA Reads and Interactive Visualization

Consensus sequences from satDNA repeats identified by TAREAN/RE2, rDNA contigs, and assembled mitogenomes were used as a reference for mapping the entire HTS dataset from each sample. All reads belonging to these three classes of sequences were separately retrieved using Geneious mapper tool with medium-low sensitivity. SatDNA, rDNA, and mitochondrial reads retrieved were separately used as input for comparative graph-based clustering using RepeatExplorer2. Interactive visualization of cluster graphs was performed with the R package SeqGrapheR, which provides a simple graphical user interface for interactive visualization of sequence clusters. SeqGrapheR enabled the selection of species-specific reads from cluster graphs allowing simultaneous viewing of the graph layout (Novak et al., 2010).

Alignment-Based Phylogenetic Sequence Comparison and Dating

Phylogenetic analyses were initially performed on a subset of diploid samples to avoid the possible uncertain relationships of polyploid specimens in a bifurcating tree. Then, the analysis including the polyploids was conducted using network approaches to account for possible inconsistencies (see below). Alignment of complete plastomes and nuclear 35S rDNA regions were performed with MAFFT v7.222 (Katoh and Standley, 2013) as a Geneious v. 9.1.8 plugin (Kearse et al., 2012). Phylogenetic relationships were inferred using Maximum likelihood with IQ-TREE2 (Nguyen et al., 2015; Minh et al., 2020) and Bayesian Inference (BI) approach implemented in BEAST v.1.8.3 (Drummond et al., 2012). A Markov Chain Monte Carlo (MCMC) run was conducted, sampling every 1,000 generations for 10,000,000 generations using the model GTR. The run was evaluated in Tracer v.1.6 (Rambaut et al., 2014) to determine that the estimated sample size (ESS) for each relevant parameter was >200, and a burn-in of 25% was applied. The majority-rule consensus tree and posterior probability (PP) were visualized and edited in FigTree v.1.4.2 (Rambaut, 2014). Splitstree4 (Huson and Bryant, 2006) was used to generate relationship networks for datasets containing diploids and polyploids, based on the standard function of maximum parsimony. As the outgroup for 1) plastome phylogenetic comparisons, we used the available plastome of *Arachis hypogaea* (KJ468094); 2) rDNA comparisons, we used available ITS1-5.8S-ITS2 regions on NCBI for *Stylosanthes* and *Arachis* species, and 3) 35S rDNA comparisons, the SRA file accession no. DRR056349 from *A. hypogaea*, where the assembly of 35S rDNA was performed as described above for *Stylosanthes*.

Divergence time estimates were performed in BEAST v.1.10.4 (Drummond and Rambaut, 2007) fixing the tree topology of the Bayesian analyses. An uncorrelated relaxed lognormal clock (Drummond et al., 2006) and a Yule Process speciation model (Gernhard, 2008) were applied. Two independent runs of 10,000,000 generations each were performed, sampling every 10,000 generations for the full plastome alignment. In order to verify the effective sampling of all parameters and assess the convergence of independent chains, we examined their posterior distributions in Tracer v.1.6, and the MCMC sampling was considered sufficient at an ESS >200. 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. (Drummond et al., 2012). Calibrations were performed using the secondary calibrations of Särkinen et al. (2012) for the *Arachis*/*Stylosanthes* divergence approx. 12.4 million years ago (Mya).

Repeat-Based Alignment and Assembly Free Phylogenetic Analysis

To access the phylogenetic signal of diverse repeat class and to avoid loose information we decided to use a recently developed approach that is able to infer phylogenetic trees out of HTS data without the need for alignment using the alignment and assembly free (AAF) tool (Fan et al., 2015). AAF constructs phylogenies directly from unassembled genome sequence data, bypassing both genome assembly and alignment. Using mathematical calculations, models of sequence evolution, and simulated sequencing of published genomes, AAF addresses both evolutionary and sampling issues caused by direct reconstruction, including homoplasy, sequencing errors, and incomplete sequencing coverage. Thus, it calculates the statistical properties of the pairwise distances between genomes, allowing it to optimize parameter selection and perform bootstrapping. Since this approach only needs a set of

reads per sample it makes the analysis quite flexible, where we can examine the phylogenetic signal of different sets of sequencing data. Thus, we have made phylogenetic inferences for different sets of data, 1- satellitome reads, which comprises all reads mapped to the consensus satellite DNA of each sample, 2- repeat reads, comprising all reads that were clustered with RE2, 3- all reads, comprising a random subsample from each sample, generated with the *reformat.sh* tool (BBMap – Bushnell B. – sourceforge.net/projects/bbmap/)

To check whether satDNA repeats found in the different species are also present in the other species in lower abundance and to identify the different families, we have compiled all consensus sequences from the TAREAN output, which consisted of 54 consensus sequences in total. We have used this file as a reference to uniquely map the total amount of HTS reads from each sample. Mapped reads were grouped by each consensus that they mapped. First, we collected all these reads by species, that were assumed to be a sum of all satellite reads from each sample (satellitome reads). These reads were concatenated in FASTA files and subsequently analyzed in a comparative analysis to test whether identified satDNA repeats are shared among the species and how they group in different clusters. Genomic abundances were then inferred by the number of reads mapped to each satDNA repeat. We have considered only satDNA repeats showing at least 0.01% of genomic abundance in at least one of the samples. For our comparative analysis, we also considered the genomic abundance obtained from our mapping strategy instead of the RepeatExplorer estimations.

Slide Preparation

For cytogenetic analysis, seeds were germinated and root tips were collected and pretreated with 8-hydroxyquinoline for 20 h at 10°C, fixed in ethanol:acetic acid (3:1; v/v) from 2 to 24 h at room temperature and stored at –20°C. The fixed roots were washed in distilled water and digested in 2% cellulase (Onozuka) and 20% pectinase (Merck) at 37°C for 90 min. Then apical meristems were squashed in 45% acetic acid under a coverslip. The coverslip was removed in liquid nitrogen.

Probe Labeling and *in situ* Hybridization

In order to localize the satDNA repeats identified in the *S. scabra* complex, the repeats were amplified by PCR and labeled with Cy3-dUTP (GE Healthcare) or with digoxigenin-11-dUTP (Merck). All probes were labeled by nick translation (Merck). These labeled probes were used for FISH. Digoxigenin-labeled probes were detected with sheep anti-digoxigenin FITC conjugate (Merck) and amplified with rabbit anti-sheep FITC conjugate (Bio-Rad). FISH was performed according to (Marques et al., 2018). The hybridization mix contained 50% of formamide (v/v), 10% dextran sulfate (w/v), 2 × SSC, and 50 ng of each probe. The final hybridization stringency was estimated to be 76%. The slides were mounted with 4′,6-diamidino-2-phenylindole (DAPI, 4 μg/ml)/Vectashield (Vector) 1:1 (v/v) and analyzed under a Leica epifluorescence microscope and the Leica Las AF software. Overlapping, processing of images for brightness and contrast were performed using Adobe Photoshop® CC 2019.

RESULTS

Mitochondrial Genomes

A circularized mitochondrial genome assembly was obtained for all nine samples and varied in length from 350,377 bp in *S. macrocephala* to 523,870 bp in *S. seabrana* (**Supplementary Table S1**). An extensive variation in the order of approximately 200 Kb was found among *Stylosanthes* mitogenomes with the following increasing order: *S. macrocephala* SMA 2707 (350,377 bp), *S. viscosa* SVI 2702 (353,136 bp), *S. capitata* SCA 2708 (384,410 bp), *S. pilosa* SPI 2706 (433,649 bp), *S. capitata* SCA 2705 (456,448 bp), *S. guianensis* SGU 2710 (468,896 bp), *S. scabra* SSC 2703 (492,899 bp), *S. hamata* SHA 2701 (503,967 bp) and *S. seabrana* SSE 2709 (523,870 bp) (**Supplementary Table S1**). Mitogenomes features of each studied species, including the number of transfer RNA (tRNA), ribosomal RNA (rRNA), and protein-coding genes from the annotated regions are shown in **Figure 2** and **Supplementary Table S2**. Mitogenome maps for each sample are provided on **Supplementary Figure S1**.

To access mitochondrial genome structure variation in *Stylosanthes*, the syntenic relationship was analyzed within all nine *Stylosanthes* mitochondrial genomes using SyMAP (**Figure 1**). In general, low conservation of synthetic blocks (considered from the gene order) was observed in the analyzed *Stylosanthes* species (**Figure 1A**). We investigated comparatively the relationship of the mitogenomes of each allotetraploid with the other species of the genus (**Figures 1B–D**). Relatively high linearity was observed among the mitogenomes of the allotetraploids *S. scabra* SSC 2703 and *S. capitata* SCA 2705, and between them and the phylogenetically close diploids *S. hamata* SHA 2701 and *S. seabrana* SSE 2709 (**Figures 1B,C**). Surprisingly, no evidence of linearity between the mitogenome of the two *S. capitata* samples (SCA 2705 and SCA 2708) and its putative diploid progenitors *S. pilosa* SPI 2706 (E genome) or *S. macrocephala* SMA 2707 (D genome) was observed (**Figures 1D**).

Chloroplast Genomes

Plastomes of all species are very similar in length varying from 156,244 bp in *S. viscosa* to 156,502 bp in *S. hamata* and *S. scabra* (**Figure 3**; **Supplementary Figure S2**). No major macrostructural rearrangements were detected in the *Stylosanthes* genomes analyzed here. The potential of these plastomes for determining the (maternal) origin of allopolyploids has been explored from a phylogenetic point of view (see below).

Genomic Repetitive Fraction Characterization and Comparative Analysis of Repeat Abundance

Individual clustering analysis with RepeatExplorer2 revealed that all nine *Stylosanthes* species shared a similar amount of repetitive sequences in their genomes (over 50%). In the present study, we have identified different families of repetitive elements, belonging to Class I (retroelements) and II (DNA transposons) mobile

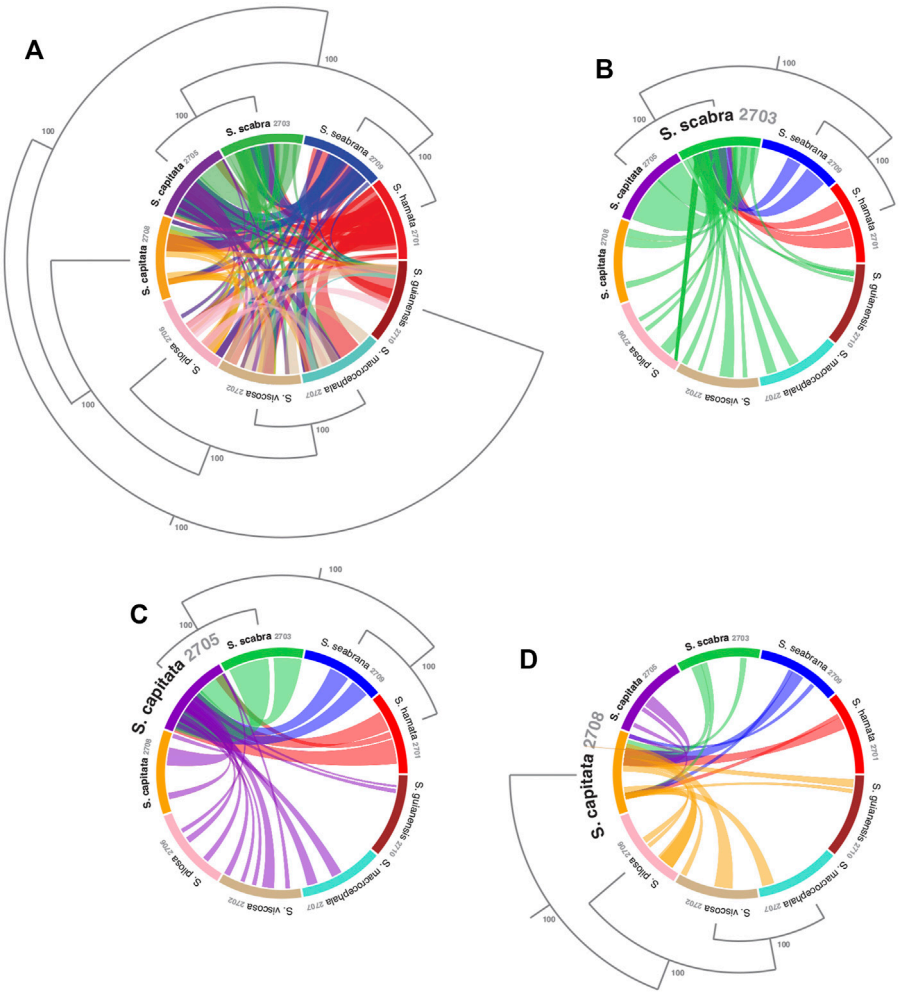

FIGURE 2 | Visualizing sequence similarity among mitogenomes of *Stylosanthes* species with a focus on the relationships among the allotetraploids. **(A)** Mitogenomes synteny all-to-all. Synteny of *S. scabra* mitogenome **(B)**, *S. capitata* (SCA 2705) **(C)** and *S. capitata* (SCA 2708) **(D)** to the other species. Phylogenetic relationships were based on the topology of whole plastome alignment analyzed by Bayesian inference. Values in the nodes indicate posterior probabilities.

elements, as well as 5S and 35S rDNA and satDNA repeats (**Table 2**). The sample that showed the most repeat diversity was *S. macrocephala* (SMA 2707), with a total of 17 different classes of repeats, and the one with the lowest diversity was *S. guianensis* (SGU 2710), with only seven different classes (see **Table 2**).

Athila (LTR – Ty3/gypsy) retroelements were by far the most abundant class of TEs found in all genomes, showing in all samples over 24% of genomic abundance (**Table 2**). Despite the high abundance of Athila found in *Stylosanthes* genomes, no clear relationship between diploids and allopolyploids was observed. The second most abundant class of repeat found in all genomes was the SIRE family, which belongs to the LTR–Ty1/copia clade (**Table 2**). In contrast to Athila, SIRE (LTR–Ty1/copia) abundance showed a stronger correlation between diploids and allopolyploids, where *S. scabra* complex showed a clear higher abundance of these repeats compared to *S. capitata* complex (**Supplementary Figure S3**). Moreover, the RE2 comparative analysis showed that indeed both total repeat and satellitome composition have similar content and abundance among phylogenetically related species (**Supplementary Figure S4**).

Phylogenetic Relationships Based on Different Approaches

In order to evaluate the use of alignment and assembly free analyses in the characterization of *Stylosanthes* allopolyploids, we have compared different approaches with more conventional alignment-based ones. Firstly, phylogenomic trees were constructed only for diploids species (*S. hamata* SHA 2701, *S. seabrana* SSE 2709, *S. pilosa* SPI 2706, *S. macrocephala* SMA 2707, and *S. viscosa* SVI 2702) based on alignment-dependent approaches. For Bayesian Inference (BI)/Maximum Likelihood (ML) we analyzed the following data sets: whole plastomes (**Figure 4A**) and nuclear rDNA sequences alignments (**Figure 4B**). To explore the AAF approach, that analyzes directly NGS data, we analyzed

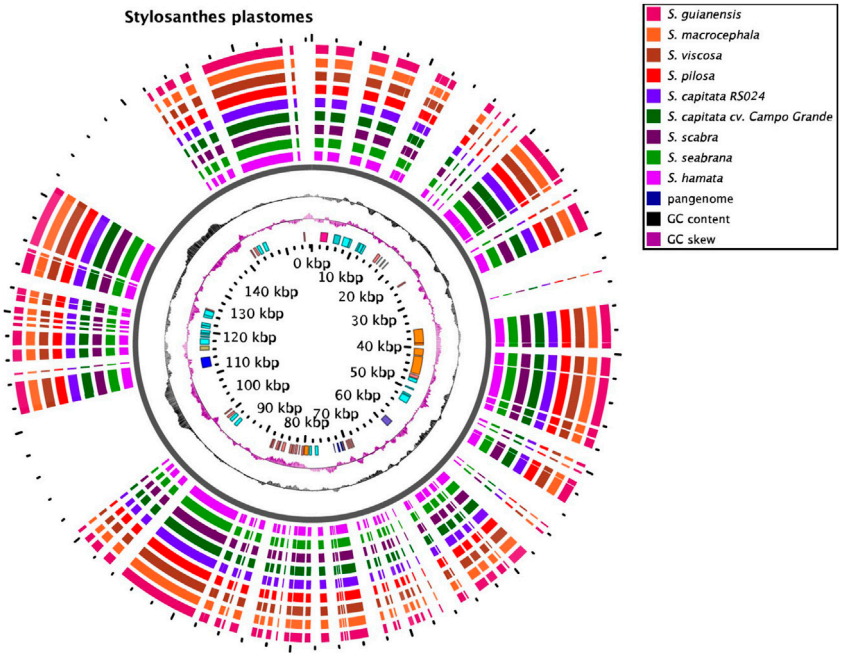

FIGURE 3 | Plastome features of Stylosanthes. Annotated sequences are shown in the inner circle.

mapped reads from both plastomes and rDNA clusters as well as mitogenome, satellite and total reads (**Figure 4C**). All the diploid trees showed the same topology with high support showing *S. guianensis* SGU 2710 as the first diverging lineage of *Stylosanthes* and two main clades: *S. hamata* SHA 2701 + *S. seabrana* SSE 2709 (**Figure 4** in blue) and *S. pilosa* SPI 2706 (*S. macrocephala* SMA 2707 + *S. viscosa* SVI 2702) (**Figure 4** in red). Then, phylogenetic analyzes were performed including also the tetraploid samples (*S. capitata* SCA 2705, *S. capitata* SCA 2708, and *S. scabra* SSC 2703). In this case, we compared the plastidial and rDNA topologies using BI/ML, Splitstree4 network (considering potential reticulated evolution of rDNA), and AAF. We founded two main topologies. In the first one [named here as plastome-like] the allotetraploids *S. capitata* SCA 2705 and *S. scabra* SSC 2703 were positioned in the *S. hamata* + *S. seabrana* clade and the other sample of *S. capitata* (SCA 2708) was related to the *S. macrocephala* + *S. pilosa* + *S. viscosa* clade (**Figures 5A–C**). The second topology [named here as rDNA-like] oppositely positioned the allotetraploids: *S. capitata* SCA 2708 with *S. hamata* + *S. seabrana* and *S. capitata* SCA 2705 and *S. scabra* SSC 2703 with *S. macrocephala* + *S. pilosa* + *S. viscosa* clade (**Figures 5D,E**). Surprisingly, the topologies generated by AAF using different sets of reads (mitogenome, satellite, and total reads) were plastome-like (**Figure 5F** and **Supplementary Figure S5**). Because both *S. scabra* (SSC 2703) and *S. capitata* (SCA 2705) showed very similar genome structures, we were interested to learn if they had similar origin times. Therefore, we dated the plastome phylogeny (**Figure 5A**; **Supplementary Figure S6**). Indeed both *S. scabra* (SSC 2703) and *S. capitata* (SCA 2705) allopolyploids revealed a very recent origin time (0.61 Mya) compared to a more ancient origin for the *S. capitata* (SCA 2708), which revealed an origin time of at least 4.49 Mya.

Satellite DNAs Characterization

The abundance of satDNA repeats varied in a species-specific manner from 0.17 to 4.8%, being the higher values observed in *S. hamata* and *S. scabra*, which showed 3.26% e 4.83%, respectively (**Table 3**). Despite the variance in genomic abundances of satDNA, there was no positive correlation between abundance and number of different satDNA familes found. For instance, *S. hamata* presented only two different satDNA repeats, with high genomic abundance, while *S. pilosa* for instance showed the highest number of different satDNA repeats, but relatively low genomic abundance (**Table 3**). Clearly, species of the clade I showed very few satDNA repeats (1–2) in *S. hamata* and *S. seabrana*, while species of the clade II showed higher numbers of different satDNA repeats (**Figure 6**).

Figure 6 shows the satellitome diversity in *Stylosanthes* and the genomic abundance of each satDNA consensus repeat in each species. Based on the comparative analysis of the RepeatExplorer and AAF phylogeny we have identified a total of 28 satDNAs. For instance, a tandem repeat found only in species of the clade I (*S. hamata* and *S. seabrana*) is called "StyloSat1-variant I" with a monomer length of 165 bp. Thus, based on the grouping pattern of each satDNA we were able to identify possible synapomorphies in some clades and establish relationships between diploids and allopolyploids (**Figure 6**). Taking advantage of this approach, we could identify that the presence of StyloSat1-variant I in the allopolyploids *S. capitata* SCA 2705 and *S. scabra* SSC 2703 reveals the likely participation of species of clade I (*S. hamata*

TABLE 2 | Genomic abundances (%) of the main repetitive sequences identified in the genomes of *Stylosanthes* and *A. hypogaea*. Bold values indicate the sum of each individual group of repeats as well as the total sum of repeat abundance.

Classes		Family	Genomic abundance (%)									
			S. hamata (SHA 2701)	**S. viscosa (SVI 2702)**	**S. scabra (SSC 2703)**	**S. capitata (SCA 2705)**	**S. pilosa (SPI 2706)**	**S. macrocephala (SMA 2707)**	***S. capitata* RS024 (SCA 2708)**	***S. seabrana* (SSE 2709)**	**S. guianensis (SGU 2710)**	***A. hypogaea***
LTR Ty1/Copia		SIRE	14,950	9,882	13,880	11,680	5,084	1,564	4,871	12,250	0.275	1,0.097
		Ikeros						0.022				0.356
		Bianca	1,208		2,418	1,938	1,133	0.449	1,679	0.750	0.027	0.541
		Ale		0.138	0.048	0.187	0.065	0.192	0.024			0.159
		TAR		0.042				0.138				0.022
		Ivana		0.015	0.157	0.222	0.021	0.115	0.049			
		Tork	0.204	0.756	0.495	0.757	0.575	1,100	0,0.398	0.276		0.185
		Total	**16,362**	**10,833**	**16,998**	**14,784**	**6,878**	**3,58**	**7,021**	**13,276**	**0,302**	**2,36**
LTR Ty3/Gypsy	non-chromovirus	Athila	39,450	31,890	31,730	31,780	24,750	27,220	34,670	37,940	35,220	46,170
		Ogre	0.294	0.106	0.310	0.728	0.397	0.177	0.028	0.088		
		Retand	0.286	1,904	0.401	1,303	1,959	3,021	1,573	0.433	1,424	2,929
	chromovirus	Tekay	0.151	4,999	0.738	4,002	9,799	6,794	5,919	0.194	0.920	3,565
		Galadriel	0.085		0.128	0.147		0.056	0.018	0.087		0.052
		CRM	0.014			0.118	0.091		0.092	0.037		
		Total	**40,28**	**38,899**	**33,307**	**38,078**	**36,996**	**37,268**	**42,3**	**38,779**	**37,564**	**52,716**
LTR non-classified		LTR		2,112								
		Ty1_copia							0.016			
Non-LTR		pararetrovirus		0.264		0.112	0.119		0.055			0.299
		LINE	0.033	0.311	0.422	0.730	0.107	0.367	0.378			0.369
Total			**56,675**	**52,419**	**50,727**	**53,704**	**44,1**	**41,215**	**49,77**	**52,055**	**37,866**	**55,744**
DNA transposons		MuDR_Mutator	0.255	1,400	0.333	0.811	1,148	1,393	0.909	0.155	0.744	0.379
		EnSpm_CACTA	0.206	0.618	0.119	0.555	0.548	0.375	0.276	0.419	1,182	
		PIF_Harbinger					0.030	0.025	0.016			0.014
		hAT						0.041				
		Helitron		0.030		0.025						0.039
DNA transposons non-classified		Classe_1	0.536	0.285	0.617	0.420	0.558	0.959	0.319	0.520	0.033	
Total			**0,997**	**2,333**	**1,069**	**1,811**	**2,284**	**2,793**	1,52	1,094	**1,959**	**0,432**
rDNA		45S_rDNA	0.426	0.666	0.312	0.706	0.981	3,551	1,070	1,236	0.658	0.805
		5S_rDNA	0.018	0.051	0.034	0.093	0.152	0.185	0.107	0.064	0.139	0.270
Satellite DNA			3,269	1,038	4,835	0.450	0.852	1,079	0.931	0.173	0.693	1,484
Non-classified			5,459	6,002	11,360	10,290	5,464	12,660	11,180	10,010	17,730	8,778
Total of Repeats			**66,844**	**62,509**	**68,337**	**67,054**	**53,833**	**61,483**	**64,578**	**64,632**	**59,045**	**67,513**

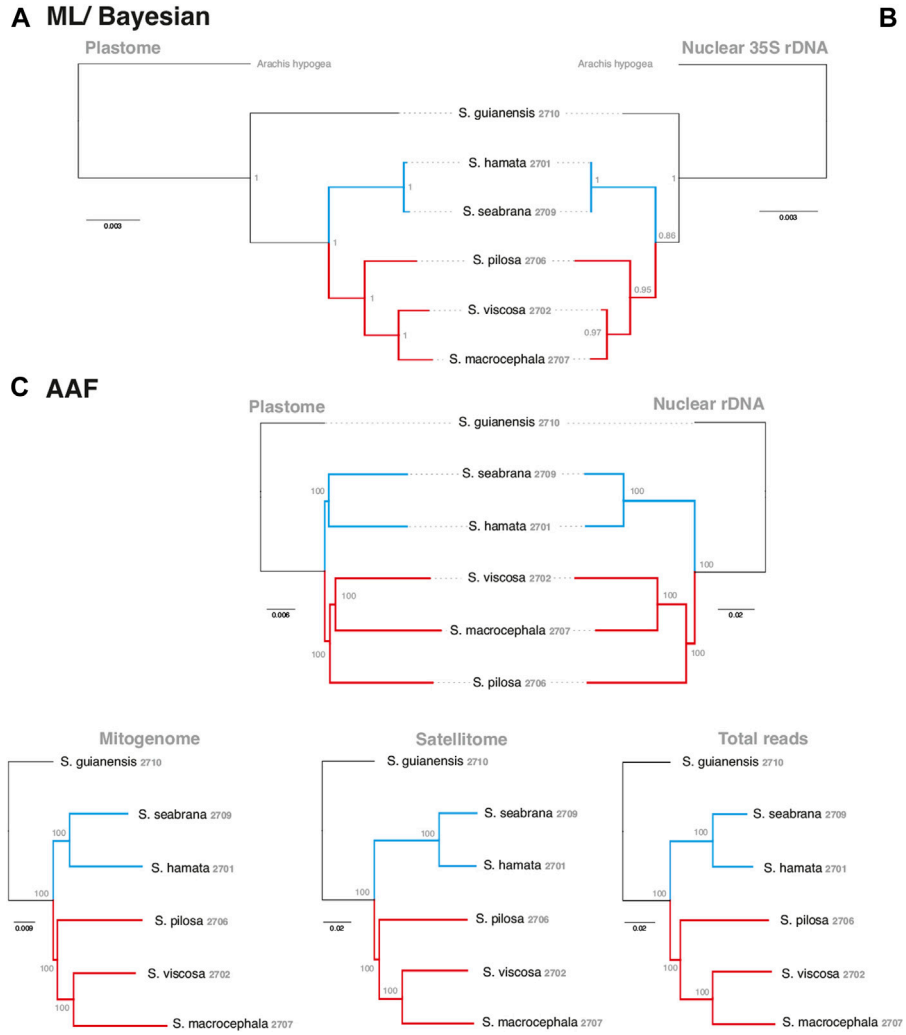

FIGURE 4 | Alignment-based **(A–B)** and AAF **(C)** phylogenies of the *Stylosanthes* diploid species based on different datasets. Each dataset used in the respective inference is indicated on the top of each tree. Blue branches indicate A genome-specific and red branches group other genome types. Numbers indicate specific codes for each sample (see **Table 1**).

+ *S. seabrana*) in the origin of these hybrids. On the other hand, the presence of the clade II specific satDNAs (e.g. StyloSat1-variant II, StyloSat2-497bp, StyloSat3-460bp, and StyloSat12-378bp) in these allopolyploids indicates the additional contribution of species from clade II to the origin of these hybrids. The allotetraploid *S. capitata* SCA 2708 showed a satellitome composition more similar to species from clade II, specifically to *S. pilosa* SPI 2706, to which it shares the satDNAs StyloSat1-variant III, StyloSat5-111bp e StyloSat14-1115bp among others.

To test whether the differential accumulation of satDNA repeats observed in *Stylosanthes* allopolyploids correspond to their origin, we selected three satDNA repeats exclusively found in *S. viscosa* (StyloSat1-variant II-165bp and StyloSat18-129bp), *S. hamata*, and *S. seabrana* (StyloSat1-varian I-165bp) and test for their presence and distribution in *S. scabra* 2703. Based on our above analysis of satDNA sequences, we know that StyloSat1-variant I-165bp and StyloSat1-variant II-165bp actually belongs to the same family of satDNA sharing 70% of sequence similarity after alignment of their consensus sequences (**Figure 6**; **Supplementary Figure S7**). After FISH hybridization StyloSat18-129bp showed proximal signals in a single pair of *S. viscosa* chromosomes while StyloSat1-variant II-165bp was seen at the centromeric region of all chromosome pairs (**Figure 7A**). StyloSat1-varian I-165bp showed signals in centromeric regions of all *S. hamata* (**Figure 7B**) and *S. seabrana* (**Figure 7C**) chromosomes, reinforcing the genomic similarity of these two species. In *S. scabra* we could also confirm the presence of a single chromosome pair showing centromeric signals for StyloSat18-129bp (**Figure 7D**). Remarkably, StyloSat1-variant I-165bp and StyloSat1-variant II-165bp satDNA repeats were also seen at centromeric regions of *S. scabra* chromosomes, where most chromosomes showed either one of the repeats (**Figure 7E**). However, two chromosome pairs

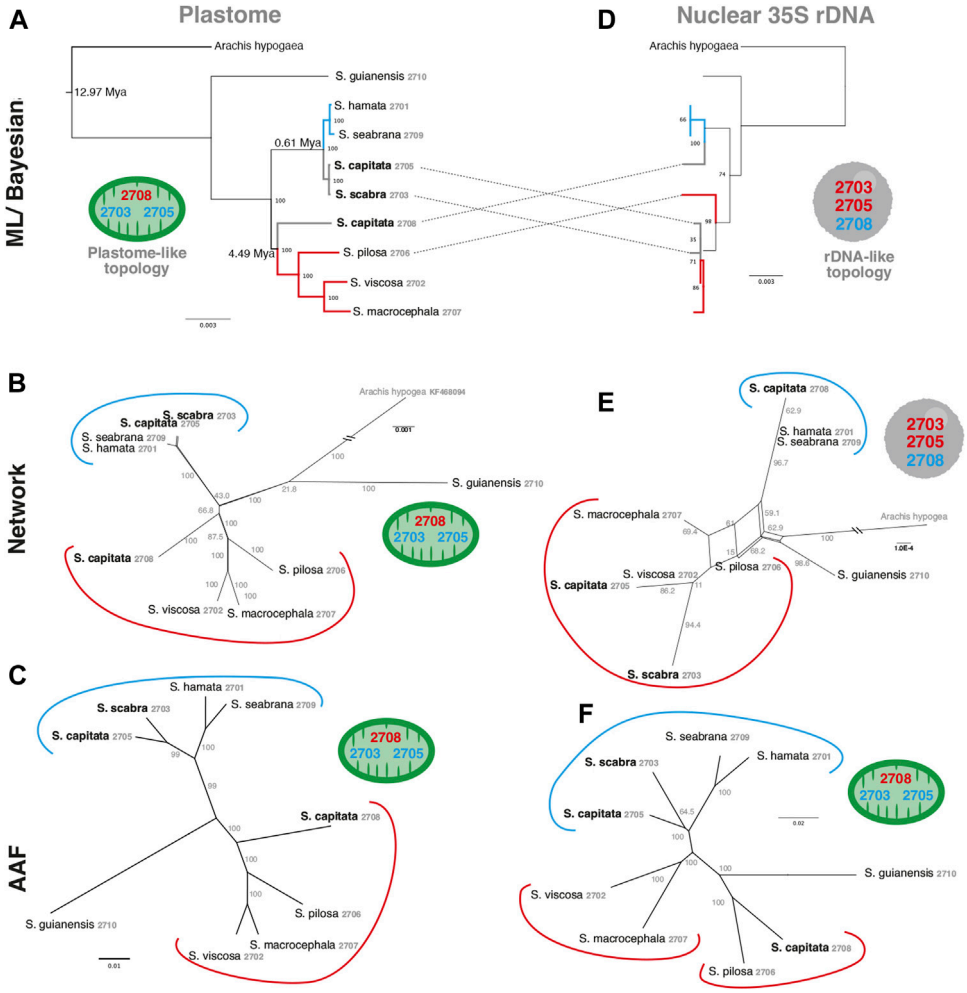

FIGURE 5 | A comparison of alignment-based **(A**,**D)**, network analysis **(B**,**E)** and AAF **(C**,**F)** phylogenies of plastome **(left)** and rDNA **(right)** sequences including *Stylosanthes* allopolyploids. Ages of allopolyploid origins are indicated on **Figure 5A**. Blue branches indicate A genome-specific and red branches group other genome types. Numbers indicate specific codes for each sample (see **Table 1**). Green and gray circles indicate chloroplast-like or nuclear-like topologies, respectively.

TABLE 3 | Comparison of genomic abundance (%) of satDNA repeats, by mapping all reads to consensus sequences identified by the TAREAN tool.

Species	SatDNA genomic abundance (%)	Number of satDNA repeats	Genome
S. hamata (SHA 2701)	2.90	2	A
S. viscosa (SVI 2702)	1.06	7	B
S. scabra (SSC 2703)	3.22	7	AB
S. capitata (SCA 2705)	1.91	7	AB
S. pilosa (SPI 2706)	0.86	8	E
S. macrocephala (SMA 2707)	1.64	8	D
S. capitata (SCA 2708)	1.00	7	DE
S. seabrana (SSE 2709)	0.66	1	A
S. guianensis (SGU 2710)	0.50	4	G

showed both repeats sitting at their centromeric regions (**Figure 6E**, arrowheads). Although in these two pairs the signals are found on their (peri)centromeric regions they do not overlap completely, suggesting that they are indeed present in the same chromosome and it is not a cross-hybridization artefact. Also, if a cross-hybridization had occurred, we should expect it for most chromosomes and not for these two pairs specifically.

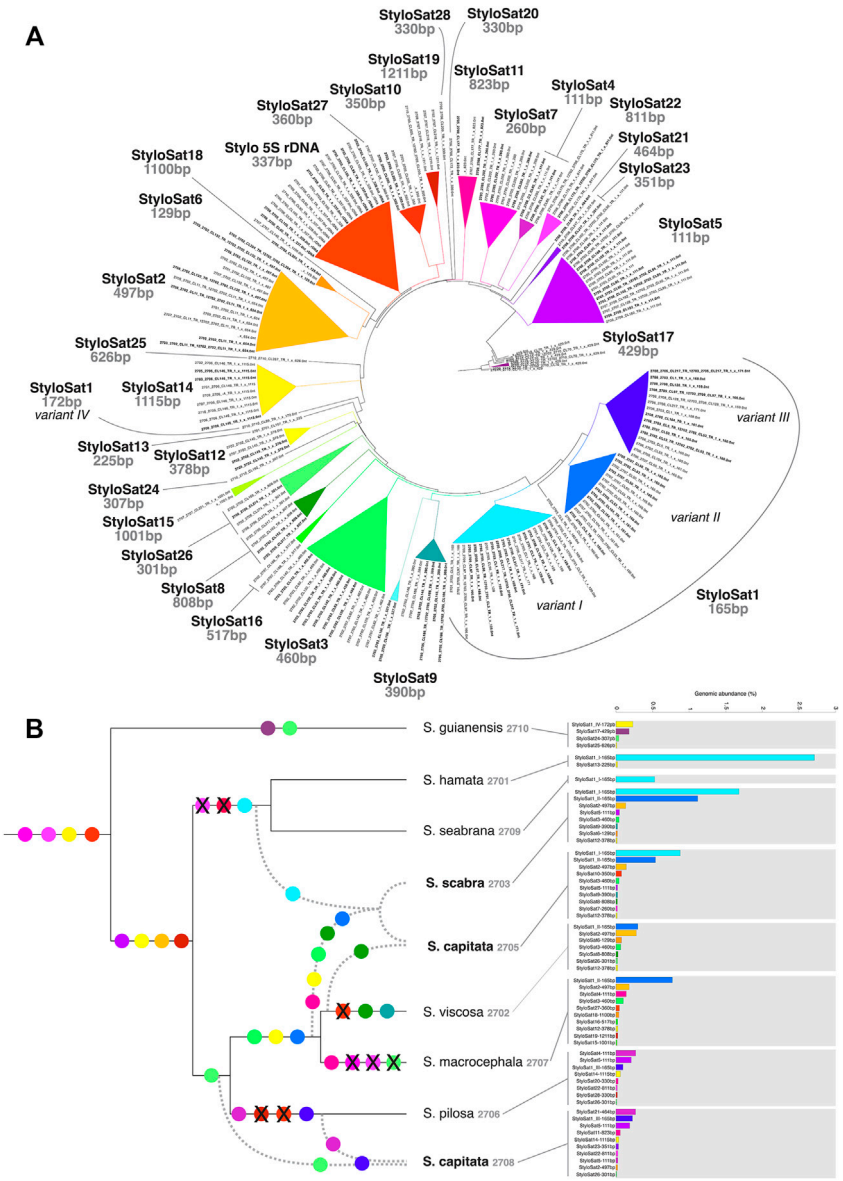

FIGURE 6 | Satellitome characterization of *Stylosanthes*. **(A)** AAF tree showing the grouping of all different satDNA families found in *Stylosanthes*. StyloSat1 family is the most abundant and shows specific lineage similarity, which is conserved in allopolyploids. **(B)** Overview of satDNA showing more than 0.01% of genomic abundance in each sample identified by TAREAN. SatDNA consensus of each sample obtained by TAREAN/RepeatExplorer analysis were used as reference to retrieve repeat reads mapped uniquely to each consensus. The draw tree reflects the most parsimonious species relationships based on all our analyses.

DISCUSSION

The Use of Repeat-Based AAF to Understand Phylogenetic Relationships in Stylosanthes: Clarity in Diploids Versus Uncertainty in Allopolyploids

Our data show a strong congruence in the phylogenetic relationships of *Stylosanthes* diploid species, regardless of the phylogenetic reconstruction method (BI, ML, and AAF) or dataset (rDNA, plastome, mitogenome, satellitome, or total reads). This suggests that, at the diploid level, the reconstruction of fully bifurcate topology results from an expected hierarchical process of speciation. The diploid clades are morphologically supported. *Stylosanthes guianensis* differs from the other species by the fruit having only one fertile article, a scanty rostrum, and the epidermis of the fruit covered with papillae. The other present two fertile articles, the lower article pubescent, the upper article without papillae and developed rostrum. *S. hamata* and *S. seabrana* share many characters as the presence of a rudimentary axis, 2 inner bracteoles, ellipsoid spikes, uncinate rostrum, viscid bristles on the stems and erect

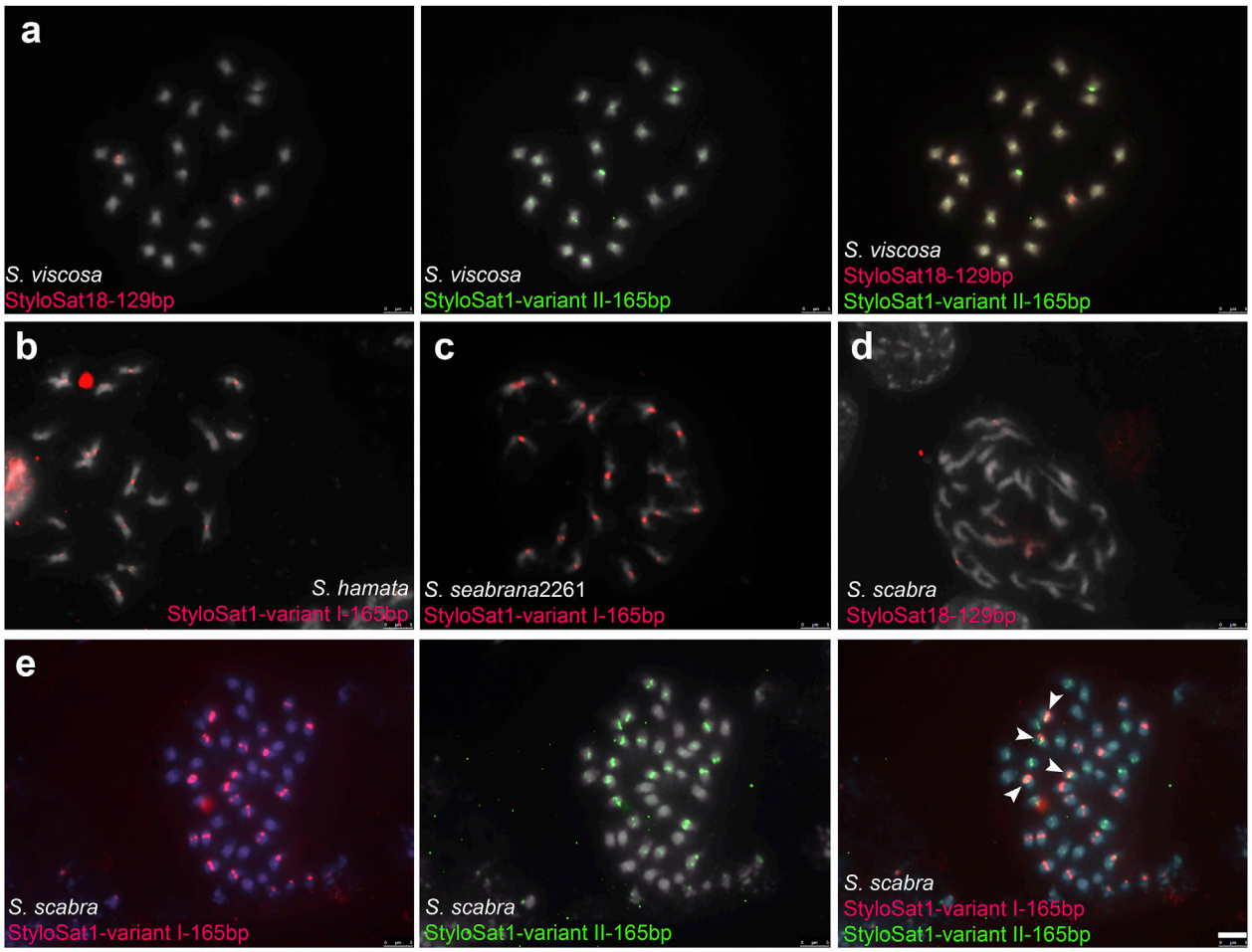

FIGURE 7 | FISH in Stylosanthes scabra complex. **(A)** FISH in *S. viscosa* with *S. viscosa*-specific satDNA repeats StyloSat18-129bp and StyloSat1-variant II-165bp. FISH in *S. hamata* **(B)** and *S. seabrana* **(C)** with *S. hamata/seabrana*-specific satDNA repeat StyloSat1-variant I-165bp. FISH in *S. scabra* with StyloSat18-129bp **(D)** and with StyloSat1-variant I-165bp and StyloSat1-variant II-165bp **(E)** satDNA repeats. Bar = 5 μm.

habit. They differ mainly by the rostrum exceeding the loment length in *S. hamata*, while in *S. seabrana*, the rostrum is shorter (Maass and Mannetje, 2002; Costa, 2006). *Stylosanthes macrocephala*, *S. viscosa*, and *S. pilosa* share fruits with prominent reticulation (Costa, 2006). On the other hand, *S. macrocephala* and *S. pilosa* presents the rudimentary axis, while it's absent in *S. viscosa*. *Stylosanthes viscosa* and *S. pilosa* present coiled rostrum, while it is uncinate in *S. macrocephala*. These results support other phylogenetic studies that indicate that the presence of the axis is not a valuable character to determine infrageneric classifications as it was previously used in taxonomic revisions (Mohlenbrock, 1958; Vander Stappen et al., 1999).

The inclusion of allopolyploids results in low support values, phylogenetic uncertainty, and incongruence between nuclear and organellar topologies, as reported in other studies (Marques et al., 2018; Souza et al., 2019). Establishing synapomorphies for these clades including the allopolyploids is very challenging regarding the different topologies between plastid and nuclear markers. Furthermore, the non-monophyletism of *S. capitata* brings difficulties to recognize morphological traits unique to the clades. However, analyzing the plastome phylogeny (**Figure 5**), we can observe that the clade *S. hamata* + *S. seabrana* + *S. scabra* share exclusive features like uncinate rostrum in the fruit, the upper article densely pubescent, with long non-glandular trichomes, and the outer bracteole trifid. The clade *S. capitata* + *S. macrocephala* + *S. viscosa* + *S. pilosa* can be supported by the bifid outer bracteole and persistent leaflet at the bracts. The bracts of *S. macrocephala* and *S. capitata* are wider than longer, meanwhile the bracts of *S. pilosa* and *S. viscosa* don't reach more than 10 mm. Perhaps the inclusion of more species in the phylogeny may shed light to the synapomorphies of these clades. Nevertheless, the plastome topology better represents the morphology than the nuclear tree.

A series of phylogenetic studies using the Sanger sequencing approach (based on nuclear ITS or plastid loci) has failed to achieve well-resolved topologies for the genus *Stylosanthes* (Liu et al., 1999; Vander Stappen et al., 1999; Vander Stappen et al., 2002), which is probably related to the inclusion of these allopolyploids in the analyzes and low diversification of the sequences and markers used. Approximately 15% of speciation

events in angiosperms have been estimated to be associated with polyploidy (Wood et al., 2009) and allopolyploidy has long been considered to be one of the most frequent events (Barker et al., 2016; Doyle and Sherman-Broyles, 2017; Oxelman et al., 2017). Failure to account for allopolyploidy when reconstructing the past evolution of groups where it has occurred inevitably will lead to inaccurate phylogenetic hypotheses. For cases of uncertain phylogenetic relationships, as in *Stylosanthes*, network analysis has been proposed, considering that the algorithms in this type of analysis assume reticulated evolution (Oxelman et al., 2017; Souza et al., 2019).

Alignment-based phylogenies are still the main method to reconstruct phylogenetic relationships among species (Delsuc et al., 2005; Yang and Rannala, 2012), whereas they are a limited approach in allopolyploid-rich groups. Although well-supported species relationships can be resolved with such approach, a clear limitation when aligning huge numbers or highly divergent sequences, like repeats, is still a major barrier. Thus, most alignment-based phylogenies are based on single genes or more robust ones based on target capture datasets or whole organelle alignments (Saarela et al., 2018; Johnson et al., 2019; Liu et al., 2019). More recently, alternative methods based on alignment and assembly free phylogenetic reconstructions have emerged to deal with this problem (Fan et al., 2015; Lau et al., 2019; Sarmashghi et al., 2019). However, to date, only a single tool is able to construct phylogenetic trees as well as report supporting values for these trees, the AAF algorithm (Fan et al., 2015). Remarkably, the use of alignment and assembly free phylogenies has not yet been explored specifically for repeat sequences. Repeat-based phylogenies have recently become the focus of several studies and are quite accurate in reporting the right phylogenetic relationship among species. The application of those methods is normally based on either repeat abundance (Dodsworth et al., 2015; McCann et al., 2020) or repeat similarity (Vitales et al., 2020). Here, we have shown that AAF-based phylogenomics can be applied as a powerful tool to analyze both sets of WGS reads as well as repeat- and organelle-derived reads. All our AAF phylogenies revealed strongly supported clades, mostly in agreement with alignment-based phylogenies. Furthermore, we propose that the AAF approach could be nicely combined with the output of different repeat identification tools, like RepeatExplorer, where isolated reads from specific repeat lineages can be combined to generate repeat-based phylogenies (e.g., satDNA, LTR retrotransposons, total repeats, etc.). AAF phylogenies can also be applied in order to identify satDNA families and overall satDNA diversification in specific groups as well as other repeat families, as shown here.

The existence of incongruent nuclear (rDNA-like) and organellar (plastome-like) phylogenetic topologies was already demonstrated for *Stylosanthes* (Marques et al., 2018). However, it was remarkable that different repeat sets (mitogenome, satellitome, total reads, and total repeats) revealed similar plastome-like topologies [except 5S rDNA reads that generated a unique topology (**Supplementary Figure S5**)]. Assuming that this topology reflects a maternal genealogy, we found evidence of maternal bias in the repeat abundances. As part of the "genomic shock" experienced by combining two divergent subgenomes within one neoallopolyploid nucleus, there are predictions regarding the level of genome reorganization, and sequence loss/retention based on the direction of the cross (Dodsworth et al., 2020). The maternal subgenome is expected to be favored, relative to the paternal subgenome, due to its compatibility within the maternal cytoplasm, as suggested by Nuclear Cytoplasmatic Interaction (NCI) hypothesis (Lim et al., 2004). This potentially leads to specific degradation of various elements from the paternal subgenome, as predicted by this NCI hypothesis (Lim et al., 2004; Renny-Byfield et al., 2011; Renny-Byfield et al., 2012).

Apparently, the allopolyploid age can influence the restructuring of the genome and turnover of repetitive elements. In older tetraploids, repeat dynamics are much more variable, impacting for example on genome size. However, in young allopolyploids repeat abundances are close to the sum of abundances expected from both parental donors (Dodsworth et al., 2020). Thus, the uncertainty in identifying the parents of an allopolyploid is proportional to the age of formation of that hybrid (McCann et al., 2018). We demonstrate here that the origin of *S. scabra* 2703 + *S. capitata* 2705 is more recent (0.61–1.71 Mya) than *S. capitata* 2708 (3.01–4.49 Mya), which may explain the greater complexity in detecting the parents of this last allotetraploid.

SatDNA Evolution in Stylosanthes

Satellite DNA has been used as an important source for subgenome identification in allopolyploids and as phylogenetic information (Gill et al., 2009; Koukalova et al., 2010). Here we found that satellite DNA showed a great diversification in the genus *Stylosanthes* and only a single family (StyloSat1) was found in all samples analyzed and seems to be conserved among the species, most likely representing the centromeric DNA. Indeed, StyloSat1 demonstrated a subgenome-specific sequence conservation and allowed us to detect the A and B subgenomes of *S. scabra*. Also, *S. hamata* and *S. seabrana* showed similar hybridization profiles with StyloSat1, although the *S. hamata* specific satDNA StyloSat13, which was observed in a single chromosome pair, was not identified in *S. seabrana*. The high abundance, phylogenetic conservation and physical location in the proximal region of the chromosomes suggests that this satellite is the centromeric DNA of these species.

Most of the other satDNA families were shared only among species with a high degree of relationship, in this case between hybrids and their genome donors. Cases of absence of satellite DNAs in certain genomes inconsistent with phylogeny (see **Figure 6**) can be explained by the library hypothesis (Fry and Salser, 1977). Remarkably, satDNA sequences showed a high phylogenetic signal in our AAF analyses and helped to identify diploid progenitors. Similarly, a recent study also reported a high phylogenetic signal for satDNA abundance among *Melampodium* L. (Asteraceae) species (McCann et al., 2020). Our AAF approach seems highly relevant and robust for repeat-based phylogeny since it takes into account both similarity and abundance as well as reporting a support value for the phylogenetic tree.

The Macrostructure of Mitochondrial Genomes and Genomes Reflect Completely Different Evolutionary Histories in Stylosanthes

Here we make available for the first time the complete mitogenomes for nine species and further provides additional six plastomes for *Stylosanthes* species, ensuring fundamental information for the next studies of systematics and genetic breeding of the genus. Mitochondrial genomes of the studied species of *Stylosanthes* have all similar genic contents despite a low density of genic sequences like most plants. Variations in genome sizes accounted for the most diversity found among the characterized *Stylosanthes* mitogenomes. These large variations are common; however, they do not directly indicate the number of functional genes, yet they are more related to intergenic regions. Indeed, the mitochondrial genome of higher plants shows extreme variations in its structure, size, and complexity (Gualberto and Newton, 2017). Specific features of mitochondrial genomes such as their high rates of recombination can account for several aspects of plant mitochondrial genome structure. Thus, playing a major role in the evolution of plant mitochondrial genomes (Arrieta-Montiel and Mackenzie, 2011; Knoop et al., 2011; Kuhn and Gualberto, 2012). Based on our synteny maps for mitogenomes, a similar situation to plastomes was observed, with the samples *S. hamata*, *S. seabrana* showing the highest synteny, confirming their close relationship. Also, *S. scabra* and *S. capitata* (SCA 2705) shared high synteny to both A genome diploid mitogenomes, confirming their maternal A genome origin. Therefore, we reinforce the use of organelle genomes for evolutionary and phylogenetic studies in this group.

The *Stylosanthes* chloroplast genomes were very conserved in genetic content, total length, and organization, as is found for other plant groups as well (Daniell et al., 2016). We confirmed our previous findings that a diploid species with A genome (*S. hamata* or *S. seabrana*) is likely the maternal genomic donor of *S. scabra*, as their plastomes shared 99.798 and 99.776% pairwise identity, respectively. Additionally, the *S. capitata* (SCA 2705) most likely has the same maternal progenitor as *S. scabra* due to a higher similarity of its organelle genetic information with A genome species *S. hamata* and *S. seabrana*, 99.803 and 99.783%, respectively. In contrast, it showed less pairwise identity with *S. pilosa* and *S. macrocephala*, 98.161%, and 98.242%, respectively. Although *S. scabra* and *S. capitata* (SCA 2705) shared the highest pairwise identity in their plastomes, the divergence in their mitogenomes suggests that these allopolyploids have independent origins. Furthermore, the *S. capitata* (SCA 2708), which was previously indicated to have DDEE genome composition (Liu and Musial, 2001) did not show the same level of sequence similarity to its putative genome donors *S. macrocephala* (98.317%, D genome) or *S. pilosa* (98.334%, E genome), but it grouped in the same clade with both and *S. viscosa* in the organelle genome phylogenies.

Genomic Repetitive Fraction Characterization and Comparative Analysis of Repeat Abundance

All nine samples of the genus *Stylosanthes* Sw. characterized in this study showed a high abundance of repetitive DNA and maintain the conservation of many sequences. Ty3/gypsy Athila elements were the most abundant in all species. Ty1/copia SIRE elements were highly abundant in species with A genome composition, clearly being a shared feature of A genome species, while in the other genomes this element showed variable lower abundances. Among all genomes, *S. guianensis* showed the highest divergence within the genus and agrees with its more distant relationship with the other species. In general, the overall repeat abundance matched the species relationships, while polyploids tend to accumulate repeats from both diploid progenitors.

In general, species with A genome had more abundance of total satDNA but less satDNA diversity compared to the species with other genome types. SatDNA abundances varied greatly between the two very closely related *S. hamata* and *S. seabrana*. Although the major family of *Stylosanthes* satDNA (StyloSAT1) was found in both species, its abundance varied from 3% in *S. hamata* to 0.7% in *S. seabrana*, suggesting that an amplification or deamplification has occurred since the separation of these two species. However, our FISH analysis confirmed the presence of this satDNA at the centromeric regions of all chromosomes in both species.

Taxonomic Conflicts

Concerning the samples that compose the cv. Campo Grande, *S. capitata* (SCA 2705) and *S. macrocephala* (SMA 2707), which has been marketed by Embrapa Beef Cattle since 2000 (EMBRAPA, 2000), we suggest that a taxonomic review should be carried out since our studies show some inconsistencies with these taxonomical designations. *S. capitata* is morphologically similar to *S. macrocephala* and *S. pilosa* by the craspedodromous leaflets, suborbicular bracts, wider than longer, two long, fertile loments, and uncinate rostrum. *Stylosanthes macrocephala* and *S. capitata* were even considered synonymous in some studies (Vanni, 2017). Our results corroborate the distinction of the two species, but brings attention to the proximity of *S. capitata* (SCA 2708) and *S. pilosa*. Both species are very alike, sometimes hard to differentiate. The main characteristic to distinguish them is the indument that is pilose in *S. pilosa* with long bristles and trichomes (Ferreira and Costa, 1977). Further taxonomic studies may elucidate the difference between them. In our study, we found that *S. capitata* (SCA 2705) is more genetically similar to *S. scabra*. The two species are distinguishable by bract width (wider in *S. capitata*, with more than 10 veins) and stem indument (scabrous with short bristles in *S. scabra*). Therefore, a taxonomic review is necessary so that consumers are sure of the product they are purchasing, and also so that the company can adapt the recommendations for that particular cultivar, according to the species that actually compose it.

Despite the divergence between some of the results we obtained and those in the literature, as in the case of *S. seabrana*, it cannot be said whether it is a distinct species or a synonym of *S. hamata* in the process of speciation. Taking all our findings into consideration we confirm the complexity and difficulty of studies in the genus *Stylosanthes* and reinforces the need for more in-depth studies that review the taxonomy and phylogenetics of the group, and phylogenetics of the group in the light of genomic data.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MASO performed repeat analysis. TN, MAS and DFG performed organelle genome assembly, annotation and analysis. IC and BV-L performed FISH experiments. RSO, MFS, and DSG collected, identified and provided the plant material. SSMS and GSAL helped with manuscript writing and discussion. CCSA, GS and AM performed phylogenetic analysis, designed the project and wrote the manuscript.

FUNDING

This work was supported by the Alagoas State Research Support Foundation (FAPEAL) and by the National Council for Scientific and Technological Development (CNPq) grant number 408100/2018-4. A DCR fellowship was granted to AM by the CNPq. AM is currently supported by the Max-Planck-Society. The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.727314/full#supplementary-material

REFERENCES

Arrieta-Montiel, M. P., and Mackenzie, S. A. (2011). Plant Mitochondrial Genomes and Recombination. *Plant Mitochondria* 1, 65–82. doi:10.1007/978-0-387-89781-3_3

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

Cameron, D. F., and Chakraborty, S. (2004). "Forage Potential of *Stylosanthes* in Different Production Systems," in *High-yielding Anthracnose-Resistant Stylosanthes for Agricultural Systems*. Editor C. Chakraborty (Canberra, Australia: Australian Centre for International Agricultural Research), 27–38.

Cardoso, D., Pennington, R. T., De Queiroz, L. P., Boatwright, J. S., Van Wyk, B.-E., Wojciechowski, M. F., et al. (2013). Reconstructing the Deep-Branching Relationships of the Papilionoid Legumes. *South Afr. J. Bot.* 89, 58–75. doi:10.1016/j.sajb.2013.05.001

Costa, N. M. S., and Ferreira, M. B. (1984). "Some Brazilian Species of *Stylosanthes*," in *The Biology and Agronomy of Stylosanthes*. Editors H. M. Stace and L. A. Edye (LondonNew York: Academic Press), 49–72. doi:10.1016/b978-0-12-661680-4.50007-x

Costa, N. M. S. (2006). *Revisão Do Género Stylosanthes Sw*. Lisboa, Portugal: Dissertation, Universidade Técnica de Lisboa, 494.

da Costa, L. C., and Valls, J. F. M. (2010). "*Stylosanthes* Sw.," in *Catálogo de plantas e fungos do Brasil*. Editor R. C. E. A. Forzza (Rio de Janeiro: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro), 1090–1091.

Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast Genomes: Diversity, Evolution, and Applications in Genetic Engineering. *Genome Biol.* 17, 134. doi:10.1186/s13059-016-1004-2

Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the Reconstruction of the Tree of Life. *Nat. Rev. Genet.* 6, 361–375. doi:10.1038/nrg1603

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: De Novo Assembly of Organelle Genomes from Whole Genome Data. *Nucleic Acids Res.* 45, e18. doi:10.1093/nar/gkw955

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., Christenhusz, M. J. M., Conran, J. G., Guignard, M. S., Knapp, S., Struebig, M., et al. (2020). Extensive Plastid-Nuclear Discordance in a Recent Radiation of *Nicotiana* Section *Suaveolentes* (Solanaceae). *Bot. J. Linn. Soc.* 193, 546–559. doi:10.1093/botlinnean/boaa024

Dodsworth, S., Pokorny, L., Johnson, M. G., Kim, J. T., Maurin, O., Wickett, N. J., et al. (2019). Hyb-Seq for Flowering Plant Systematics. *Trends Plant Sci.* 24, 887–891. doi:10.1016/j.tplants.2019.07.011

Doyle, J. J., and Sherman-Broyles, S. (2017). Double Trouble: Taxonomy and Definitions of Polyploidy. *New Phytol.* 213, 487–493. doi:10.1111/nph.14276

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed Phylogenetics and Dating with Confidence. *Plos Biol.* 4, e88. doi:10.1371/journal.pbio.0040088

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

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

EMBRAPA (2000). *Estilosantes Campo Grande: estabelecimento, manejo e produção animal*. Campo Grande: Embrapa Gado de Corte.

Fan, H., Ives, A. R., Surget-Groba, Y., and Cannon, C. H. (2015). An Assembly and Alignment-free Method of Phylogeny Reconstruction from Next-Generation Sequencing Data. *BMC Genomics* 16, 522. doi:10.1186/s12864-015-1647-5

Ferreira, M. B., and Costa, N. M. S. (1977). Novas espécies do gênero *Stylosanthes* para Minas Gerais, Brasil. *Anais Congr. Soc. Bot. Brasil* 28, 77–100.

Franco, A. L., Figueredo, A., Pereira, L. M., Sousa, S. M., Souza, G., Carvalho, M. A., et al. (2020). Low Cytomolecular Diversification in the Genus *Stylosanthes* Sw. (Papilionoideae, Leguminosae). *Genet. Mol. Biol.* 43, e20180250. doi:10.1590/1678-4685-GMB-2018-0250

Fry, K., and Salser, W. (1977). Nucleotide Sequences of HS-α Satellite DNA from Kangaroo Rat *Dipodomys ordii* and Characterization of Similar Sequences in Other Rodents. *Cell* 12, 1069–1084. doi:10.1016/0092-8674(77)90170-2

Gastony, G. J., and Yatskievych, G. (1992). Maternal Inheritance of the Chloroplast and Mitochondrial Genomes in Cheilanthoid Ferns. *Am. J. Bot.* 79, 716–722. doi:10.1002/j.1537-2197.1992.tb14613.x

Gernhard, T. (2008). The Conditioned Reconstructed Process. *J. Theor. Biol.* 253, 769–778. doi:10.1016/j.jtbi.2008.04.005

Gill, N., Findley, S., Walling, J. G., Hans, C., Ma, J., Doyle, J., et al. (2009). Molecular and Chromosomal Evidence for Allopolyploidy in Soybean. *Plant Physiol.* 151, 1167–1174. doi:10.1104/pp.109.137935

Greiner, S., Sobanski, J., and Bock, R. (2014). Why Are Most Organelle Genomes Transmitted Maternally?. *Bioessays* 37, 80–94. doi:10.1002/bies.201400110

Gualberto, J. M., and Newton, K. J. (2017). Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annu. Rev. Plant Biol.* 68, 225–252. doi:10.1146/annurev-arplant-043015-112232

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, 176. doi:10.1186/s12864-017-3555-3

Han, F., Fedak, G., Guo, W., and Liu, B. (2005). Rapid and Repeatable Elimination of a Parental Genome-specific DNA Repeat (pGc1R-1a) in Newly Synthesized Wheat Allopolyploids. *Genetics* 170, 1239–1245. doi:10.1534/genetics.104.039263

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. Biodiversity* 5, 277–289. doi:10.1017/s147720000700240x

Huson, D. H., and Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* 23, 254–267. doi:10.1093/molbev/msj030

Jankowiak, K., Rybarczyk, A., Wyatt, R., Odrzykoski, I., Pacak, A., and Szweykowska-Kulinska, Z. (2005). Organellar Inheritance in the Allopolyploid moss *Rhizomnium pseudopunctatum*. *Taxon* 54, 383–388. doi:10.2307/25065367

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., et al. (2019). A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using K-Medoids Clustering. *Syst. Biol.* 68, 594–606. doi:10.1093/sysbio/syy086

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780. doi:10.1093/molbev/mst010

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

Kim, H. T., Lim, K. B., and Kim, J. S. (2019). New Insights on *Lilium* Phylogeny Based on a Comparative Phylogenomic Study Using Complete Plastome Sequences. *Plants (Basel)* 8, 1. doi:10.3390/plants8120547

Knoop, V., Volkmar, U., Hecht, J., and Grewe, F. (2011). Mitochondrial Genome Evolution in the Plant Lineage. *Plant Mitochondria* 1, 3–29. doi:10.1007/978-0-387-89781-3_1

Koukalova, B., Moraes, A. P., Renny-Byfield, S., Matyasek, R., Leitch, A. R., and Kovarik, A. (2010). Fall and Rise of Satellite Repeats in Allopolyploids of *Nicotiana* over C. 5 Million Years. *New Phytol.* 186, 148–160. doi:10.1111/j.1469-8137.2009.03101.x

Kühn, K., and Gualberto, J. M. (2012). Recombination in the Stability, Repair and Evolution of the Mitochondrial Genome. *Mitochondrial Genome Evol.* 63, 215–252. doi:10.1016/b978-0-12-394279-1.00009-0

Lau, A.-K., Dörrer, S., Leimeister, C.-A., Bleidorn, C., and Morgenstern, B. (2019). Read-SpaM: Assembly-free and Alignment-free Comparison of Bacterial Genomes with Low Sequencing Coverage. *BMC Bioinformatics* 20, 638. doi:10.1186/s12859-019-3205-7

Lim, K. Y., Matyasek, R., Kovarik, A., and Leitch, A. R. (2004). Genome Evolution in Allotetraploid *Nicotiana*. *Biol. J. Linn. Soc.* 82, 599–606. doi:10.1111/j.1095-8312.2004.00344.x

Liu, C. J., and Musial, J. M. (2001). The Application of Chloroplast DNA Clones in Identifying Maternal Donors for Polyploid Species of *Stylosanthes*. *Theor. Appl. Genet.* 102, 73–77. doi:10.1007/s001220051619

Liu, C. J., Musial, J. M., and Thomas, B. D. (1999). Genetic Relationships Among *Stylosanthes* Species Revealed by RFLP and STS Analyses. *Theor. Appl. Genet.* 99, 1179–1186. doi:10.1007/s001220051322

Liu, Y., Johnson, M. G., Cox, C. J., Medina, R., Devos, N., Vanderpoorten, A., et al. (2019). Resolution of the Ordinal Phylogeny of Mosses Using Targeted Exons from Organellar and Nuclear Genomes. *Nat. Commun.* 10, 1485. doi:10.1038/s41467-019-09454-w

Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013). OrganellarGenomeDRAW-a Suite of Tools for Generating Physical Maps of Plastid and Mitochondrial Genomes and Visualizing Expression Data Sets. *Nucleic Acids Res.* 41, W575–W581. doi:10.1093/nar/gkt289

Ma, Z. Y., Chandra, A., Musial, J. M., and Liu, C. J. (2004). Molecular Evidence that *Stylosanthes angustifolia* Is the Third Putative Diploid Progenitor of the Hexaploid *S. Erecta* (Fabaceae). *Plant Syst. Evol.* 248, 171–176. doi:10.1007/s00606-004-0179-2

Maass, B. L., and Mannetje, L. t. (2002). *Stylosanthes seabrana* (Leguminosae: Papilionoideae), a New Species from Bahia, Brazil. *Novon* 12, 497–500. doi:10.2307/3393129

Maass, B. L., and Sawkins, M. (2004). "History, Relationships and Diversity Among *Stylosanthes* Species of Commercial Significance," in *High-yielding Anthracnose-Resistant Stylosanthes for Agricultural Systems*. Editor C. Chakraborty (Canberra, Australia: Australian Centre for International Agricultural Research), 9–26.

Mannetje, L. T. (1984). "Considerations on the Taxonomy of the Genus *Stylosanthes*," in *The Biology and Agronomy of Stylosanthes*. Editors H. M. Stace and L. A. Edye (North Ryde, NSW, Australia: Academic Press Australia), 1–21. doi:10.1016/b978-0-12-661680-4.50006-8

Marques, A., Moraes, L., Aparecida Dos Santos, M., Costa, I., Costa, L., Nunes, T., et al. (2018). Origin and Parental Genome Characterization of the allotetraploid *Stylosanthes scabra* Vogel (Papilionoideae, Leguminosae), an Important Legume Pasture Crop. *Ann. Bot.* 122, 1143–1159. doi:10.1093/aob/mcy113

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. USA* 112, 13633–13638. doi:10.1073/pnas.1512255112

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., Macas, J., Novák, P., Stuessy, T. F., Villaseñor, J. L., 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

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534. doi:10.1093/molbev/msaa015

Mohlenbrock, R. H. (1958). A Revision of the Genus *Stylosanthes*. *Ann. MO. Bot. Gard* 44, 299–355. doi:10.2307/2394648

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

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

Novák, P., Neumann, P., and Macas, J. (2010). Graph-based Clustering and Characterization of Repetitive Sequences in Next-Generation Sequencing Data. *BMC Bioinformatics* 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

Oxelman, B., Brysting, A. K., Jones, G. R., Marcussen, T., Oberprieler, C., and Pfeil, B. E. (2017). Phylogenetics of Allopolyploids. *Annu. Rev. Ecol. Evol. Syst.* 48, 543–557. doi:10.1146/annurev-ecolsys-110316-022729

Rambaut, A. (2014). *Fig Tree*. 1.4.2 ed..

Rambaut, A., Suchard, M. A., Xie, W., and Drummond, A. J. (2014). *TRACER*. 1.6 ed..

Reboud, X., and Zeyl, C. (1994). Organelle Inheritance in Plants. *Heredity* 72, 132–140. doi:10.1038/hdy.1994.19

Renny-Byfield, S., Kovařík, A., Chester, M., Nichols, R. A., Macas, J., Novák, P., et al. (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., Chester, M., Kovarik, 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

Saarela, J. M., Burke, S. V., Wysocki, W. P., Barrett, M. D., Clark, L. G., Craine, J. M., et al. (2018). A 250 Plastome Phylogeny of the Grass Family (Poaceae): Topological Support under Different Data Partitions. *Peerj* 6, e4299. doi:10.7287/peerj.4299v0.1/reviews/1

Santos-Garcia, M. O., Karia, C. T., Resende, R. M., Chiari, L., Vieira, M. L., Zucchi, M. I., et al. (2012). Identification of *Stylosanthes* Guianensis Varieties Using Molecular Genetic Analysis. *Aob Plants* 2012, pls001. doi:10.1093/aobpla/pls001

Särkinen, T., Pennington, R. T., Lavin, M., Simon, M. F., and Hughes, C. E. (2012). Evolutionary Islands in the Andes: Persistence and Isolation Explain High Endemism in Andean Dry Tropical Forests. *J. Biogeogr.* 39, 884–900. doi:10.1111/j.1365-2699.2011.02644.x

Sarmashghi, S., Bohmann, K., P. Gilbert, M. T., Bafna, V., and Mirarab, S. (2019). Skmer: Assembly-free and Alignment-free Sample Identification Using Genome Skims. *Genome Biol.* 20, 34. doi:10.1186/s13059-019-1632-4

Soderlund, C., Bomhoff, M., and Nelson, W. M. (2011). SyMAP v3.4: a Turnkey Synteny System with Application to Plant Genomes. *Nucleic Acids Res.* 39, e68. doi:10.1093/nar/gkr123

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

Stace, H. M., and Edye, L. A. (1984). *The Biology and Agronomy of Stylosanthes*. Academic Press.

Vander Stappen, J., De Laet, J., Gama-López, S., Van Campenhout, S., and Volckaert, G. (2002). Phylogenetic Analysis of *Stylosanthes* (Fabaceae) Based on the Internal Transcribed Spacer Region (ITS) of Nuclear Ribosomal DNA. *Plant Syst. Evol.* 234, 27–51. doi:10.1007/s00606-002-0193-1

Vander Stappen, J., Weltjens, I., Munaut, F., and Volckaert, G. (1999). Interspecific and Progeny Relationships in the Genus *Stylosanthes* Inferred from Chloroplast DNA Sequence Variation. *Comptes Rendus de l'Académie des Sci. - Ser. - Sci. de la Vie* 322, 481–490. doi:10.1016/s0764-4469(99)80098-5

Vanni, R. O. (2017). - the Genus Stylosanthes (Fabaceae, Papilionoideae, Dalbergieae) in South America. *Bol. Soc. Argent. Bot.* 52, 549–585. doi:10.31055/1851.2372.v52.n3.18033

Vieira, M. L. C., Aguiar-Perecin, M. L. R. d., and Martins, P. S. (1993). A Cytotaxonomic Study in Twelve Brazilian Taxa of Stylosanthes Sw., Leguminosae. *Cytologia* 58, 305–311. doi:10.1508/cytologia.58.305

Vitales, D., Garcia, S., and Dodsworth, S. (2020). Reconstructing Phylogenetic Relationships Based on Repeat Sequence Similarities. *Mol. Phylogenet. Evol.* 147, 106766. doi:10.1016/j.ympev.2020.106766

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The Frequency of Polyploid Speciation in Vascular Plants. *Proc. Natl. Acad. Sci.* 106, 13875–13879. doi:10.1073/pnas.0811575106

Xie, D.-F., Tan, J.-B., Yu, Y., Gui, L.-J., Su, D.-M., Zhou, S.-D., et al. (2020). Insights into Phylogeny, Age and Evolution of Allium (Amaryllidaceae) Based on the Whole Plastome Sequences. *Ann. Bot.* 125, 1039–1055. doi:10.1093/aob/mcaa024

Yang, Z., and Rannala, B. (2012). Molecular Phylogenetics: Principles and Practice. *Nat. Rev. Genet.* 13, 303–314. doi:10.1038/nrg3186

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 September 2021
doi: 10.3389/fgene.2021.728670

Genomic Differences Between the Sexes in a Fish Species Seen Through Satellite DNAs

*Carolina Crepaldi[1], Emiliano Martí[1], Évelin Mariani Gonçalves[1], Dardo Andrea Martí[2] and Patricia Pasquali Parise-Maltempi[1]**

[1]Departamento de Biologia Geral e Aplicada, Instituto de Biociências (IB), Universidade Estadual Paulista (UNESP), Rio Claro, Brazil, [2]Laboratorio de Genética Evolutiva, Instituto de Biología Subtropical (IBS), Universidad Nacional de Misiones (UNaM), CONICET, Posadas, Argentina

Neotropical fishes have highly diversified karyotypic and genomic characteristics and present many diverse sex chromosome systems, with various degrees of sex chromosome differentiation. Knowledge on their sex-specific composition and evolution, however, is still limited. Satellite DNAs (satDNAs) are tandemly repeated sequences with pervasive genomic distribution and distinctive evolutionary pathways, and investigating satDNA content might shed light into how genome architecture is organized in fishes and in their sex chromosomes. The present study investigated the satellitome of *Megaleporinus elongatus*, a freshwater fish with a proposed $Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$ multiple sex chromosome system that encompasses a highly heterochromatic and differentiated W_1 chromosome. The species satellitome comprises of 140 different satDNA families, including previously isolated sequences and new families found in this study. This diversity is remarkable considering the relatively low proportion that satDNAs generally account for the *M. elongatus* genome (around only 5%). Differences between the sexes in regards of satDNA content were also evidenced, as these sequences are 14% more abundant in the female genome. The occurrence of sex-biased signatures of satDNA evolution in the species is tightly linked to satellite enrichment associated with W_1 in females. Although both sexes share practically all satDNAs, the overall massive amplification of only a few of them accompanied the W_1 differentiation. We also investigated the expansion and diversification of the two most abundant satDNAs of *M. elongatus,* MelSat01-36 and MelSat02-26, both highly amplified sequences in W_1 and, in MelSat02-26's case, also harbored by Z_2 and W_2 chromosomes. We compared their occurrences in *M. elongatus* and the sister species *M. macrocephalus* (with a standard ZW sex chromosome system) and concluded that both satDNAs have led to the formation of highly amplified arrays in both species; however, they formed species-specific organization on female-restricted sex chromosomes. Our results show how satDNA composition is highly diversified in *M. elongatus*, in which their accumulation is significantly contributing to W_1 differentiation and not satDNA diversity per se. Also, the evolutionary behavior of these repeats may be associated with genome plasticity and satDNA variability between the sexes and between closely related species, influencing how seemingly homeologous heteromorphic sex chromosomes undergo independent satDNA evolution.

Keywords: satellitome, concerted evolution, satDNA evolution, neotropical fish, fish sex chromosomes, megaleporinus, anostomidae

OPEN ACCESS

Edited by:
Ricardo Utsunomia, Federal Rural University of Rio de Janeiro, Brazil

Reviewed by:
Pablo Mora Ruiz, University of South Bohemia in České Budějovice, Czechia
Alexandr Sember, Academy of Sciences of the Czech Republic (ASCR), Czechia
Sebastián Pita, Universidad de la República, Uruguay

****Correspondence:***
Patricia Pasquali Parise-Maltempi
patricia.parise@unesp.br

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *21 June 2021*
Accepted: *13 September 2021*
Published: *30 September 2021*

Citation:
Crepaldi C, Martí E, Gonçalves ÉM, Martí DA and Parise-Maltempi PP (2021) Genomic Differences Between the Sexes in a Fish Species Seen Through Satellite DNAs. Front. Genet. 12:728670. doi: 10.3389/fgene.2021.728670

INTRODUCTION

Among the repetitive fraction of a eukaryotic genome, satellite DNAs (satDNAs) are one of the most abundant elements, characterized as tandemly organized sequences that can be amplified into multiple copies in the genome (Charlesworth et al., 1994; Plohl et al., 2012). Selective forces act loosely on satDNAs, as they are prone to accumulate random mutations and rapidly diverge from each other. This ultimately leads to large arrays of diversified satellite composition, as they can vary in sequence length, nucleotide composition, genomic position and chromosome distribution (Ugarković and Plohl, 2002; Garrido-Ramos, 2017). Repeats within the same satDNA family, however, show lower divergence rates as they do not evolve independently, but via what is called "concerted evolution" (Dover, 1982; Elder and Turner, 1995; Thakur et al., 2021). These satDNAs are submitted to the process of molecular drive, which causes sequence homogenization for species-specific mutations and results in repetitions evolving in concert with each other (Dover, 1982; Dover, 1986; Elder and Turner, 1995; Ugarković and Plohl, 2002; Thakur et al., 2021). The rates at which these sequences expanded, homogenized and eventually fixed in the genome vary for each satDNA family. These levels of variation depend on a number of factors, such as mutation rate, array size and structure, chromosomal structure and recombination rates (Ohta and Dover, 1984; Elder and Turner, 1995; Plohl et al., 2012). SatDNA evolution encompasses the duality of combining stable homogeneous arrays fixed in a genome and the high dynamism of rapidly replaceable sequences (i.e., turnover) (Ugarković and Plohl, 2002). These features still place satDNA evolutionary perspective under necessary evaluation. Nonetheless, what is certainly known so far is how the evolutionary mechanisms governing these sequences are different in comparison to other genomic elements (Plohl et al., 2012; Thakur et al., 2021).

Heterochromatin is a poorly understood genomic component, perhaps given the specific nature of its repetitive content (composed mostly of satDNAs) (Charlesworth et al., 1994; Garrido-Ramos, 2017). Sex chromosomes are a good example of genomic entities that can experience heterochromatin expansions and contractions along their evolution, as they can be submitted to rapid diversification after colonization by repetitive sequences and they might gradually differentiate from their homologs, becoming heteromorphic sex chromosomes (Chalopin et al., 2015; Palacios-Gimenez et al., 2015; Wright et al., 2016; Yano et al., 2016; Sember et al., 2018; Charlesworth, 2021; Kratochvíl et al., 2021). The rapid spread of repetitive sequences in a heterogametic (Y or W) sex chromosome may occur during the initial phase of its existence and it may facilitate the expansion of regions with ceased recombination, which further helps XY or ZW counterparts to differentiate from each other (Charlesworth, 1991; Bergero and Charlesworth, 2009; Bachtrog et al., 2011; Schartl et al., 2016).

Teleost fishes compose the most speciose group of vertebrates, which present equally diversified spectrum of sex chromosome occurrences scattered throughout the taxa (Devlin and Nagahama, 2002; Volff, 2005; Godwin and Roberts, 2018; Sember et al., 2021). Teleosts display variable degrees of molecular differentiation in their sex chromosomes, even among closely related species; which contrasts with the more uniform sex systems found in birds and therian mammals (Ellegren, 2011; Schartl et al., 2016; Charlesworth, 2019). Having said that, not much is known about what molecular mechanisms underlie this large compendium of sex chromosome occurrences or the evolutionary dynamics and genomic features of teleosts as a whole.

Megaleporinus is a Neotropical fish genus with a conserved karyotype of 2n = 54 chromosomes, female heterogamety (ZW) and, in *M. elongatus* case, a hypothetical $Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$ multiple sex chromosome system (Parise-Maltempi et al., 2007; Parise-Maltempi et al., 2013). The heteromorphic sex chromosomes for both systems (W and W_1) are highly heterochromatic and they constitute the largest elements in the female karyotype, which makes the understanding of *Megaleporinus* species genomic traits and sex chromosome systems a coveted approach (Nakayama et al., 1994; Parise-Maltempi et al., 2007; Hashimoto et al., 2009; Ramirez et al., 2017a; Ramirez et al., 2017b; Dulz et al., 2020). Heterochromatin content of the genus has sparked interest in previous works that specifically targeted repetitive DNA occurrence in the heteromorphic sex chromosomes (da Silva et al., 2012; Marreta et al., 2012; de Borba et al., 2013; Parise-Maltempi et al., 2013; Poltronieri et al., 2014). However, only recently was it possible to effectively quantify satDNA content and trace their evolutionary features within *Megaleporinus* (Utsunomia et al., 2019; Crepaldi and Parise-Maltempi, 2020).

Thus, in the present study we used low-coverage genomic DNA data yielded from Illumina paired-end sequencing to assess and further investigate the genomic organization of the highly diversified satellitome in *Megaleporinus elongatus*, firstly presented by Crepaldi and Parise-Maltempi (2020). Cytogenomic and haplotype analyses were also used to further investigate possible sex-specific patterns of these satDNAs and their evolutionary pathways. We complemented our survey with an in-depth analysis of W_1-located satDNAs in both *M. elongatus* and its sister species *M. macrocephalus*, focusing specifically on the two most abundant and quantitatively relevant elements of *M. elongatus* satellitome. By complementing previous satDNA and FISH results (Crepaldi and Parise-Maltempi, 2020) with new additions to satellitome data and haplotype networks, we managed to trace a possible evolutionary pathway for these satDNAs in *M. elongatus* and understand how they contributed to the recent differentiation of the heteromorphic sex chromosomes and their possible role in the multiple sex chromosome differentiation. Altogether, our study aimed to integrate thorough genomic analyses to previously published data for the *Megaleporinus* genus and to provide new information regarding the evolution of repetitive genomic composition in the group, as well as satDNA evolutionary differences between the sexes and closely related species.

MATERIALS AND METHODS

Species Sampling and Chromosome Preparation

Chromosomal preparations and tissue samples from three males and three females from *Megaleporinus elongatus* (Anostomidae)

were analyzed. All samples were already available at the Animal Cytogenetics Laboratory in UNESP Rio Claro, Brazil from previous studies (da Silva et al., 2012; Marreta et al., 2012; Parise-Maltempi et al., 2013; Crepaldi and Parise-Maltempi, 2020). All procedures for sampling, material handling and analysis were authorized and approved by the Animal Ethics Committee (Comitê de Ética no Uso de Animais (CEUA) (protocol number 3524, approval code 09/2017), by the Brazilian Institute of Environment and Renewable Natural Resources (Instituto Brasileiro do Meio Ambiente e Recursos Naturais Renováveis - IBAMA) (19833-1) and the Brazilian College of Animal Experimentation (Colégio Brasileiro de Experimentação Animal - 016/04-CEEA). Mitotic chromosomes were obtained from anterior kidney cell suspensions, according to Foresti and Toledo (1981).

Genome Sequencing and Computational Satellitome Analysis

In the present study, previously sequenced libraries from each sex of *M. elongatus* using Illumina® Hiseq™ 2000 platform (female) and Illumina® Hiseq™ 4000 platform (male) (Crepaldi and Parise-Maltempi, 2020) were used, which provided 1.9 Gb of sequence data and yielded 19,289,312 paired-end trimmed reads for the female, and 1.8 Gb of data and 17,837,098 paired-end trimmed reads for the male library.

The sequenced libraries were used for comparative analysis regarding their satDNA content. We focused both on the species' satellitome as a whole and on the differences between the sexes, in search for sequences that could be more representative in the female and probably enriched in the heteromorphic W_1 sex chromosome. To perform a high-throughput analysis, the satMiner bioinformatic protocol for satDNA prospection in both libraries was used (Ruiz-Ruano et al., 2016) available at GitHub (https://github.com/fjruizruano/satminer). The satMiner protocol uses several rounds of clustering in RepeatExplorer (Novák et al., 2013) to identify and extract satDNA sequences, and each round includes filtering out reads matching previously assembled contigs with deconseq 0.4.3 (Schmieder and Edwards, 2011), in order to identify and extract as many sequences as possible, even with low abundance in the genome. We then started with a library sampling of 200,000 reads, incrementing this number by two in each consequent round of RepeatExplorer clustering. For each round, we selected clusters with spherical shaped graphs for putative satDNA. Each selected cluster was manually analyzed for their internal contig structure and tandem repetitions were investigated using the dotplot tool implemented in Geneious v4.8 (Drummond et al., 2015) and Tandem Repeat Finder (TRF) (Benson, 1999).

The clustering and filtering steps were repeated 13 times for the female library and 9 times for the male one, adding new filtered reads in each iteration until we could no longer detect new satDNAs in neither. A parallel homology search was performed in both male and female rounds using previously detected satDNA consensus sequences by Crepaldi and Parise-Maltempi (2020) as a custom library to match ones that might have previously been isolated and physically mapped.

After satDNA mining, all-against-all alignments with RepeatMasker (Smit et al., 2013) were performed to search for homologous satDNAs, and by comparing all monomers from all clusters we were able to classify them into superfamilies, families or variants (Ruiz-Ruano et al., 2016). Non-redundant consensus library was used for each satDNA family to check for possible similarities with published sequences deposited in Genbank and Repbase employing BLAST (http://www.ncbi.nlm.gov/Blast/) and Censor (http://www.girinst.org/censor) searches.

All satDNA families were numbered in order of decreasing abundance in the female genome and assigned following the nomenclature proposed by Ruiz-Ruano et al. (2016). All satDNAs that were previously isolated by Crepaldi and Parise-Maltempi (2020) were properly renamed also following this criterion. Sequences are deposited in GenBank under accession numbers MZ546645–MZ546784.

RepeatMasker with rmblast engine was used to determine abundance and average nucleotide divergence (Kimura-2-parameter, K2P) for each variant in both sexes. We estimated the genomic abundance for every satDNA in the male and female libraries as the number of nucleotides aligned to the reference consensus divided by the library size (in bp). With this data we generated repeat landscapes for the relative abundance (Y-axis) at 1% intervals of K2P distance from the consensus (X-axis), using the script calcDivergencFromAlign.pl (from RepeatMasker utils). A subtractive landscape was subsequently generated to evaluate which satDNA families differ between both libraries, in turn providing the first indications of which satDNA are more prominent in one sex in comparison to the other.

Physical Mapping via FISH in Male Metaphases

We selected the same previously isolated and amplified 52 satDNAs in the female (Crepaldi and Parise-Maltempi, 2020) and amplified them via PCR in *M. elongatus* males following the protocol described by the authors. The PCR products were confirmed via Sanger sequencing.

The sequences of each satDNA obtained through PCR were labeled by nick translation with digoxigenin-11-dUTP (Roche®) or biotin-14-dATP (Invitrogen®). Fluorescence *in situ* hybridization (FISH) experiments were performed following the method described by Pinkel et al. (1986), with small adjustments described by Crepaldi and Parise-Maltempi (2020). The resulting slides were visualized under an Olympus® BX51 fluorescence microscope, with a digital camera Olympus® DP71 attached, and the images were captured using the DP Controller camera software. For each slide, a minimum of 20 metaphases were analyzed and photographed to confirm the FISH results.

Sex-Biased Ratio

The different enrichment of all satDNAs across the sexes was determined by generating a female to male ratio as we calculated the quotient between the abundance values of each satDNA family. This data complemented the subtractive landscape by providing more between-sexes differences, as satDNA families

with Female/Male (F/M) ratio higher than one were considered more abundant in females (as the threshold to determine it as more prevalent in this sex).

RepeatProfiler pipeline (Negm et al., 2021) was applied to generate comparative variant-enhanced profiles of selected satDNAs, which provide a summary of variant sites that are relative to the consensus sequences and may uncover sex-specific signatures of sequence variants and possible point mutations in satDNAs of interest. The 22 satDNA families that are most enriched in each sex considering the F/M ratio were selected for profiling. RepeatProfiler relies on Bowtie2 (Langmead and Salzberg, 2012) for mapping the reads to the consensus monomers of the selected sequences and the pipeline generated a PDF file for each selected satellite with the variant-enhanced profiles for both sexes as an output. We applied Bowtie2 default parameters for RepeatProfiler.

After analyzing the resulting profiles, we generated individual landscapes for each selected female-biased satDNAs to confirm different amplification and divergence of their copies in male and female genomes.

Retrieving satDNA Monomers From Raw Reads of *M. elongatus* and *M. macrocephalus*

Comparative cytogenetic analysis of the two most abundant satDNAs (MelSat01-36 and MelSat02-26) revealed particular characteristics that prompted us to investigate them further, such as differences in clustered patterns in the W chromosome in *M. macrocephalus* and W_1 in *M. elongatus* and occurrence in autosomal pairs in both males and females. MelSat02-26 is specifically interesting given it is a fragment of Le*Spe*I, a repetitive marker used for the *M. elongatus* hypothetical multiple sex chromosome system (Parise-Maltempi et al., 2007; Crepaldi and Parise-Maltempi, 2020).

Monomers from Illumina reads representing MelSat01-36 and MelSat02-26 were extracted from both sexes of *M. elongatus* and *M. macrocephalus* libraries. Thus, Minimum Spanning Trees (MSTs) of these satDNA families were generated to trace the diversification patterns of copies between sexes and species. First, we mapped reads from raw Illumina libraries from the two selected satDNA, of both sexes of *M. elongatus* and *M. macrocephalus*, with a custom script (https://github.com/fjruizruano/satminer/blob/master/mapping_blat_gs.py).

Libraries of *M. macrocephalus* were retrieved from Sequence Read Archive (SRA) under the accession numbers SRR7263033 and SRR7263034. Mapped reads were then extracted from SAM files and aligned separately using MUSCLE (Edgar, 2004) with default parameters. The resulting alignment files of each satDNA were used as input in PHYLOViZ 2.0 (Nascimento et al., 2017) to construct MSTs on the bases of pairwise differences. RepeatMasker was used to determine abundance and average nucleotide divergence (K2P) for MelSat01-36 and MelSat02-26 in both sexes of *M. macrocephalus*. Then, the genomic abundance of these two satDNAs in reference to this species library sizes (in bp) was calculated, and this data was used to generate comparative landscapes between *M. elongatus* and *M. macrocephalus*.

RESULTS

General Satellitome Analysis in *M. elongatus*

A total of 140 different satDNA families (308 variants) were uncovered for *M. elongatus* as a whole, with a predominance of short repeat unit lengths (RUL) ranging from 11 to 245 bp (average of 43 bp) for both sexes. A + T content of consensus sequences varied between 30.8 and 80.4% (60% on average), which indicated a slight tendency towards A + T rich content for the majority of satDNAs (**Supplementary Table S1**). The homology analysis of repeat units revealed only one satDNA superfamily (SF1) comprising MelSat09-60 and MelSat113-60 (76.7%) (**Supplementary Figure S1**) and present in both sexes with relatively similar abundances. Both sequences share the same RUL (60 bp) and have high divergence values in both sexes (**Supplementary Table S1**).

SatDNA genome proportion ranged from 0.0001 to 0.484%, with the three most abundant satDNAs (MelSat01-36, MelSat02-26 and MelSat03-177) showing an abundance higher than 0.3%. SatDNA abundance in the present work regards the values determined by RepeatMasker that applied a substantial amount of reads in comparison to the pool of randomly selected reads analysed previously (and solely via RepeatExplorer output) by Crepaldi and Parise-Maltempi (2020). MelSat01-36 and MelSat02-26 remained the two most abundant families as previously described (Crepaldi and Parise-Maltempi, 2020) (0.484 and 0.413% respectively), but MelSat03-177 (previously the family MelSat07-177 located in the centromeric area) ranked third in abundance after this thorough analysis, representing 0.338% of the genome.

Therefore, the satellitome of *M. elongatus* consists primarily of the top three most prevalent satDNA, which that comprise almost 25% of the whole compendium of satellites, in addition to 10 families with abundance between 0.1 and 0.3% and a remainder of rare or very low abundant sequences (**Supplementary Table S1**). Overall divergence values were relatively variable for the species as a whole, ranging from 2.85 to 33.02% (average divergence for the species was 10.73%).

Searches in GenBank resulted in 46 families from *M. elongatus* with positive results with deposited *M. macrocephalus* satDNA sequences (Utsunomia et al., 2019). BLAST results and subsequent alignments for MelSat03-177 showed high similarities with centromeric sequences from other Characiformes (Utsunomia et al., 2017, Utsunomia et al.2019). MelSat40-52 (previously named MelSat49-52 (Crepaldi and Parise-Maltempi, 2020) also showed positive results for other satellite sequences in Characiformes, a conserved sequence with active transcription in *Characidium gomesi* (dos Santos et al., 2021).

FISH analysis was performed in male metaphases for *M. elongatus* for the same satDNA previously mapped in the female (Crepaldi and Parise-Maltempi, 2020). From the 52

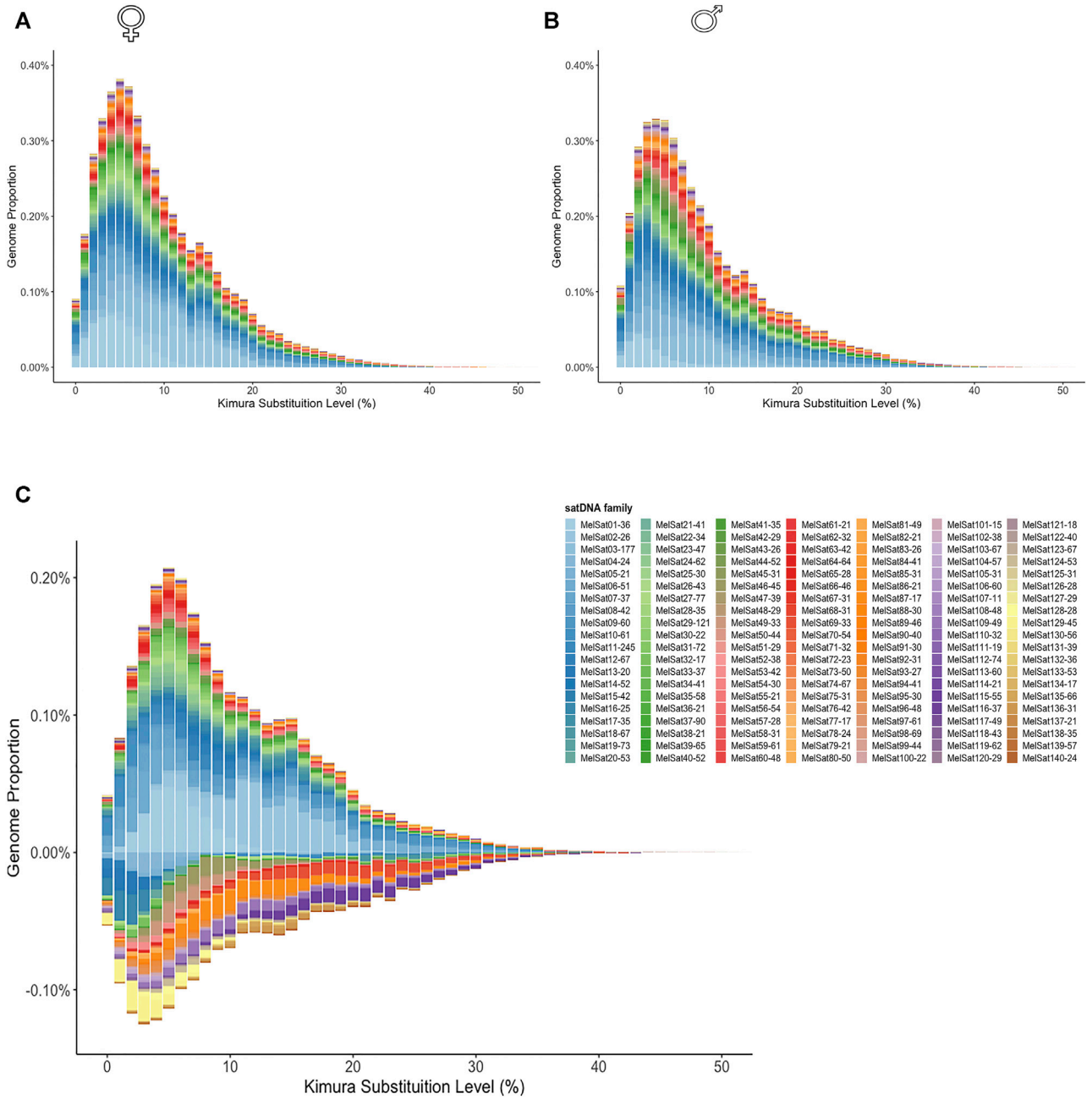

FIGURE 1 | Repeat landscapes (abundance vs. divergence) for satDNAs identified in female **(A)** and male **(B)** *M. elongatus*. The graphs show, for each color-coded element, the sequence divergence according to Kimura distance (x-axis) in relation to their copies in the genome (y-axis). Copies clustered to the left (lower divergence) potentially correspond to recent copies occurring in the genome. A subtractive landscape **(C)** was obtained by subtracting male library (negative values) from the female one (positive values). Notice the diversity of satDNA families also present in the male genome in the subtractive landscape.

satDNAs positively amplified via PCR, only 5 (MelSart02-26, MelSat22-34, MelSat29-121, MelSat44-52 and MelSat52-38) successfully hybridized in this sex (**Supplementary Figure S2**), in one autosomal pair each and, for MelSat02-26, in Z_2Z_2. Each satDNA presented a single band in the chromosomes, and positioned either in the telomeric or in the centromeric regions. No satDNA was mapped exclusively in the male, as all five presented FISH bands in female chromosomes (Crepaldi and Parise-Maltempi, 2020).

Satellitome Differences Between the Sexes

The total satDNA composition corresponded to 4.83 and 4.23% of the female and male genomes, respectively. Average satDNA family divergence was lower in the female (10.38%) than in the

male genome (11.07%), and some satDNAs had quite higher divergence values for the male in comparison to the female, such as MelSat66-46 (33.02 and 7.38%, for male and female respectively) and MelSat64-64 (22.44 and 6.34%). It was later found these satDNAs are highly female-biased according to the F/M ratio (**Supplementary Table S1**) and with insignificant abundance values for the male library, which sparked our interest to investigate further these differences, described hereinafter.

We generated individual repeat landscapes for female and male *M. elongatus* (**Figure 1A** and **Figure 1B**, respectively) and a subtractive landscape (**Figure 1C**). The subtractive landscape revealed higher proportions of several abundant satDNA families in the female library, such as MelSat01-36 and MelSat02-26, which are located in the W1 sex chromosome (Crepaldi and Parise-Maltempi, 2020). Interestingly, we could also confirm that the male library presents a higher variety of male-biased satDNAs in comparison to the female (**Figure 1C**), which are located in the W_1 sex chromosome (Crepaldi and Parise-Maltempi, 2020). Interestingly, also through the subtractive repeat landscape we could confirm that the male library presents a higher variety of male-biased satDNAs in comparison to the female (**Figure 1C**).

The computational analysis revealed that, except for MelSat131-39, which is present in the female genome only, all remaining 139 satDNA families are shared between the sexes. From these, however, 124 are differently enriched across sexes as 45 satDNAs had a F/M ratio higher than 1, suggesting an enrichment in the female library, while 79 were deemed male-biased (F/M ratio lower than 1). Despite having less overall female-biased satDNAs in numbers, several families had ratios ranging from 13 up to 347 in the female, denoting an expressive enrichment in this sex compared to the male library (**Supplementary Table S2**). In short, the female library has the most abundant satDNAs, enriched in the W_1 (Crepaldi and Parise-Maltempi, 2020); however, the greater diversity of satDNAs is male-biased, comprised mostly of the very rare and less abundant families of the satellitome.

Variant Profiles for Sex-Biased satDNAs

RepeatProfiler pipeline was applied in order to evaluate the sequences of the most sex-biased satDNAs and how they might differ between the sexes. The output provided variant-enhanced profiles for each selected satDNA, summarizing sex-specific signatures for some sequences (**Supplementary Figures S3, S4**). All male-biased satDNAs (**Supplementary Figure S3**) have relatively similar divergent values for both sexes and, with few exceptions (such as MelSat140-24, MelSat137-21 and MelSat92-31), they did not present such apparent differences between sexes as the female-biased ones, showing some conservatism in comparison with male and female profiles. Female-biased satDNAs (**Supplementary Figure S4**), on the other hand, showed some reoccurring patterns: 1) most have discrepant divergent values between the sexes (with high divergent values for the male) and are either absent in males or have incomplete profiles for this sex, since the aligner failed to map high-divergent reads to consensus; 2) comparable satDNAs were prone to variation in male variants, with particular positions exhibiting almost fixed copies, since mutation profiles differed between the sexes when abundance values dropped; and 3) the only two satDNAs that presented relative homogeneity between the sexes are MelSat01-36 and MelSat02-26 (the two most abundant satDNAs in the genome).

We complemented the variant profiles results by generating individual repeat landscapes (**Supplementary Figure S5**) for the top most female-biased satDNAs, to check for differential amplification and divergence values between the sequences especially on satDNAs that are most likely specific to the female. The landscapes also evidenced the higher abundance in the female and contrasting divergences for these sequences in the male genome. Some satDNAs, such as MelSat57-28, MelSat112-74 and MelSat123-67 showed monomers with low divergences in the female and shared higher divergences between the sexes, indicating highly divergent copies comparing the genomes.

Tracing the Amplification of W-Localized satDNAs in *M. elongatus* and *M. macrocephalus*

MSTs for MelSat01-36 and MelSat02-26 were generated and complemented with comparative landscapes for these sequences individually in order to integrate their abundance and divergence values to the analysis. Different MSTs and line plots were obtained for each satDNA (**Figure 2**).

The topology for MelSat01-36 confirms the relatedness and conservatism of this satDNA, especially when considering the larger haplotype shared between both species and its co-localized pattern in FISH results (**Figure 2A**). Furthermore, due to its low divergence values in both species, it most likely had a common origin in both genomes and an equally recent amplification in the heteromorphic sex chromosomes in the females.

Some discrepancies, however, are notable for *M. elongatus*. Firstly, clusterisation is absent in the male (no FISH signals). In addition to the fact that it has the least abundance in the male of *M. elongatus* of all four analyzed individuals, it is probably gradually losing copies in its genome. This coincides with another *M. elongatus* conspicuous characteristic, perhaps the most noticeable: the female copies. MelSat01-36 is not only more abundantly represented in the female genome, but it is also presented in two separate clusters in the heteromorphic chromosome (**Figure 2A**).

For MelSat02-26, the resulting MSTs displayed a more divergent pattern and a different evolutionary scenario for this satDNA (**Figure 2B**) was observed. Firstly, no monomers were retrieved from the male *M. macrocephalus*, as its abundance was virtually non-existent in this individual. For the remaining copies, all three (female *M. macrocephalus,* male and female *M. elongatus*) presented similar divergence with rather higher values, which implies that MelSat02-26 comprised an older satDNA especially in comparison to MelSat01-36.

Both sexes of *M. elongatus* share autosomal clusters and its origins appear concomitant in male and female, especially

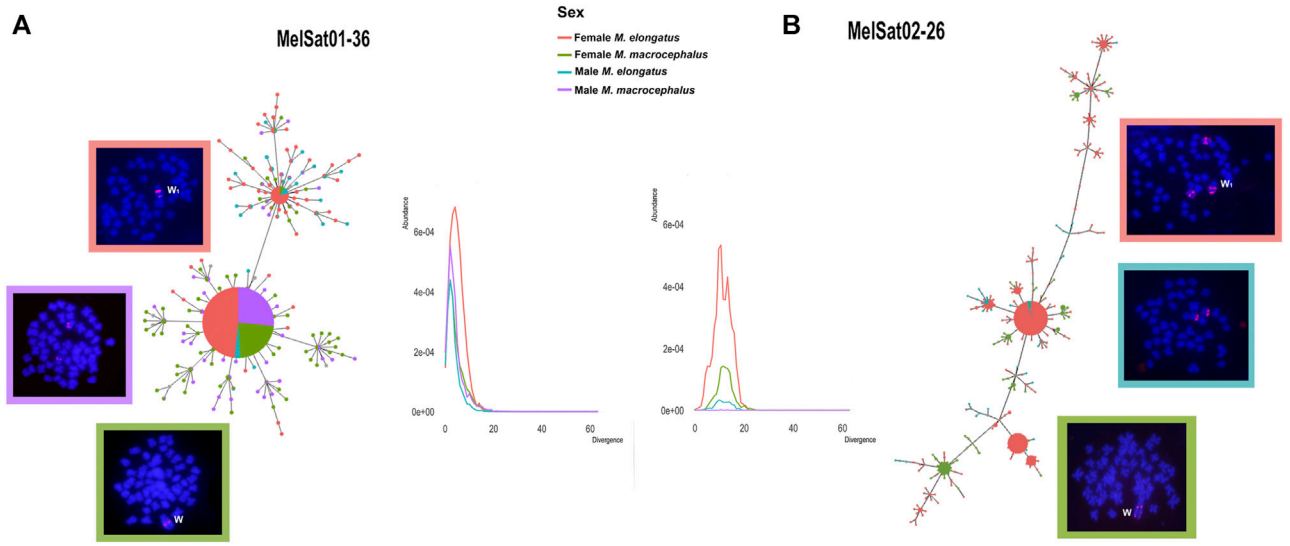

FIGURE 2 | MSTs showing the relationships between isolated monomers for MelSat01-36 **(A)** and MelSat02-26 **(B)** retrieved from *M. elongatus* and *M. macrocephalus* males and females. Respective FISH results and the heteromorphic sex chromosomes W and W_1 are indicated. Male FISH results for *M. elongatus* and *M. macrocephalus* are absent for MelSat01-36 and MelSat02-26 as no FISH signals were obtained, respectively. Abundance x divergence landscapes that shows the respective similar origins in both species are also indicated, but the substantial abundance in *M. elongatus* female for both satDNAs is apparent. Note the different divergences between the satDNAs, as MelSat02-26 has larger divergence values than MelSat01-36 and a rather dispersed organization in comparison to the latter.

considering it is clustered in the same autosomal pair and putative multiple sex chromosomes in both sexes (Parise-Maltempi et al., 2007; Crepaldi and Parise-Maltempi, 2020). There is a distinct spike in abundance for *M. elongatus* female in the line plot for this satDNA (**Figure 2B**), which, paired with the clustered pattern in FISH, shows the substantial amplification in the heteromorphic sex chromosome, which also occurred long since. It presents a different amplification trajectory between the species: although *M. macrocephalus* female also presents clusterisation in its W, it is not as considerable as in *M. elongatus* W_1 and it does not share any copies with *M. macrocephalus* male.

DISCUSSION

The present study provides relevant information regarding the satDNA content of *Megaleporinus elongatus* and its genomic features, adding notable details to the first glimpse of these sequences previously published by our group (Crepaldi and Parise-Maltempi, 2020). Using more thorough methods for satDNA prospection, we expanded the array of satellite features in the species, compared satellitomes between the sexes and uncovered the evolutionary pathway of the most abundant (and most quantitively relevant) satDNAs in the genome and their differences between species.

Firstly, some distinct characteristics are responsible for the *M. elongatus* satellitome as a whole, which showed small and highly diversified sequences, mostly recently amplified (low divergence) and with general low abundance. These characteristics were also found in other Characiformes, such as *Astyanax* (Silva et al., 2017), *Megaleporinus macrocephalus* (Utsunomia et al., 2019) and *Characidium gomesi* (Serrano-Freitas et al., 2020).

The disparity of a large satDNA diversity in *M. elongatus* (140 families) and such a scarce occurrence result in very rare sequences, organized in small arrays and with a dispersed organization in the genome. The massively amplified MelSat01-36 and MelSat02-26 contrasts with the remaining satellitome and indicate that satDNA diversification in *M. elongatus* increases as their clusters become smaller and more dispersed. Fast satDNA turnover may be another possibility, since almost all satDNA content is non-homologous in *M. elongatus*. *M. macrocephalus* presents 17 satDNA superfamilies (Utsunomia et al., 2019), but for *M. elongatus*, on the other hand, we recovered only one superfamily, which also endorses its highly diversified satDNA collection.

Also, most satDNAs were not detectable by physical mapping in neither sex, as presented in our present analyses and previous FISH results (Crepaldi and Parise-Maltempi, 2020). This confirms the majority of the satellitome for *M. elongatus* is arranged in small arrays below the detection threshold of FISH, corroborating other studies (Ruiz-Ruano et al., 2016; Bardella et al., 2020; Ferretti et al., 2020). SatDNA abundance is prone to rapidly change due to molecular mechanisms, such as dispersion and amplification (Garrido-Ramos, 2017), and can result in rapid repatterning as they expand or decrease their arrays (Plohl et al., 2008; Palacios-Gimenez et al., 2020; Sproul et al., 2020; Thakur et al., 2021) and this certainly is the case for *M. elongatus* as well.

This is a rather peculiar satellitome, especially considering our sex-biased analysis. Given the extraordinary heterochromatin content present in the female (owing to the heteromorphic W_1

sex chromosome) (Galetti et al., 1995), it was initially assumed by us that the largest set of satDNAs would be female-biased as previously observed in *M. macrocephalus* (Utsunomia et al., 2019); however, not many satDNAs (in terms of number of different families) accounted for female-biased or W_1-specific sequences in *M. elongatus*.

In what regards the male genome, satDNAs are equally abundant in qualitative terms; however, they are most likely becoming gradually lost and/or constrained instead of actually accumulating. Their low copies in the genome increase the possibility to escape homogenization mechanisms (Elder and Turner, 1995; Ugarković and Plohl, 2002; Plohl et al., 2012) as we noticed different homogenization patterns between the sexes. Repeats in larger arrays (and clustered in the W_1 chromosome) had higher homogenization rates, and low abundance sequences (especially the male-biased ones) presented a more divergent pattern. It is possible to assume some non-exclusive explanations for this scenario. Firstly, more heterogeneity can be expected among repeating units if the mutation rate is high, relative to the rate in which a variant spreads through an array (Lorite et al., 2004; Ruiz-Ruano et al., 2019). Also, the heteromorphic W_1 chromosome has a tendency to accumulate satDNAs already present in the female genome (Molina and Galetti, 2007; Crepaldi and Parise-Maltempi, 2020), which results in a quick, massive amplification in this sex in comparison to the poor clusterisation in the male counterpart (cf. Lower et al., 2018; Vondrak et al., 2020). In conclusion, satellitome diversity does not seem to be the main factor leading to W_1 heterochromatic expansion in *M. elongatus*, instead the high amplification and homogenization rates of few particularly abundant satDNAs in the female genome effectively contributed to heterogametic sex chromosome differentiation.

Our second approach regarded a more evolutionary focus in the sex chromosome system and its satDNA content in *Megaleporinus elongatus,* as sex chromosome dynamics analysed through sex-linked repetitive DNA profiles has been shown to be effective (Traut and Winking, 2001; Schemberger et al., 2011; Cioffi et al., 2012; Terencio et al., 2012; Yano et al., 2016; Yano et al., 2017; Sember et al., 2018; Deakin et al., 2019; Schemberger et al., 2019). The heteromorphic sex chromosomes (W and W_1) in *Megaleporinus* have already been deemed derived from a common ancestral chromosomic pair and differentiated from the Z chromosomes via massive heterochromatinization (Galetti and Foresti, 1986; Marreta et al., 2012; Parise-Maltempi et al., 2013; de Barros et al., 2018). However, variation in repetitive content within the genus could be caused by the expansion and contraction of these sequences (Utsunomia et al., 2019; Crepaldi and Parise-Maltempi, 2020).

Regardless of the library size of satDNAs, usually one or very few satDNA families are the most predominant in each species (Ruiz-Ruano et al., 2016; Silva et al., 2017; Bardella et al., 2020; da Silva et al., 2020) and the best candidates were none other but the two most abundant satDNA in *M. elongatus*, MelSat01-36 and MelSat02-26. Both satDNAs have concomitant origins in both sexes of *M. elongatus* and *M. macrocephalus,* but they have different evolutionary pathways in each species and concerted evolution might be acting separately for each of them.

MelSat01-36 is shared by W and W_1, but it is clustered only in males of one species (with prominent FISH signals being present in *M. macrocephalus* but absent in *M. elongatus*). Its two distinct clusters in female *M. elongatus* prompted us to combine all of our data on this satDNA and attempt to trace its relatedness between the species. All individuals presented similarly low divergence values for this satDNA, indicating a recent amplification for both species. Although identical sequences for MelSat01-36 are shared between the species, copy number can vary dramatically even among related organisms, without necessarily varying in their nucleotide sequences (Ugarković and Plohl, 2002; Louzada et al., 2020). Given our MST for this satDNA (**Figure 2**), ancestral species for *M. elongatus* and *M. macrocephalus* most likely presented the same conformation as *M. macrocephalus,* with one cluster in the ancestral W sex chromosome in the female and another cluster in a male autosomal chromosome. For *M. elongatus*, the male has lower copy numbers and a dispersed organization, since it did not present any FISH signals. This loss of autosomal clusters might have contributed to the homogenization in W_1 in *M. elongatus,* and generated the enriched and duplicated regions in this chromosome, fixating in the female genome as described for many other organisms (Dalíková et al., 2017; Palacios-Gimenez et al., 2017; da Silva et al., 2020; Ferretti et al., 2020; Rovatsos et al., 2021).

MelSat02-26, on the other hand, is an older satDNA, with higher divergence values for both species and it is practically absent in male *M. macrocephalus*. Despite shared between the heteromorphic W and W_1, it has a much less conserved pattern (**Figure 2**) in comparison to MelSat01-36, with dispersed clusters and almost none shared haplotypes between the females. An ancestral library for this satDNA cluster was most likely present in both the autosomes and the heteromorphic chromosome, much like MelSat01-36, however remaining highly amplified on both sexes of *M. elongatus.* It most likely had more time to diverge between the species and fixate in *M. elongatus* as well-defined clusters, as it is also co-localized in the putative multiple sex chromosomes (Z_2 and W_2). What previously seemed like having different accumulation in terms of time for each chromosome (Crepaldi and Parise-Maltempi, 2020), now, with more intimate evolutionary details, we can see that MelSat02-26 seems to have emerged equally in all *M. elongatus* chromosomes; along with its exclusive clusterisation pattern in the female Z_2 and W_2, shared with no other autosomal satDNA in the species and very particular to this sex.

The dynamics of the most abundant satDNAs in *M. elongatus* demonstrates not only their intimate evolution with the highly differentiated heteromorphic W_1 sex chromosome, but also confirms that not only general molecular satDNA evolution is at play; particular characteristics of the species also concomitantly generate diversified satellitomes (Ugarković and Plohl, 2002; Lorite et al., 2004, Lorite et al., 2017; Richard et al., 2008; Palacios-Gimenez et al., 2017; Cabral-de-Mello et al., 2021). The ZW-based multiple sex chromosome system in *M. elongatus* shares a conserved heteromorphic W chromosome with the remaining *Megaleporinus*; but as seen by the two most prevalent satDNAs in the species their interspecific differences show distinct paces for W/W_1 differentiation.

SatDNA evolution seems to be not only recent but very fast-paced in the genus, with high sequence turnover rates, and this might contribute to an equally fast and independent differentiation process in their young sex chromosome systems. A clear outcome of this is the highly amplified MelSat02-26 in Z_2 and W_2 in *M. elongatus* female, an interesting satDNA on its own as it partially represents a bigger repetitive sequence and molecular marker (Le*Spe*I) for the putative multiple chromosomes (Parise-Maltempi et al., 2007; Crepaldi and Parise-Maltempi, 2020). While MelSat02-26 is underrepresented in the female Z_1 chromosome, the unusual female-specific clustered pattern shared by W1 and the novel elements (W_2 and Z_2) might indicate its spread via ectopic recombination, which has most likely helped this satDNA expand into large, homogeneous arrays (Ugarković and Plohl, 2002; Lower et al., 2018).

The rare $Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$ system of *M. elongatus* is so far only found in one other fish species (*Ancistrus dolichopterus*) (de Oliveira et al., 2008; Favarato et al., 2016), and remains a puzzling occurrence even among multiple sex chromosome systems as whole (Sember et al., 2021). Analyzing other *Megaleporinus* as well as the sister genus *Leporinus* (comprising species without heteromorphic sex chromosomes) will eventually broaden our understanding of satDNA diversity and sex chromosome evolution as a whole in this group.

In the present study, we used cytogenomic approaches for satDNA analysis and visualization of their male and female heterogeneity. The results corroborate the recent spike of highly clustered repetitive content in the female genome and showed it has very distinctive characteristics even in comparison to closely-related species. Also, the recent burst of repetitive satDNA in the multiple sex chromosome system is still an ongoing process under the molecular mechanisms for satDNA evolution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories and accession number(s) can be found in the article.

ETHICS STATEMENT

The animal study was reviewed and approved by Sao Paulo State University Animal Ethics Committee (protocol number 3524, approval code 09/2017).

AUTHOR CONTRIBUTIONS

CC contributed to the work's conception, data acquisition, analysis and interpretation, and wrote the manuscript. EM contributed to bioinformatic data acquisition, analysis and interpretation. CC and EG conducted cytogenetic analysis. DM contributed to bioinformatic data resources. PPP-M conceptualized, funded and coordinated the research. All authors revised the manuscript.

FUNDING

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP #2017/00195_7 and FAPESP #2018/12906-8).

ACKNOWLEDGMENTS

The authors would like to thank Dr. Diogo Cavalcanti Cabral-de-Mello and his laboratory students for the valuable suggestions and constructive comments, and critical reading of an earlier version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.728670/full#supplementary-material

REFERENCES

Bachtrog, D., Kirkpatrick, M., Mank, J. E., McDaniel, S. F., Pires, J. C., Rice, W., et al. (2011). Are all sex chromosomes created equal. *Trends Genet.* 27 (9), 350–357. doi:10.1016/j.tig.2011.05.005

Bardella, V. B., Milani, D., and Cabral-de-Mello, D. C. (2020). Analysis of *Holhymenia Histrio* genome provides insight into the satDNA evolution in an insect with holocentric chromosomes. *Chromosome Res.* 28 (3), 369–380. doi:10.1007/s10577-020-09642-1

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

Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* 24 (2), 94–102. doi:10.1016/j.tree.2008.09.010

Cabral-de-Mello, D. C., Zrzavá, M., Kubíčková, S., Rendón, P., and Marec, F. (2021). The role of satellite DNAs in genome architecture and sex chromosome evolution in Crambidae moths. *Front. Genet.* 12, 661417. doi:10.3389/fgene.2021.661417

Caetano de Barros, L., Piscor, D., Parise-Maltempi, P. P., and Feldberg, E. (2018). Differentiation and Evolution of the W Chromosome in the Fish Species of *Megaleporinus* (Characiformes, Anostomidae). *Sex. Dev.* 12 (4), 204–209. doi:10.1159/000489693

Chalopin, D., Volff, J.-N., Galiana, D., Anderson, J. L., and Schartl, M. (2015). Transposable elements and early evolution of sex chromosomes in fish. *Chromosome Res.* 23 (3), 545–560. doi:10.1007/s10577-015-9490-8

Charlesworth, B., Sniegowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371 (6494), 215–220. doi:10.1038/371215a0

Charlesworth, B. (1991). The evolution of sex chromosomes. *Science* 251 (4997), 1030–1033. doi:10.1126/science.1998119

Charlesworth, D. (2021). When and how do sex-linked regions become sex chromosomes. *Evolution* 75 (3), 569–581. doi:10.1111/evo.14196

Charlesworth, D. (2019). Young sex chromosomes in plants and animals. *New Phytol.* 224 (3), 1095–1107. doi:10.1111/nph.16002

Cioffi, M. B., Moreira-Filho, O., Almeida-Toledo, L. F., and Bertollo, L. A. C. (2012). The contrasting role of heterochromatin in the differentiation of sex

chromosomes: an overview from Neotropical fishes. *J. Fish Biol.* 80 (6), 2125–2139. doi:10.1111/j.1095-8649.2012.03272.x

Crepaldi, C., and Parise-Maltempi, P. P. (2020). Heteromorphic sex chromosomes and their DNA content in fish: an insight through satellite DNA accumulation in *Megaleporinus elongatus*. *Cytogenet. Genome Res.* 160 (1), 38–46. doi:10.1159/000506265

da Silva, E. L., de Borba, R. S., and Parise-Maltempi, P. P. (2012). Chromosome mapping of repetitive sequences in Anostomidae species: implications for genomic and sex chromosome evolution. *Mol. Cytogenet.* 5 (1), 45. doi:10.1186/1755-8166-5-45

da Silva, M. J., Fogarin Destro, R., Gazoni, T., Narimatsu, H., Pereira dos Santos, P. S., Haddad, C. F. B., et al. (2020). Great abundance of satellite DNA in *Proceratophrys* (Anura, Odontophrynidae) revealed by genome sequencing. *Cytogenet. Genome Res.* 160 (3), 141–147. doi:10.1159/000506531

Dalíková, M., Zrzavá, M., Kubíčková, S., and Marec, F. (2017). W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). *Chromosome Res.* 25 (3), 241–252. doi:10.1007/s10577-017-9558-8

de Oliveira, R. R., Feldberg, E., Dos Anjos, M. B., and Zuanon, J. (2008). Occurrence of multiple sexual chromosomes (XX/XY_1Y_2 and $Z_1Z_1Z_2Z_2/Z_1Z_2W_1W_2$) in catfishes of the genus *Ancistrus* (Siluriformes: Loricariidae) from the Amazon basin. *Genetica* 134 (2), 243–249. doi:10.1007/s10709-007-9231-9

Deakin, J. E., Potter, S., O'Neill, R., Ruiz-Herrera, A., Cioffi, M. B., Eldridge, M. D. B., et al. (2019). Chromosomics: bridging the gap between genomes and chromosomes. *Genes* 10, 627. doi:10.3390/genes10080627

Devlin, R. H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208 (3), 191–364. doi:10.1016/S0044-8486(02)00057-1

dos Santos, R. Z., Calegari, R. M., Silva, D. M. Z. d. A., Ruiz-Ruano, F. J., Melo, S., Oliveira, C., et al. (2021). A long-term conserved satellite DNA that remains unexpanded in several genomes of Characiformes fish is actively transcribed. *Genome Biol. Evol.* 13 (2). doi:10.1093/gbe/evab002

Dover, G. A. (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet.* 2, 159–165. doi:10.1016/0168-9525(86)90211-8

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299 (5879), 111–117. doi:10.1038/299111a0

Drummond, A., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., et al. (2015). *Geneious v4. 8. 2009*. [Online]. Available: http://www.geneious.com (Accessed 10, 2020).

Dulz, T. A., Azambuja, M., Nascimento, V. D., Lorscheider, C. A., Noleto, R. B., Moreira-Filho, O., et al. (2020). Karyotypic diversification in two *Megaleporinus* species (Characiformes, Anostomidae) inferred from *in situ* localization of repetitive DNA sequences. *Zebrafish* 17 (5), 333–341. doi:10.1089/zeb.2020.1918

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5 (1), 113. doi:10.1186/1471-2105-5-113

Elder, J. F., and Turner, B. J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 70 (3), 297–320. doi:10.1086/419073

Ellegren, H. (2011). Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat. Rev. Genet.* 12 (3), 157–166. doi:10.1038/nrg2948

Favarato, R. M., da Silva, M., de Oliveira, R. R., Artoni, R. F., Feldberg, E., and Matoso, D. A. (2016). Cytogenetic diversity and the evolutionary dynamics of rDNA genes and telomeric sequences in the *Ancistrus* genus (Loricariidae: Ancistrini). *Zebrafish* 13 (2), 103–111. doi:10.1089/zeb.2015.1140

Ferretti, A. B. S. M., Milani, D., Palacios-Gimenez, O. M., Ruiz-Ruano, F. J., and Cabral-de-Mello, D. C. (2020). High dynamism for neo-sex chromosomes: satellite DNAs reveal complex evolution in a grasshopper. *Heredity* 125 (3), 124–137. doi:10.1038/s41437-020-0327-7

Foresti, F., Toledo, A., and Toledo F°, S. A. (1981). Polymorphic nature of nucleolus organizer regions in fishes. *Cytogenet. Cel Genet* 31 (3), 137–144. doi:10.1159/000131639

Galetti Jr., P. M., Jr, Lima, N. R. W., and Venere, P. C. (1995). A monophyletic ZW sex chromosome system in *Leporinus* (Anostomidae, Characiformes). *Cytologia* 60 (4), 375–382. doi:10.1508/cytologia.60.375

Galetti, Jr., P. M., Jr, and Foresti, F. (1986). Evolution of the ZZ/ZW system in *Leporinus* (Pisces, Anostomidae). *Cytogenet. Cel Genet* 43 (1-2), 43–46. doi:10.1159/000132296

Garrido-Ramos, M. (2017). Satellite DNA: an evolving topic. *Genes* 8, 230. doi:10.3390/genes8090230

Godwin, J., and Roberts, R. (2018). "Environmental and Genetic Sex Determining Mechanisms in Fishes," in *Transitions Between Sexual Systems: Understanding the Mechanisms of, and Pathways Between, Dioecy, Hermaphroditism and Other Sexual Systems*. Editor J. L. Leonard (Cham: Springer International Publishing), 311–344. doi:10.1007/978-3-319-94139-4_11

Hashimoto, D. T., Parise-Maltempi, P. P., Laudicina, A., Bortolozzi, J., Senhorini, J. A., Foresti, F., et al. (2009). Repetitive DNA probe linked to sex chromosomes in hybrids between Neotropical fish *Leporinus macrocephalus* and *Leporinus elongatus* (Characiformes, Anostomidae). *Cytogenet. Genome Res.* 124 (2), 151–157. doi:10.1159/000207523

Kratochvíl, L., Stöck, M., Rovatsos, M., Bullejos, M., Herpin, A., Jeffries, D. L., et al. (2021). Expanding the classical paradigm: what we have learnt from vertebrates about sex chromosome evolution. *Phil. Trans. R. Soc. B* 376 (1833), 20200097. doi:10.1098/rstb.2020.0097

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

Lorite, P., Carrillo, J. A., Tinaut, A., and Palomeque, T. (2004). Evolutionary dynamics of satellite DNA in species of the Genus *Formica* (Hymenoptera, Formicidae). *Gene* 332, 159–168. doi:10.1016/j.gene.2004.02.049

Lorite, P., Muñoz-López, M., Carrillo, J. A., Sanllorente, O., Vela, J., Mora, P., et al. (2017). Concerted evolution, a slow process for ant satellite DNA: study of the satellite DNA in the *Aphaenogaster* genus (Hymenoptera, Formicidae). *Org. Divers. Evol.* 17 (3), 595–606. doi:10.1007/s13127-017-0333-7

Louzada, S., Lopes, M., Ferreira, D., Adega, F., Escudeiro, A., Gama-Carvalho, M., et al. (2020). Decoding the Role of Satellite DNA in Genome Architecture and Plasticity-An Evolutionary and Clinical Affair. *Genes* 11, 72. doi:10.3390/genes11010072

Lower, S. S., McGurk, M. P., Clark, A. G., and Barbash, D. A. (2018). Satellite DNA evolution: old ideas, new approaches. *Curr. Opin. Genet. Dev.* 49, 70–78. doi:10.1016/j.gde.2018.03.003

Marreta, M. E., Faldoni, F. L. C., and Parise-Maltempi, P. P. (2012). Cytogenetic mapping of the W chromosome in the genus *Leporinus* (Teleostei, Anostomidae) using a highly repetitive DNA sequence. *J. Fish Biol.* 80 (3), 630–637. doi:10.1111/j.1095-8649.2011.03199.x

Molina, W. F., and Galetti, P. M. (2007). Early replication banding in *Leporinus* species (Osteichthyes, Characiformes) bearing differentiated sex chromosomes (ZW). *Genetica* 130 (2), 153–160. doi:10.1007/s10709-006-9002-z

Nakayama, I., Foresti, F., Tewari, R., Schartl, M., and Chourrout, D. (1994). Sex chromosome polymorphism and heterogametic males revealed by two cloned DNA probes in the ZW/ZZ fish *Leporinus elongatus*. *Chromosoma* 103 (1), 31–39. doi:10.1007/bf00364723

Nascimento, M., Sousa, A., Ramirez, M., Francisco, A. P., Carriço, J. A., and Vaz, C. (2017). PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics* 33 (1), 128–129. doi:10.1093/bioinformatics/btw582

Negm, S., Greenberg, A., Larracuente, A. M., and Sproul, J. S. (2021). RepeatProfiler: a pipeline for visualization and comparative analysis of repetitive DNA profiles. *Mol. Ecol. Resour.* 21 (3), 969–981. doi:10.1111/1755-0998.13305

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 (6), 792–793. doi:10.1093/bioinformatics/btt054

Ohta, T., and Dover, G. A. (1984). The cohesive population genetics of molecular drive. *Genetics* 108 (2), 501–521. doi:10.1093/genetics/108.2.501

Palacios-Gimenez, O. M., Dias, G. B., de Lima, L. G., Kuhn, G. C. e. S., Ramos, É., Martins, C., et al. (2017). High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis*. *Sci. Rep.* 7 (1), 6422. doi:10.1038/s41598-017-06822-8

Palacios-Gimenez, O. M., Marti, D. A., and Cabral-de-Mello, D. C. (2015). Neo-sex chromosomes of *Ronderosia bergi*: insight into the evolution of sex chromosomes in grasshoppers. *Chromosoma* 124 (3), 353–365. doi:10.1007/s00412-015-0505-1

Palacios-Gimenez, O. M., Milani, D., Song, H., Marti, D. A., López-León, M. D., Ruiz-Ruano, F. J., et al. (2020). Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. *Genome Biol. Evol.* 12 (3), 88–102. doi:10.1093/gbe/evaa018

Parise-Maltempi, P., da Silva, E., Rens, W., Dearden, F., O'Brien, P. C., Trifonov, V., et al. (2013). Comparative analysis of sex chromosomes in *Leporinus* species (Teleostei, Characiformes) using chromosome painting. *BMC Genet.* 14 (1), 60. doi:10.1186/1471-2156-14-60

Parise-Maltempi, P. P., Martins, C., Oliveira, C., and Foresti, F. (2007). Identification of a new repetitive element in the sex chromosomes of *Leporinus elongatus*

(Teleostei: Characiformes: Anostomidae): new insights into the sex chromosomes of *Leporinus*. *Cytogenet. Genome Res.* 116 (3), 218–223. doi:10.1159/000098190

Pinkel, D., Straume, T., and Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci.* 83 (9), 2934–2938. doi:10.1073/pnas.83.9.2934

Plohl, M., Luchetti, A., Meštrović, N., and Mantovani, B. (2008). Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409 (1-2), 72–82. doi:10.1016/j.gene.2007.11.013

Plohl, M., Meštrovic, N., and Mravinac, B. (2012). Satellite DNA evolution. *Genome Dyn.* 7, 126–152. doi:10.1159/000337122

Poltronieri, J., Marquioni, V., Bertollo, L. A. C., Kejnovsky, E., Molina, W. F., Liehr, T., et al. (2014). Comparative Chromosomal Mapping of Microsatellites in *Leporinus* Species (Characiformes, Anostomidae): Unequal Accumulation on the W Chromosomes. *Cytogenet. Genome Res.* 142 (1), 40–45. doi:10.1159/000355908

Ramirez, J. L., Birindelli, J. L., Carvalho, D. C., Affonso, P. R. A. M., Venere, P. C., Ortega, H., et al. (2017a). Revealing hidden diversity of the underestimated neotropical ichthyofauna: DNA barcoding in the recently described genus *Megaleporinus* (Characiformes: Anostomidae). *Front. Genet.* 8, 149. doi:10.3389/fgene.2017.00149

Ramirez, J. L., Birindelli, J. L. O., and Galetti, P. M., Jr. (2017b). A new genus of Anostomidae (Ostariophysi: Characiformes): Diversity, phylogeny and biogeography based on cytogenetic, molecular and morphological data. *Mol. Phylogenet. Evol.* 107, 308–323. doi:10.1016/j.ympev.2016.11.012

Richard, G.-F., Kerrest, A., and Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* 72 (4), 686–727. doi:10.1128/MMBR.00011-08

Rovatsos, M., Marchal, J. A., Giagia-Athanasopoulou, E., and Sánchez, A. (2021). Molecular composition of heterochromatin and its contribution to chromosome variation in the *Microtus thomasi/Microtus atticus* species complex. *Genes* 12, 807. doi:10.3390/genes12060807

Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., and Camacho, J. P. M. (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* 6, 28333. doi:10.1038/srep28333

Ruiz-Ruano, F. J., Navarro-Domínguez, B., Camacho, J. P. M., and Garrido-Ramos, M. A. (2019). Characterization of the satellitome in lower vascular plants: the case of the endangered fern *Vandenboschia speciosa*. *Ann. Bot.* 123 (4), 587–599. doi:10.1093/aob/mcy192

Schartl, M., Schmid, M., and Nanda, I. (2016). Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. *Chromosoma* 125 (3), 553–571. doi:10.1007/s00412-015-0569-y

Schemberger, M. O., Bellafronte, E., Nogaroto, V., Almeida, M. C., Schühli, G. S., Artoni, R. F., et al. (2011). Differentiation of repetitive DNA sites and sex chromosome systems reveal closely related group in Parodontidae (Actinopterygii: Characiformes). *Genetica* 139 (11), 1499–1508. doi:10.1007/s10709-012-9649-6

Schemberger, M. O., Nascimento, V. D., Coan, R., Ramos, É., Nogaroto, V., Ziemniczak, K., et al. (2019). DNA transposon invasion and microsatellite accumulation guide W chromosome differentiation in a Neotropical fish genome. *Chromosoma* 128 (4), 547–560. doi:10.1007/s00412-019-00721-9

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLOS ONE* 6 (3), e17288. doi:10.1371/journal.pone.0017288

Sember, A., Bertollo, L. A. C., Ráb, P., Yano, C. F., Hatanaka, T., de Oliveira, E. A., et al. (2018). Sex chromosome evolution and genomic divergence in the fish *Hoplias malabaricus* (Characiformes, Erythrinidae). *Front. Genet.* 9, 71. doi:10.3389/fgene.2018.00071

Sember, A., Nguyen, P., Perez, M. F., Altmanová, M., Ráb, P., and Cioffi, M. d. B. (2021). Multiple sex chromosomes in teleost fishes from a cytogenetic perspective: state of the art and future challenges. *Phil. Trans. R. Soc. B* 376 (1833), 20200098. doi:10.1098/rstb.2020.0098

Serrano-Freitas, É. A., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Utsunomia, R., Araya-Jaime, C., Oliveira, C., et al. (2020). Satellite DNA content of B chromosomes in the characid fish *Characidium gomesi* supports their origin from sex chromosomes. *Mol. Genet. Genomics* 295 (1), 195–207. doi:10.1007/s00438-019-01615-2

Silva, D. M. Z. d. A., Utsunomia, R., Ruiz-Ruano, F. J., Daniel, S. N., Porto-Foresti, F., Hashimoto, D. T., et al. (2017). High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Sci. Rep.* 7 (1), 12726. doi:10.1038/s41598-017-12939-7

Smit, A. F. A., Hubley, R., and Green, P. (2013). *Repeat Masker*. *2013*. [Online]. Available: http://www.repeatmasker.org/(Accessed 10, 2020).

Splendore de Borba, R., Lourenço da Silva, E., and Parise-Maltempi, P. P. (2013). Chromosome mapping of retrotransposable elements Rex1 and Rex3 in *Leporinus* Spix, 1829 species (Characiformes: Anostomidae) and its relationships among heterochromatic segments and W sex chromosome. *Mobile Genet. Elem.* 3 (6), e27460. doi:10.4161/mge.27460

Sproul, J. S., Khost, D. E., Eickbush, D. G., Negm, S., Wei, X., Wong, I., et al. (2020). Dynamic evolution of euchromatic satellites on the X chromosome in *Drosophila melanogaster* and the simulans clade. *Mol. Biol. Evol.* 37 (8), 2241–2256. doi:10.1093/molbev/msaa078

Terencio, M. L., Schneider, C. H., Gross, M. C., Nogaroto, V., de Almeida, M. C., Artoni, R. F., et al. (2012). Repetitive sequences associated with differentiation of W chromosome in *Semaprochilodus taeniurus*. *Genetica* 140 (10-12), 505–512. doi:10.1007/s10709-013-9699-4

Thakur, J., Packiaraj, J., and Henikoff, S. (2021). Sequence, chromatin and evolution of satellite DNA. *Ijms* 22 (9), 4309. doi:10.3390/ijms22094309

Traut, W., and Winking, H. (2001). Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Res.* 9 (8), 659–672. doi:10.1023/A:1012956324417

Ugarković, Đ., and Plohl, M. (2002). Variation in satellite DNA profiles-causes and effects. *Embo J.* 21 (22), 5955–5959. doi:10.1093/emboj/cdf612

Utsunomia, R., Ruiz-Ruano, F. J., Silva, D. M. Z. A., Serrano, É. A., Rosa, I. F., Scudeler, P. E. S., et al. (2017). A glimpse into the satellite DNA library in Characidae fish (Teleostei, Characiformes). *Front. Genet.* 8, 103. doi:10.3389/fgene.2017.00103

Utsunomia, R., Silva, D. M. Z. d. A., Ruiz-Ruano, F. J., Goes, C. A. G., Melo, S., Ramos, L. P., et al. (2019). Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* 9 (1), 5856. doi:10.1038/s41598-019-42383-8

Volff, J.-N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity* 94 (3), 280–294. doi:10.1038/sj.hdy.6800635

Vondrak, T., Ávila Robledillo, L., Novák, P., Koblížková, A., Neumann, P., and Macas, J. (2020). Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. *Plant J.* 101 (2), 484–500. doi:10.1111/tpj.14546

Wright, A. E., Dean, R., Zimmer, F., and Mank, J. E. (2016). How to make a sex chromosome. *Nat. Commun.* 7, 12087. doi:10.1038/ncomms12087

Yano, C. F., Bertollo, L. A. C., Ezaz, T., Trifonov, V., Sember, A., Liehr, T., et al. (2017). Highly conserved Z and molecularly diverged W chromosomes in the fish genus *Triportheus* (Characiformes, Triportheidae). *Heredity* 118, 276–283. doi:10.1038/hdy.2016.83

Yano, C. F., Bertollo, L. A. C., Liehr, T., Troy, W. P., and Cioffi, M. d. B. (2016). W Chromosome Dynamics in *Triportheus* Species (Characiformes, Triportheidae): An Ongoing Process Narrated by Repetitive Sequences. *Jhered* 107 (4), 342–348. doi:10.1093/jhered/esw021

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 October 2021
doi: 10.3389/fgene.2021.769984

Tracking the Evolutionary Trends Among Small-Size Fishes of the Genus *Pyrrhulina* (Characiforme, Lebiasinidae): New Insights From a Molecular Cytogenetic Perspective

Renata Luiza Rosa de Moraes[1], *Francisco de Menezes Cavalcante Sassi*[1], *Luiz Antonio Carlos Bertollo*[1], *Manoela Maria Ferreira Marinho*[2,3], *Patrik Ferreira Viana*[4], *Eliana Feldberg*[4], *Vanessa Cristina Sales Oliveira*[1], *Geize Aparecida Deon*[1,5], *Ahmed B. H. Al-Rikabi*[6], *Thomas Liehr*[6]* and *Marcelo de Bello Cioffi*[1]

[1]*Laboratorio de Citogenética de Peixes, Departamento de Genética e Evolução, Universidade Federal de São Carlos (UFSCar), São Carlos, Brazil,* [2]*Museu de Zoologia da Universidade de São Paulo (MZUSP), São Paulo, Brazil,* [3]*Laboratório de Sistemática e Morfologia de Peixes, Departamento de Sistemática e Ecologia (DSE), Universidade Federal da Paraíba (UFPB), João Pessoa, Brazil,* [4]*Laboratório de Gentética Animal, Instituto Nacional de Pesquisa da Amazônia, Coordenação de Biodiversidade, Manaus, Brazil,* [5]*Laboratório de Biologia Cromossômica, Estrutura e Função, Departamento de Biologia Estrutural, Molecular e Genética, Universidade Estadual de Ponta Grossa, Ponta Grossa, Brazil,* [6]*Institute of Human Genetics, University Hospital Jena, Jena, Germany*

Miniature fishes have always been a challenge for cytogenetic studies due to the difficulty in obtaining chromosomal preparations, making them virtually unexplored. An example of this scenario relies on members of the family Lebiasinidae which include miniature to medium-sized, poorly known species, until very recently. The present study is part of undergoing major cytogenetic advances seeking to elucidate the evolutionary history of lebiasinids. Aiming to examine the karyotype diversification more deeply in *Pyrrhulina*, here we combined classical and molecular cytogenetic analyses, including Giemsa staining, C-banding, repetitive DNA mapping, comparative genomic hybridization (CGH), and whole chromosome painting (WCP) to perform the first analyses in five *Pyrrhulina* species (*Pyrrhulina* aff. *marilynae*, *Pyrrhulina* sp., *P. obermulleri*, *P. marilynae* and *Pyrrhulina* cf. *laeta*). The diploid number (2n) ranged from 40 to 42 chromosomes among all analyzed species, but *P. marilynae* is strikingly differentiated by having 2n = 32 chromosomes and a karyotype composed of large meta/submetacentric chromosomes, whose plesiomorphic status is discussed. The distribution of microsatellites does not markedly differ among species, but the number and position of the rDNA sites underwent significant changes among them. Interspecific comparative genome hybridization (CGH) found a moderate divergence in the repetitive DNA content among the species' genomes. Noteworthy, the WCP reinforced our previous hypothesis on the origin of the X_1X_2Y multiple sex chromosome system in *P. semifasciata*. In summary, our data suggest that the karyotype differentiation in *Pyrrhulina* has been driven by major structural rearrangements, accompanied by high dynamics of repetitive DNAs.

Keywords: fishes, repetitive DNAs, karyotype evolution, sex chromosomes, evolution

OPEN ACCESS

Edited by:
Ricardo Utsunomia, Federal Rural University of Rio de Janeiro, Brazil

Reviewed by:
František Marec, Biology Centre of the Czech Academy of Sciences (CAS), Institute of Entomology, Czechia
Daniel Garcia-Souto, University of Vigo, Spain
Camilla Di Nizo, Zoological Research Museum Alexander Koenig (LG), Germany

***Correspondence:**
Thomas Liehr
Thomas.Liehr@med.uni-jena.de

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *03 September 2021*
Accepted: *22 September 2021*
Published: *06 October 2021*

Citation:
de Moraes RLR, Sassi FdMC, Bertollo LAC, Marinho MMF, Viana PF, Feldberg E, Oliveira VCS, Deon GA, Al-Rikabi ABH, Liehr T and Cioffi MdB (2021) Tracking the Evolutionary Trends Among Small-Size Fishes of the Genus Pyrrhulina *(Characiforme, Lebiasinidae): New Insights From a Molecular Cytogenetic Perspective. Front. Genet. 12:769984. doi: 10.3389/fgene.2021.769984*

INTRODUCTION

Characiformes comprise a very diverse and abundant freshwater order (Nelson et al., 2016), in which the family Lebiasinidae is represented by 75 valid species (Fricke et al., 2021) widely distributed across South and Central America (Weitzman and Weitzman, 2003). The phylogenetic relationships of the Lebiasinidae remained in doubt for a long time, but more recent phylogenetic analysis indicate their proximity to the Ctenoluciidae (Calcagnotto et al., 2005; Oliveira et al., 2011), which was also reinforced by the different studies (Arcila et al., 2017; Betancur-R et al., 2019; Melo et al., 2021). Most Lebiasinidae species reach about 60 mm of Standard Length (SL), but miniature species, not surpassing a maximum of 26 mm SL, is found within the Pyrrhulininae, whereas medium-sized species up to 150 mm SL can be found within Lebiasininae (Weitzman and Weitzman, 2003).

Because of their small sizes and difficulties in obtaining good chromosomal preparations, species of Lebiasinidae were, for a long time, little analyzed in terms of cytogenetics, with scarce references mainly on the chromosomal number of few species (Scheel, 1973; Oliveira et al., 1991; Arai, 2011). However, this scenario has recently undergone significant changes with the methodological advance of cytogenetics and its applicability among small to miniature fishes of *Pyrrhulina*, *Lebiasina*, *Copeina*, and *Nannostomus* genus (de Moraes et al., 2017, de Moraes et al.,2019; Sassi et al., 2019; Toma et al., 2019; Sassi et al., 2020; Sember et al., 2020).

Pyrrhulina is one of the most speciose genera of the subfamily Pyrrhulininae, with 19 valid small species (Fricke et al., 2021), ranging from 30.4 to 85 mm SL (Weitzman and Weitzman, 2003; Netto-Ferreira and Marinho, 2013). The genus is among the most problematic, with many poorly known species, species complexes, and old taxonomic problems (Netto-Ferreira and Marinho, 2013). The first *Pyrrhulina* species to have some chromosomal data evidenced was *Pyrrhulina* cf. *australis*, with 2n = 40 chromosomes, mainly acrocentric ones (Oliveira et al., 1991). Taxonomic boundaries of *P. australis* are still poorly defined, demonstrated in subsequent studies (de Moraes et al., 2017; de Moraes et al.,2019) of two morphotypes. Both *P. australis* and *Pyrrhulina* aff. *australis* showed similar data 2n = 40 (4st + 36a), distinct from *P. brevis*, 2n = 42 (2sm + 4st + 36a), with no evidence of heteromorphic sex chromosomes in the three species (de Moraes et al., 2017; de Moraes et al., 2019). Another species, *P. semifasciata*, was analyzed, presenting 2n = 42 (4st + 38a) in females, and 2n = 41 (1m + 4st + 36a) in males, the latter with three unpaired chromosomes because of a multiple $X_1X_1X_2X_2/X_1X_2Y$ sex chromosome system (de Moraes et al., 2019). This occurrence was also confirmed by comparative genomic hybridizations (CGH) and whole-chromosome painting (WCP), with some indications that the Y chromosome originated by centric fusions of non-homologous acrocentric chromosomes (de Moraes et al., 2019).

To improve the knowledge of the evolutionary processes within the genus *Pyrrhulina*, we combined classical and molecular cytogenetic analyses, including Giemsa staining, C-banding, repetitive DNA mapping, comparative genomic hybridization (CGH), and whole chromosome painting (WCP to perform the first analyses in five *Pyrrhulina* species (*Pyrrhulina* aff. *marilynae*, *Pyrrhulina* sp., *P. obermulleri*, *P. marilynae* and *Pyrrhulina* cf. *laeta*). The results highlighted relationships and particular evolutionary paths at the chromosomal and genomic levels among the species. In addition, the hypothesis on the origin of the multiple sex chromosome system in *P. semifasciata* is validated.

MATERIALS AND METHODS

Animals

The collection sites, number, and sex of the specimens investigated are presented in **Figure 1**, **Table 1**. Part of the sampling (**Figure 1**, white circles) resembles the one previously analyzed by de Moraes et al. (2017), de Moraes et al. (2019) with different cytogenetic and molecular methods. Animals were collected with the authorization of the Brazilian environmental agency ICMBIO/SISBIO (license no. 48628-14) and SISGEN (A96FF09). All species were properly identified by morphological criteria, and the specimens were deposited in the fish collection of the Museu de Zoologia da Universidade de São Paulo (MZUSP) under the voucher numbers (119077, 119079, 123073, 123080) and the Universidade Federal da Paraíba (UFPB) museum under the voucher number (12079, 12080, 12082 and 12083). Experiments followed ethical and anesthesia conducts and were approved by the Ethics Committee on Animal Experimentation of the Universidade Federal de São Carlos (process number CEUA 1853260315).

Chromosomal Preparations and Analysis of the Constitutive Heterochromatin

Mitotic chromosomes were obtained from kidney cells by the protocol described in Bertollo et al. (2015). The distribution of constitutive heterochromatin was observed by the C-banding methodaccording to (Sumner, 1972).

Repetitive DNA Mapping with Fluorescence *in situ* Hybridization (FISH)

The 5S rDNA probe included 120 base pairs (bp) of the 5S rDNA gene coding region and 200 bp of non-transcribed spacer (NTS) (Pendás et al., 1994). The 18S rDNA probe was composed of a 1,400-bp-long segment of the 18S rDNA coding region (Cioffi et al., 2009). Both probes were directly labeled with the Nick-Translation Mix Kit (Jena Bioscience, Jena, Germany)—18S rDNA with ATTO488-dUTP and 5S rDNA with ATTO550-dUTP, according to the manufacturer's instructions. The $(CA)_{15}$, $(GA)_{15}$, $(CGG)_{10}$ microsatellite probes were directly labeled with Cy3 during the synthesis, according to Kubat et al. (2008). In addition, since it contains the lowest 2n, telomeric $(TTAGGG)_n$ sequence was also used as probe in *P. marylinae*. This probe was generated by PCR in the absence of a template according to Ijdo et al. (1991) and later labeled with ATTO550-dUTP with the Nick-Translation Mix Kit (Jena

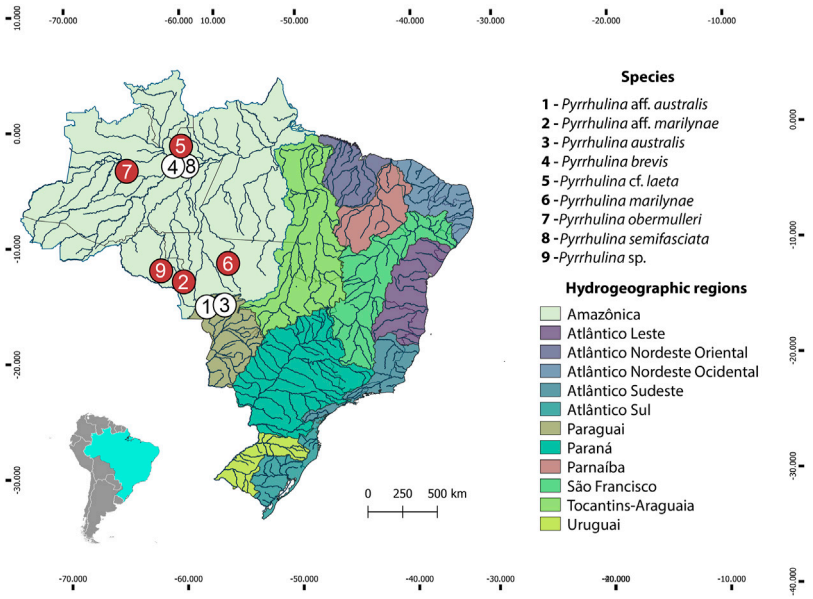

FIGURE 1 | Brazilian collection sites of the *Pyrrhulina* species cytogenetically investigated in the present study (red circles) and the ones previously cytogenetically analyzed (white circles: data from (de Moraes et al., 2017; de Moraes et al., 2019).

TABLE 1 | Geographical coordinates, sample size, and diploid number of *Pyrrhulina* (Characiformes, Lebiasinidae) species collected in Brazil.

Species	Locality	Sample size	2n (Sex)	References
Pyrrhulina aff. *australis*	Rio Sepotuba, Lambari D'Oeste—MT (15°11′28.0″S 57°41′30.7″W)	16♂ 22♀	40♂♀	de Moraes et al. (2017)
Pyrrhulina aff. *marilynae*	Igarapé 12 de Outubro, Comodoro—MT (12°58′41.0″S 60°00′34.0″W)	14♂ 10♀	40♂♀	This study
P. australis	Barra do Bugres—MT (15°04′27.5″S 57°11′05.4″W)	18♂ 30♀	40♂♀	de Moraes et al. (2017)
P. brevis	Reserva Florestal Adolpho Ducke, Manaus—AM (2°58′20.7″S 59°55′53.0″W)	17♂ 13♀	42♂♀	de Moraes et al. (2019)
Pyrrhulina cf. *laeta*	Presidente Figueiredo—AM (1°59′10.8″S 60°03′40.8″W)	07♂ 05♀	42♂♀	This study
P. marilynae	Ipiranga do Norte—MT (11°36′02.0″S 55°56′27.0″W)	14♂ 08♀	32♂♀	This study
P. obermulleri	Tefé—AM (3°25′50.7″S 64°44′54.8″W)	21♂ 12♀	42♂♀	This study
P. semifasciata	Careiro—AM (3°51′00.0″S 60°04′00.0″W)	12♂ 09♀	41♂42♀	de Moraes et al. (2019)
Pyrrhulina sp	Represa, Alto Alegre dos Parecis—RO (12°11′58.0″S 61°46′47.7″W)	19♂ 29♀	40♂♀	This study

Bioscience, Jena, Germany). FISH experiments followed the methodology described in Yano et al. (2017). Metaphase chromosomes were treated with RNAse A (40 µg/ml) for 1.5 h at 37°C and the DNA denatured in 70% formamide/2× SSC at 72°C for 3.15 min. A hybridization mixture (2.5 ng/µL probes, 50% deionized formamide, 10% dextran sulfate) was then dropped on the slides, and the hybridization process was performed overnight at 37°C in a moist chamber. The first post-hybridization wash was performed with 1× SSC for 5 min at 65°C, followed by the second one performed with 4xSSC/Tween for 5 min, at room temperature. Chromosomes were then counterstained with DAPI, and the slides were mounted with an antifade solution (Vectashield from Vector Laboratories, Burlingame, CA).

FISH for Whole Chromosome Painting

As *P. semifasciata* represents the only *Pyrrhulina* species that harbors an X_1X_2Y multiple sex system, a Y-chromosome probe, named PSEMI-Y, was previously prepared by microdissection, as described in (de Moraes et al., 2019) Male and female metaphases of *P. marilynae*, *Pyrrhulina* aff. *marilynae*, *Pyrrhulina* sp., *P. obermulleri*, *Pyrrhulina* cf. *laeta* were used for Zoo-FISH experiments with the PSEMI-Y probe, according to procedures described in Yano et al. (2017). The hybridization was performed for 72 h at 37°C in a moist chamber, with post-hybridization washes with 1xSSC for 5 min at 65°C, and in 4xSSC/Tween (RT). 10 µg of male-derived C_0t-1 DNA from *P. semifasciata* was used as suppressor in each experiment. Chromosomes were stained with DAPI (1.2 µg/ml) and the slides were mounted with an antifade solution, as described above.

Probes for Comparative Genomic Hybridization

The genomic DNAs (gDNAs) from male and female specimens of *P. marilynae*, *Pyrrhulina* aff. *marilynae*, *Pyrrhulina* sp., *P. obermulleri*, *Pyrrhulina* cf. *laeta*, *P. australis*, *Pyrrhulina* aff.

australis, *P. brevis*, and *P. semifasciata* were extracted from liver tissue by the standard phenol-chloroform-isoamyl alcohol method (Sambrook and Russell, 2001). For intraspecific comparisons, the male-derived gDNAs of all species were labeled with ATTO550-dUTP and the female gDNAs with ATTO 488-dUTP, by nick translation (Jena Bioscience, Jena, Germany). The repetitive sequences were blocked using unlabeled C_0t-1 DNA in all experiments, according to (Zwick et al., 1997). The final hybridization mixture for each slide (20 μL) was composed of male- and female-derived gDNAs (500 ng each), plus 25 μg of female-derived C_0t-1 DNA from the respective species. The probe was ethanol-precipitated, and the dry pellets were mixed in a hybridization mixture containing 50% formamide, 2× SSC, 10% SDS, 10% dextran sulfate, and Denhardt´s buffer, pH 7.0.

For interspecific comparisons, the gDNA of male specimens of *P. australis* (Paus), *Pyrrhulina* aff. *australis* (Pafa), *P. semifasciata* (Psem), *P. brevis* (Pbre), *P. marilynae* (Pmar), *Pyrrhulina* aff. *marilynae* (Pafm), *Pyrrhulina* sp. (Psp), *P.obermulleri* (Pobe) and *Pyrrhulina* cf. *laeta* (Pcfl) were hybridized against metaphase chromosomes of *P. marilynae*. This species was selected since it harbors the lowest 2n = 32 until now register for the genus, coupled with a remarkable karyotype differentiation. For this purpose, male-derived gDNA of *P. marilynae* was labeled with ATTO 550-dUTP, while the gDNAs of the other species were labeled with ATTO 488-dUTP (*P. australis*, *Pyrrhulina* aff. *marilynae*, *P. brevis* and *P. obermulleri*) or ATTO 425-dUTP (*Pyrrhulina* aff. *australis*, *Pyrrhulina* sp., *P. semifasciata* and *Pyrrhulina* cf. *laeta*), both through nick translation (Jena Bioscience, Jena, Germany).

The interspecific comparisons were divided into a set of four slides. In the first slide, the final probe mixture was composed of 500 ng of male-derived gDNA plus 10 μg of male-derived C_0t-1 DNA of each of the following species: *P. marilynae*, *P. australis*, and *Pyrrhulina* aff. *australis*. In the second slide, the final probe mixture was composed of 500 ng of male-derived gDNA plus 10 μg of male-derived C_0t-1 DNA of each one of the following species: *P. marilynae*, *Pyrrhulina* aff. *marilynae* and *Pyrrhulina* sp. In the third slide, the final probe mixture was composed of 500 ng of male-derived gDNA plus 10 μg of male-derived C_0t-1 DNA of each one of the following species: *P. marilynae*, *P. brevis*, and *P. semifasciata*. Finally, in the fourth slide, the final probe mixture was composed of 500 ng of male-derived gDNA plus and 10 μg of male-derived C_0t-1 DNA of each one of the following species: *P. marilynae*, *P. obermulleri*, and *Pyrrhulina* cf. *laeta*. The chosen ratio of probe vs. C_0t-1 DNA amount was based on previous experiments performed in our fish studies (de Moraes et al., 2019; Toma et al., 2019; Sassi et al., 2020). The CGH experiments followed the methodology described in Sember et al. (2018).

Microscopy and Images Processing

To confirm the diploid number, karyotype structure and FISH results inat least 30 metaphase spreads were analyzed per individual. The microscopy images were captured using an Olympus BX50 epifluorescence microscope (Olympus Corporation, Ishikawa, Japan) coupled with a CoolSNAP camera, and the images were processed using Image-Pro Plus 4.1 Software (Media Cybernetics, Silver Spring, MD, United States). Final images were optimized and arranged using Adobe Photoshop, version CC 2020. Chromosomes were classified as metacentric (m), submetacentric (sm), subtelocentric (st), or acrocentric (a), according to their arm ratios (Levan, 1964). As the males and females results showed no differences, only male metaphases were represented in the figures.

RESULTS

Karyotypes and Heterochromatin Distribution

The diploid number ranged from 2n = 40 to 42 among the following four species: *Pyrrhulina* sp. (2n = 40; 2st+38a), *Pyrrhulina* aff. *marilynae* (2 = 40; 40a), *P. obermulleri* (2n = 42; 2m/sm+8st+32a) and *Pyrrhulina* cf. *laeta* (2n = 42; 2m/sm+4st+36a), the two latter also sharing a characteristic small metacentric/submetacentric pair. On the other hand, *P. marilynae* differed by presenting a very distinct karyotype composition (2n = 32; 8m/sm+4st+20a). These results represent the first cytogenetic data for the abovementioned species. The constitutive heterochromatin was distributed at the pericentromeric region of several chromosome pairs in *P. marilynae* and *Pyrrhulina* aff. *marilynae*. In its turn, *Pyrrhulina* sp., *P. obermulleri*, and *Pyrrhulina* cf. *laeta* presented a remarkable series of interstitial and pericentromeric C-bands, in addition to telomeric ones (**Figure 2**). In our sampling, we did not observe any karyotype differences between males and females.

Chromosomal Mapping of Repetitive DNA Sequences

All the five species differ by the distribution of the multigene rDNA families. *Pyrrhulina* sp. and *P. marilynae* were the only species with only one chromosome pair bearing 18S rDNA sites, found at the telomeric region of acrocentric pairs 4 and 9, respectively. Six to twelve centromeric or telomeric sites occur in the other three species, including bitelomeric sites in *Pyrrhulina* aff. *marilynae* (pair 11) and *Pyrrhulina* cf. *laeta* (pairs 6 and 13). As for the 5S rDNA, from six to twelve centromeric sites occured among species, including a syntenic condition for the 5S and 18S rDNA repeats in the chromosome pair 6 of *Pyrrhulina* cf. *laeta*, the same pair that displays bitelomeric 18S rDNA signals in this species (**Figure 2**). The distribution of the microsatellites $(CA)_{15}$, $(GA)_{15}$, and $(CGG)_{10}$ does not differ significantly among species, having a preferential location in the centromeric and telomeric regions of the chromosomes, in addition to some interstitial sites. However, $(CA)_{15}$ differs quantitatively, with a greater number of conspicuous sites compared to the

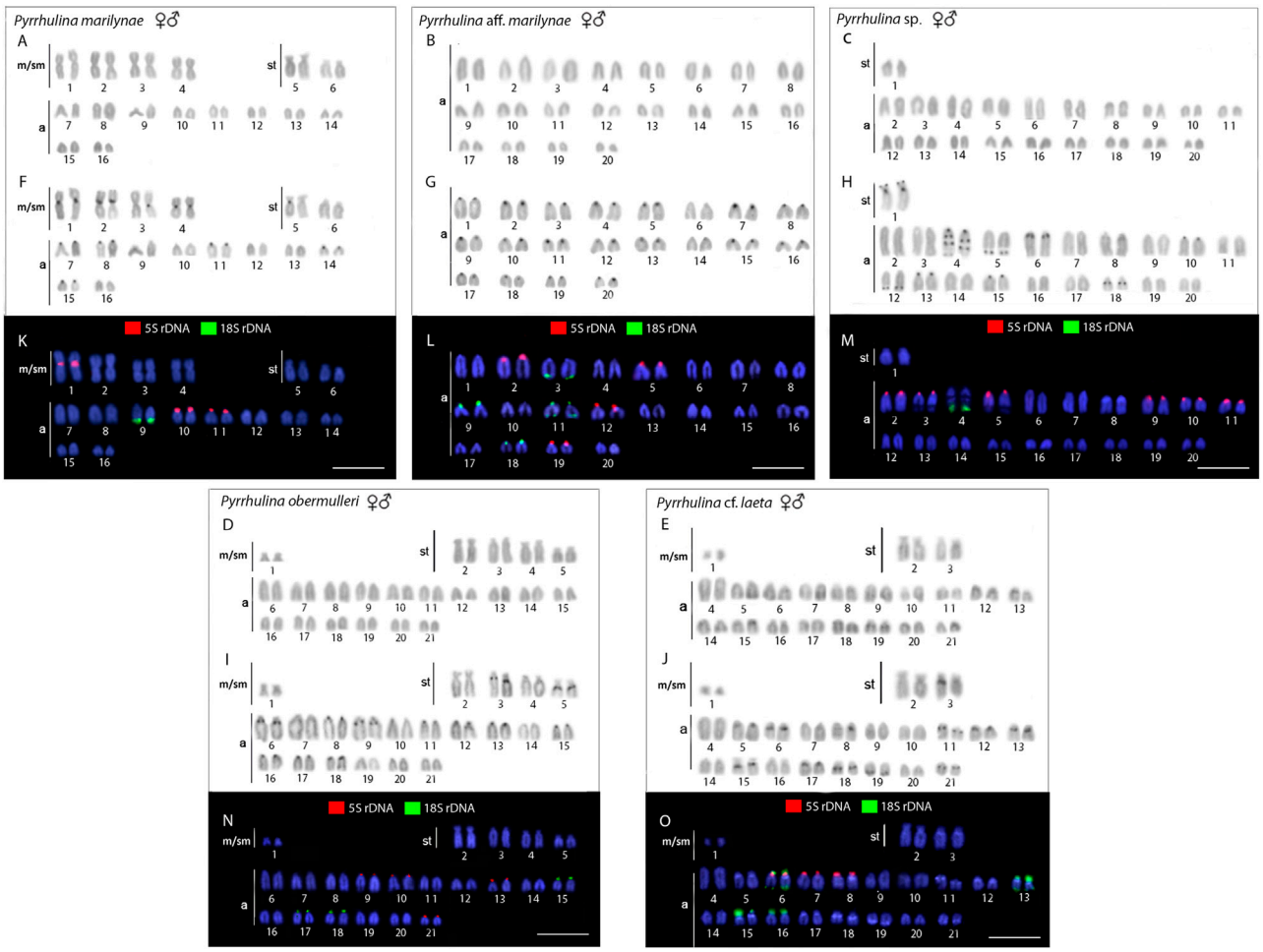

FIGURE 2 | Male and female karyotypes of *Pyrrhulina marilynae* **(A, F, and K)**, *Pyrrhulina* aff. *marilynae* **(B, G and L)**, *Pyrrhulina* sp. **(C, H and M)**, *P. obermulleri* **(D, I and N)** and *Pyrrhulina* cf. *laeta* **(E, J and O)** arranged after Giemsa staining **(A-E)**, C-banding **(F-J)**, and dual-color *in situ* hybridization (FISH) with 18S (green) and 5S (red) ribosomal DNA probes **(K-O)**. Chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI).Scale bar = 5 μm.

other microsatellites, especially in *Pyrrhulina* aff. *marilynae* and *Pyrrhulina* cf. *laeta*. In the same way, $(CGG)_{10}$ occurs in smaller amounts in the five species (**Figure 3**). The (TTAGGG)n repeats showed the expected hybridization signals on telomeres of *P. marylinae* (**Figure 4F**). Whole chromosome painting–WCP.

Two acrocentric chromosome pairs were entirely painted with the PSEMI-Y probe in *Pyrrhulina marilynae*, *P. obermulleri*, *Pyrrhulina* sp., *Pyrrhulina* aff. *marilynae* and *Pyrrhulina* cf. *laeta* (**Figures 4A–E**).

Comparative Genomic Hybridization–CGH

The interespecific genomic comparison among *Pyrrhulina marilynae* and other *Pyrrhulina* species (*P. semifasciata*, *P. australis*, *P. brevis*, *P. obermulleri*, *Pyrrhulina* aff. *australis*, *Pyrrhulina* sp., *Pyrrhulina* aff. *marilynae*, *Pyrrhulina* cf. *laeta*) revealed a high level of DNA compartmentalization, within all species presenting a distinct composition of repetitive DNA sequences and specific signals. However, *P. marilynae* shows more evident species-specific arrangements when compared to the other species. (**Figure 5**). Intraspecific genomic hybridization between males and females did not show any clustering for sex-specific sequences in all species (data not shown).

DISCUSSION

Overall, two main evolutionary trends are proposed for the karyotypic evolution of the Lebiasinidae: 1) the conservation of a plesiomorphic karyotype in the subfamily Lebiasininae, with 2n = 36 bi-armed chromosomes and, 2) high variations in diploid numbers and karyotypic structures in the subfamily Pyrrhulininae, with the predominance of acrocentric chromosomes (Sassi et al., 2020). It is noteworthy that the karyotypic structure of Lebiasininae, 2n = 36 biarmed chromosomes, is similar to that found in the sister family Ctenoluciidae (de Souza e Sousa et al., 2017; Sassi et al., 2019;

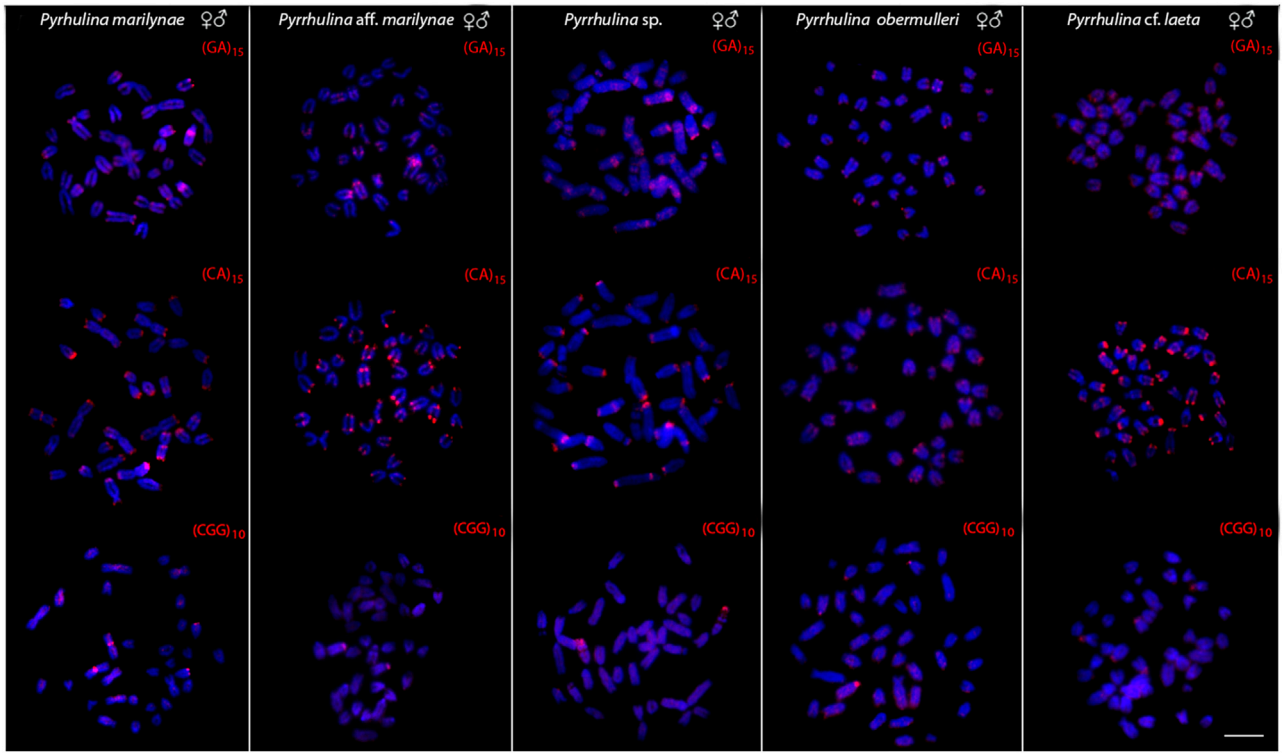

FIGURE 3 | Male and female metaphase plates of *Pyrrhulina marilynae*; *Pyrrhulina* aff. *marilynae*; *Pyrrhulina* sp.; *P. obermulleri* and *Pyrrhulina* cf. *laeta* shows the general distribution of the microsatellites $(GA)_{15}$, $(CA)_{15}$ and $(CGG)_{10}$ on chromosomes. Bar = 5 μm.

de Souza e Sousa et al., 2021). Therefore, in this scenario, the majority of the acrocentric chromosomes found in the species of the Pyrrhulininae are probably derived from rearrangements such as centric fissions (Sassi et al., 2020). However, unlike other *Pyrrhulina* species, *P. marilynae* has the smallest 2n identified in the genus so far, 2n = 32, including four typical meta/submetacentric pairs. Some exceptions within the subfamily showed secondary fusion events of

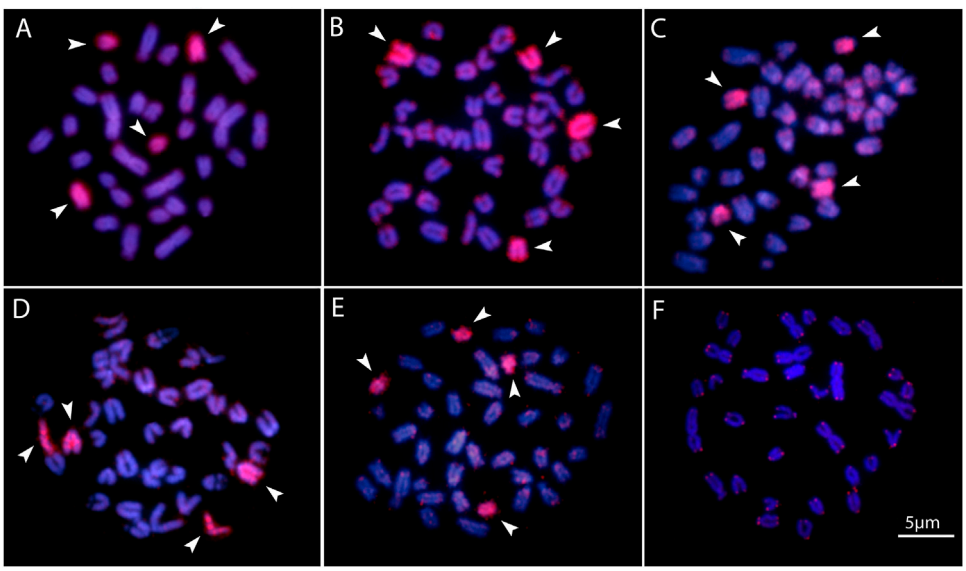

FIGURE 4 | Zoo-FISH with the PSEMI-Y probeon male metaphase plates of *P. marilynae* **(A)**, *Pyrrhulina* aff. *marilynae* **(B)**, *Pyrrhulina* cf. *laeta* **(C)**, *Pyrrhulina* sp. **(D)**, and *P. obermulleri* **(F)** shows the distribution of the telomeric (TTAGGG)n repeats in *P. marilynae*. Bar = 5 μm.

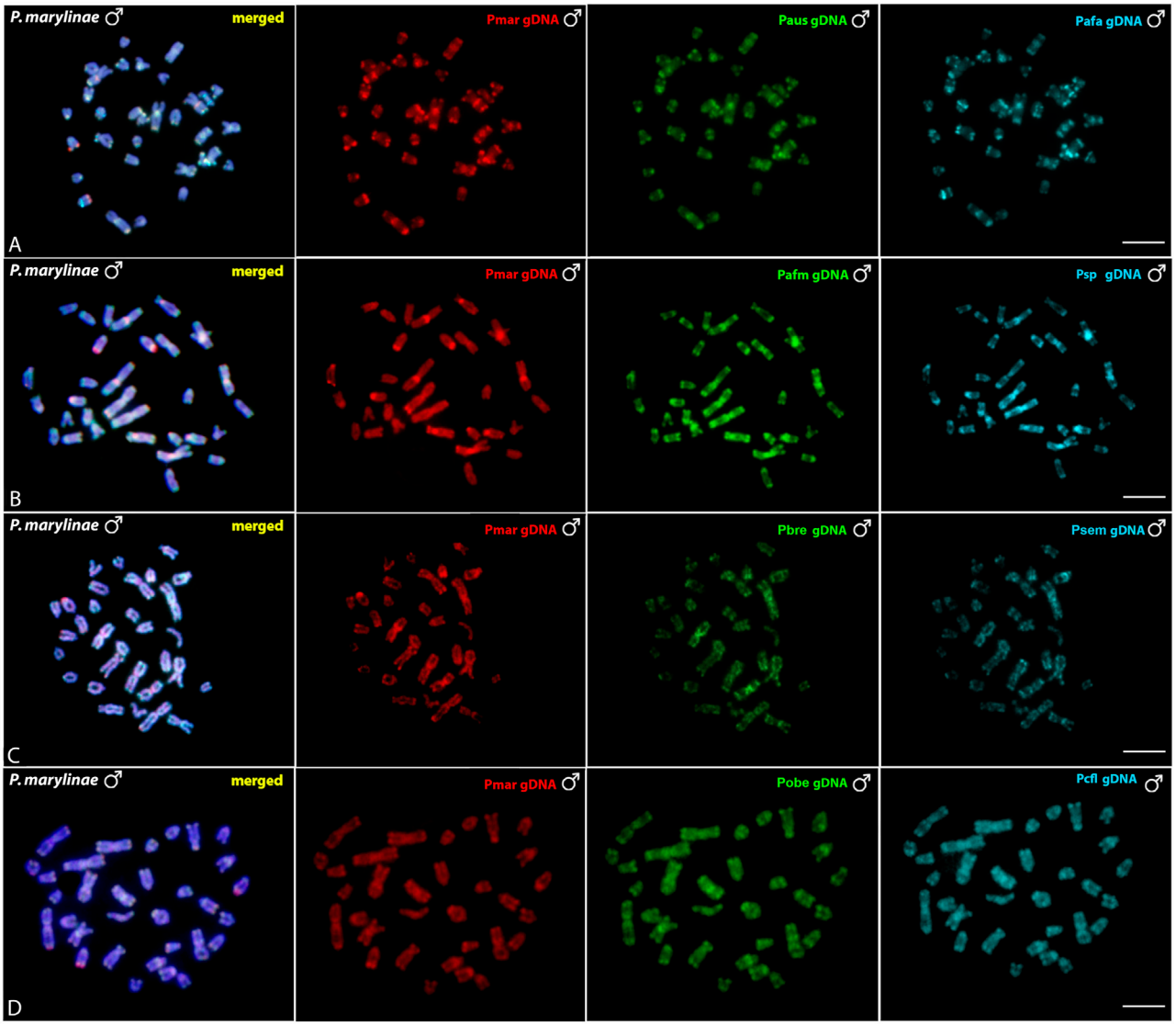

FIGURE 5 | Comparative genomic hybridization (CGH) using male-derived genomic probes from *Pyrrhulina* species hybridized onto male chromosomes of *P. marilynae*. The common genomic regions are depicted in the 1st column in each line representing the experiments A-D. Hybridization between the gDNA of *P. marilynae* (Pmar), *P. australis* (Paus) and *Pyrrhulina* aff. *australis* (Pafa) **(A)**; *P. marilynae*(Pmar), *Pyrrhulina* aff. *marilynae* (Pafm) and *Pyrrhulina* sp. (Psp) **(B)**; *P. marilynae*(Pmar), *P. brevis* (Pbre) and *P. semifasciata* (Psem) **(C)**; *P. marilynae* (Pmar), *P. obermulleri* (Pobe) and *Pyrrhulina* cf. *laeta* (Pcfl) **(D)**. Bar = 5 μm.

acrocentric chromosomes giving rise to metacentric chromosomes, reducing the diploid number as observed in *Nannostomus unifasciatus* (Sember et al., 2020). Biarmed chromosomes could also represent remnants of the ancestral karyotype condition that were maintained during the evolutionary processes. However, no ITS was found in any chromosome of *P. marilynae*, but such a scenario does not exclude the hypothesis of fusion, given that telomeric regions can be lost after the rearrangement occurs (Bolzán, 2017). Thus, to corroborate such hypotheses and to determine whether the evolutionary trajectory of karyotype change in *Pyrrhulina* is directed mainly towards centric fusions or fissions, cytogenetic data should be discussed in a larger phylogenetic framework of interspecific and intergeneric relationships of Lebiasinidae.

CGH procedures have greatly assisted cytogenetic studies (Symonová et al., 2013; Cioffi et al., 2017; Cioffi et al.,2019), as among all *Pyrrhulina* studied so far. In fact, despite showing close genomic similarities, the species also show considerable divergences, in addition to species-specific repetitive DNA and C-band patterns, thus helping to understand their differential evolutionary paths, considering the taxonomic problems still pending in this fish group. In addition, multiple and syntenic ribosomal sites are not frequently observed among fishes, but these chromosomal features are very informative cytotaxonomic markers regarding Pyrrhulininae species. Comparatively, they

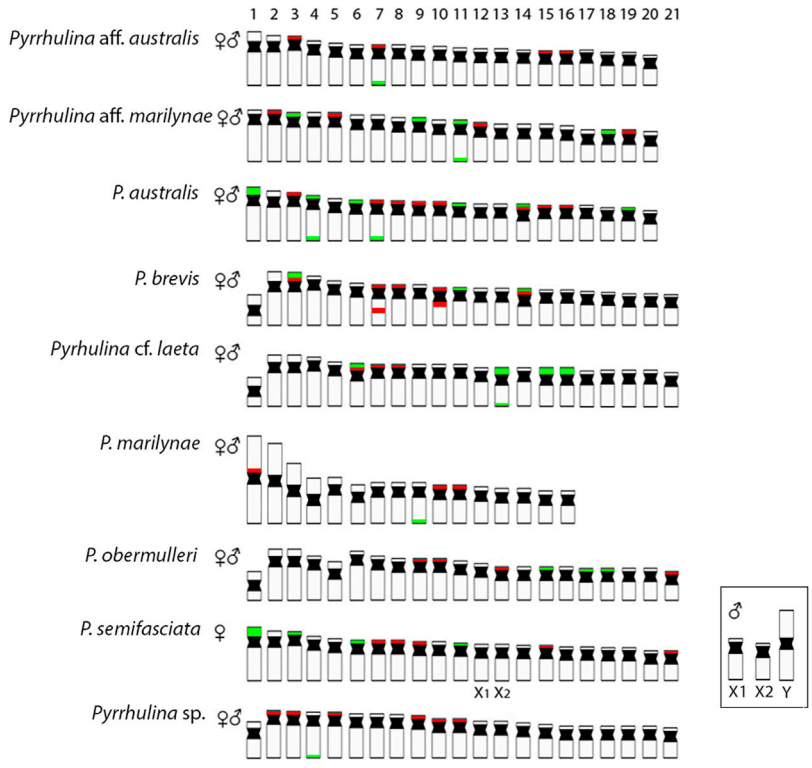

FIGURE 6 | Representative idiograms of *Pyrrhulina* species showing the distribution of the 18S (green) and 5S rDNA (red) sites on chromosomes, based on the present study and some other previous data (de Moraes et al., 2017; de Moraes et al., 2019). Bar = 5 µm.

occur more frequently among *Pyrrhulina* than in other species of this subfamily (de Moraes et al., 2017; de Moraes et al.,2019; Sassi et al., 2019; Sassi et al.,2020; Toma et al., 2019; Sember et al., 2020). Like *Pyrrhulina* aff. *australis* (de Moraes et al., 2017), *Pyrrhulina* sp., and *P. marilynae* present multiple 5S rDNA sites and only one 18S rDNA site, thus differentiating them from *Pyrrhulina* aff. *marilynae*, *P. obermulleri*, and *Pyrrhulina* cf. *laeta*, as well as from some other *Pyrrhulina* species (de Moraes et al., 2017; de Moraes et al.,2019), which have multiple 5S and 18S rDNA sites. Furthermore, the syntenic condition for the 18S/5S rDNAs in *Pyrrhulina* cf. *laeta* is shared with *P. brevis* and *P. australis*, indicating a high rDNA diversity. (**Figure 6**). In its turn, the 18S rDNA clusters are distributed on distal chromosome positions for all investigated *Pyrrhulina* species (de Moraes et al., 2017; de Moraes et al.,2019; this study), as also occur among *Copeina* (Toma et al., 2019), *Lebiasina* (Sassi et al., 2019), and *Nannostomus* (Sember et al., 2020), so as in the species of the sister family, Ctenoluciidae (de Souza e Sousa et al., 2017; de Souza e Sousa et al., 2021).

Microsatellite distribution patterns have significantly contributed to evolutionary studies in fish species, especially regarding sex chromosome differentiation (Kubat et al., 2008; Cioffi et al., 2012; Terencio et al., 2012; Kejnovský et al., 2013; Poltronieri et al., 2014; Yano et al., 2014; de Freitas et al., 2018). Among the five *Pyrrhulina* species now investigated, as well as in other previous analyzed ones (de Moraes et al., 2017; de Moraes et al.,2019), the distribution of the microsatellites did not significantly differ among them, although the $(CA)_{15}$ repeats present a greater number of more conspicuous sites than the other microsatellites, especially in *Pyrrhulina* aff. *marilynae* and *Pyrrhulina* cf. *laeta*. It is noteworthy that microsatellites have a preferential location in the telomeric and centromeric regions of fish chromosomes (Cioffi and Bertollo, 2012), as occur with the $(CA)_{15}$ and $(GA)_{15}$ motifs in *Pyrrhulina*, despite some interstitial and pericentromeric signs in *Pyrrhulina* cf. *laeta*, *P. marilynae*, *Pyrrhulina* aff. *marilynae* and *Pyrrhulina* sp., thus differentiating these species from others previously studied (de Moraes et al., 2017; de Moraes et al.,2019). Furthermore, it is also frequent that microsatellites and other repetitive sequences occur in the association among fish (Cioffi and Bertollo, 2012), such as in *Hepsetus odoe* (Carvalho et al., 2017), *Lebiasina bimaculata* (Sassi et al., 2019), and *Silurichthys phaiosoma* (Ditcharoen et al., 2020), for example. This is the scenario that also occurs in *Pyrrhulina* sp., in which the $(CGG)_{10}$ microsatellite located in the telomeric region of pair 4 shares the same chromosomal region with 18S rDNA repeats.

Fish, besides presenting high diversity in morphological and genetic characteristics, also have a variety of sex chromosome systems (Sember et al., 2021). About nine differentiated systems, involving the XX/XY and ZZ/ZW sex

chromosomes and their variations, have been identified among species, including several Neotropical ones (Sember et al., 2021). It is noteworthy that among the multiple systems, the ♀$X_1X_1X_2X_2$/♂X_1X_2Y is the most prevalent one, and commonly originated by centric or tandem fusions of the ancestral Y with an autosomal member of the karyotype, giving rise to neo-Y chromosomes, as identified in a variety of fish species (Sember et al., 2021). That includes *P. semifasciata*, the only Lebiasinidae representative highlighting heteromorphic sex chromosomes so far (de Moraes et al., 2019), in addition to a putative ZZ/ZW sex system present *in Lebiasina bimaculata* (Sassi et al., 2019). Although our intraspecific CGH results in the current analyzed species did not reveal any sex-specific differentiated region, our WCP experiment with the Y-derived probe of *P. semifasciata* entirely painted two acrocentric pairs, suggesting that putative proto-XY chromosomes may occur in these species. Thus, it supports our previous hypothesis on the origin of the *P. semifasciata* sex chromosome system through centric fusion between the non-homologous acrocentric, giving rise to the large metacentric Y chromosome. That can be considered as an apomorphy of this species when compared to others of the genus. Furthermore, the experiments also showed that although the karyotype of *P. marilynae* has large metacentric chromosomes, these do not correspond to the heteromorphic sex chromosome of *P. semifasciata* (**Figure 4**).

CONCLUSION

Our data advances the understanding of evolutionary trends of the Lebiasinidae, particularly concerning *Pyrrhulina*. Karyotypes with 2n = 40–42, with the predominance of mono-armed chromosomes, are more frequent among its species, except for *P. marilynae*, which has a smaller diploid number (2n = 32), and several atypical biarmed chromosomes, a characteristic that differentiates this species from the others analyzed in the genus. However, we cannot rule out the hypothesis that this karyotypic reduction (2n = 32) may have been generated by secondary fusions that allowed the formation of the four meta/submetacentric pairs identified in *P. marilynae*. The present data also highlighted the putative proto-XY chromosomes that may occur in these species and support the occurrence, through centric fusion, of the multiple sex chromosome system of *P. semifasciata* as an independent evolutionary event of this Lebiasinidae species. Our results highlight the importance of chromosomal data as valuable markers for understanding the evolutionary relationships among Lebiasinidae species.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by Ethics Committee on Animal Experimentation of the Universidade Federal de São Carlos (process number CEUA 1853260315).

AUTHOR CONTRIBUTIONS

RM and MC carried out the cytogenetic analysis and drafted the manuscript. TH, AA-R, and PV helped in the cytogenetics analysis, drafted and revised the manuscript. TL, GD, FS, VO, EF and MM drafted and revised the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

MC was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Proc. No. 302449/2018-3) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Proc. No. 2020/11772-8). RM and FS were supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Proc. No. 2019/25045-3 and 2020/02681-9, respectively). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Finance Code 001.

ACKNOWLEDGMENTS

The authors are grateful for FAPESP and CNPq. We also appreciate the contributions in the collection of animals from Hugmar Pains, Alany Pedrosa Gonçalves (Mamirauá Sustainable Development Reserve) and Ezequiel Aguiar de Oliveira.

REFERENCES

Arai, R. (2011). *Fish Karyotypes: A Check List*. Berlin, Germany: Springer Science & Business Media. doi:10.1007/978-4-431-53877-6

Arcila, D., Ortí, G., Vari, R., Armbruster, J. W., Stiassny, M. L. J., Ko, K. D., et al. (2017). Genome-wide Interrogation Advances Resolution of Recalcitrant Groups in the Tree of Life. *Nat. Ecol. Evol.* 1, 1–10. doi:10.1038/s41559-016-0020

Cioffi, M. B., Molina, W. F., Artoni, R. F., and Bertollo, L. A. C. (2012). Chromosomes as Tools for Discovering Biodiversity — The Case of Erythrinidae Fish Family," in *Recent Trends in Cytogenetic Studies—Methodologies and Applications*. Editor P. Tirunilai (Rijeka: IntechOpen), 125–146. doi:10.5772/35890

Bertollo, L. A. C., Cioffi, M., and Moreira-Filho, O. (2015). "Direct Chromosome Preparation from Freshwater Teleost Fishes," in *Fish Cytogenet. Tech. (Chondrichthyans Teleosts)*. Editors C. Ozouf-Costaz, E. Pisano, F. For, and L. F Almeida Toledo, 31–36. doi:10.1201/b18534-6

Betancur-R., R., Arcila, D., Vari, R. P., Hughes, L. C., Oliveira, C., Sabaj, M. H., et al. (2019). Phylogenomic Incongruence, Hypothesis Testing, and Taxonomic Sampling: The Monophyly of Characiform Fishes. *Evolution* 73, 329–345. doi:10.1111/evo.13649

Bolzán, A. D. (2017). Interstitial Telomeric Sequences in Vertebrate Chromosomes: Origin, Function, Instability and Evolution. *Mutat. Research/Reviews Mutat. Res.* 773, 51–65. doi:10.1016/j.mrrev.2017.04.002

Calcagnotto, D., Schaefer, S. A., and DeSalle, R. (2005). Relationships Among Characiform Fishes Inferred from Analysis of Nuclear and Mitochondrial Gene Sequences. *Mol. Phylogenet. Evol.* 36, 135–153. doi:10.1016/j.ympev.2005.01.004

Carvalho, P. C., de Oliveira, E. A., Bertollo, L. A. C., Yano, C. F., Oliveira, C., Decru, E., et al. (2017). First Chromosomal Analysis in Hepsetidae (Actinopterygii, Characiformes): Insights into Relationship between African and Neotropical Fish Groups. *Front. Genet.* 8, 203. doi:10.3389/fgene.2017.00203

Cioffi, M. B., and Bertollo, L. A. C. (2012). Chromosomal Distribution and Evolution of Repetitive DNAs in Fish. *Repetitive DNA* 7, 197–221. doi:10.1159/000337950

Cioffi, M. B., Martins, C., Centofante, L., Jacobina, U., and Bertollo, L. A. C. (2009). Chromosomal Variability Among Allopatric Populations of Erythrinidae Fish *Hoplias malabaricus*: Mapping of Three Classes of Repetitive DNAs. *Cytogenet. Genome Res.* 125, 132–141. doi:10.1159/000227838

Cioffi, M. d. B., Yano, C. F., Sember, A., and Bertollo, L. A. C. (2017). Chromosomal Evolution in Lower Vertebrates: Sex Chromosomes in Neotropical Fishes. *Genes* 8, 258. doi:10.3390/genes8100258

Cioffi, M. de. B., Ráb, P., Ezaz, T., Bertollo, L., Lavoué, S., Oliveira, E., et al. (2019). Deciphering the Evolutionary History of Arowana Fishes (Teleostei, Osteoglossiformes, Osteoglossidae): Insight from Comparative Cytogenomics. *Int J Mol Sci* 20, 4296. doi:10.3390/ijms20174296

de Freitas, N. L., Al-Rikabi, A. B. H., Bertollo, L. A. C., Ezaz, T., Yano, C. F., de Oliveira, E. A., et al. (2018). Early Stages of XY Sex Chromosomes Differentiation in the Fish *Hoplias malabaricus* (Characiformes, Erythrinidae) Revealed by DNA Repeats Accumulation. *Curr Genomics.* 19, 216–226. doi:10.2174/1389202918666170711160528

de Moraes, R. L. R., Bertollo, L. A. C., Marinho, M. M. F., Yano, C. F., Hatanaka, T., Barby, F. F., et al. (2017). Evolutionary Relationships and Cytotaxonomy Considerations in the GenusPyrrhulina(Characiformes, Lebiasinidae). *Zebrafish* 14, 536–546. doi:10.1089/zeb.2017.1465

de Moraes, R. L. R., Sember, A., Bertollo, L. A. C., De Oliveira, E. A., Ráb, P., Hatanaka, T., et al. (2019). Comparative Cytogenetics and Neo-Y Formation in Small-Sized Fish Species of the Genus Pyrrhulina (Characiformes, Lebiasinidae). *Front. Genet.* 10, 678. doi:10.3389/fgene.2019.00678

de Souza e Sousa, J. F., Viana, P. F., Bertollo, L. A. C., Cioffi, M. B., and Feldberg, E. (2017). Evolutionary Relationships Among Boulengerella Species (Ctenoluciidae, Characiformes): Genomic Organization of Repetitive DNAs and Highly Conserved Karyotypes. *Cytogenet. Genome Res.* 152, 194–203. doi:10.1159/000480141

Ditcharoen, S., Sassi, F. d. M. C., Bertollo, L. A. C., Molina, W. F., Liehr, T., Saenjundaeng, P., et al. (2020). Comparative Chromosomal Mapping of Microsatellite Repeats Reveals Divergent Patterns of Accumulation in 12 Siluridae (Teleostei: Siluriformes) Species. *Genet. Mol. Biol.* 43. doi:10.1590/1678-4685-GMB-2020-0091

Ijdo, J. W., Wells, R. A., Baldini, A., and Reeders, S. T. (1991). Improved Telomere Detection Using a Telomere Repeat Probe (TTAGGG)ngenerated by PCR. *Nucl. Acids Res.* 19, 4780. doi:10.1093/nar/19.17.4780

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, e45519. doi:10.1371/journal.pone.0045519

Kubat, Z., Hobza, R., Vyskot, B., and Kejnovsky, E. (2008). Microsatellite Accumulation on the Y Chromosome inSilene Latifolia. *Genome* 51, 350–356. doi:10.1139/G08-024

Levan, A., Fredga, K., and Sandberg, A. A. (1964). Nomenclature for Centromeric Position on Chromosomes. *Hereditas* 52, 201–220. doi:10.1111/j.1601-5223.1964.tb01953.x

Melo, B. F., Sidlauskas, B. L., Near, T. J., Roxo, F. F., Ghezelayagh, A., Ochoa, L. E., et al. (2021). Accelerated Diversification Explains the Exceptional Species Richness of Tropical Characoid Fishes. *Syst. Biol.*, 1–21. doi:10.1093/sysbio/syab040

Nelson, J. S., Grande, T. C., and Wilson, M. V. H. (2016). *Fishes of the World*. Hoboken, NJ: John Wiley & Sons. doi:10.2307/2412830

Netto-Ferreira, A. L., and Marinho, M. M. F. (2013). New Species of Pyrrhulina (Ostariophysi: Characiformes: Lebiasinidae) from the Brazilian Shield, with Comments on a Putative Monophyletic Group of Species in the Genus. *Zootaxa* 3664, 369–376. doi:10.11646/zootaxa.3664.3.7

Oliveira, C., Andreata, A. A., Toledo, L. F. A., and Toledo, S. A. (1991). Karyotype and Nucleolus Organizer Regions of Pyrrhulina Cf Australis (Pisces, Characiformes, Lebiasinidae). *Rev. Bras. Genética*, 14, 685–690.

Oliveira, C., Avelino, G. S., Abe, K. T., Mariguela, T. C., Benine, R. C., Ortí, G., et al. (2011). Phylogenetic Relationships within the Speciose Family Characidae (Teleostei: Ostariophysi: Characiformes) Based on Multilocus Analysis and Extensive Ingroup Sampling. *BMC Evol. Biol.* 11, 1–25. doi:10.1186/1471-2148-11-275

Pendás, A. M., Moran, P., Freije, J. P., and Garcia-Vazquez, E. (1994). Chromosomal Mapping and Nucleotide Sequence of Two Tandem Repeats of Atlantic salmon 5S rDNA. *Cytogenet. Cel Genet* 67, 31–36. doi:10.1159/000133792

Poltronieri, J., Marquioni, V., Bertollo, L. A. C., Kejnovsky, E., Molina, W. F., Liehr, T., et al. (2014). Comparative Chromosomal Mapping of Microsatellites inLeporinusSpecies (Characiformes, Anostomidae): Unequal Accumulation on the W Chromosomes. *Cytogenet. Genome Res.* 142, 40–45. doi:10.1159/000355908

R. Fricke, W. N. Eschmeyer, and R. van der Laan (Editors) (2021). *Eschmeyer's Catalog of Fishes: Genera, Species, References*. Available at: http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp (Accessed September 2, 2021).

Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning-Sambrook & Russel-Vol. 1, 2, 3*. Cold Spring Harbor, NY: Cold Springs Harb. Lab. Press.

Sassi, F. de. M. C., Oliveira, E., Bertollo, L., Nirchio, M., Hatanaka, T., Marinho, M., et al. (2019). Chromosomal Evolution and Evolutionary Relationships of Lebiasina Species (Characiformes, Lebiasinidae). *Int J Mol Sci* 20, 2944. doi:10.3390/ijms20122944

Sassi, F. d. M. C., Hatanaka, T., Moraes, R. L. R. d., Toma, G. A., Oliveira, E. A. d., Liehr, T., et al. (2020). An Insight into the Chromosomal Evolution of Lebiasinidae (Teleostei, Characiformes). *Genes* 11, 365. doi:10.3390/genes11040365

Scheel, J. J. (1973). Fish Chromosomes and Their Evolution. *Intern. Rep. Danmarks Akvar.* 22.

Sember, A., Bertollo, L. A. C., Ráb, P., Yano, C. F., Hatanaka, T., de Oliveira, E. A., et al. (2018). Sex Chromosome Evolution and Genomic Divergence in the Fish *Hoplias malabaricus* (Characiformes, Erythrinidae). *Front. Genet.* 9, 71. doi:10.3389/fgene.2018.00071

Sember, A., de Oliveira, E. A., Ráb, P., Bertollo, L. A. C., de Freitas, N. L., Viana, P. F., et al. (2020). Centric Fusions behind the Karyotype Evolution of Neotropical Nannostomus Pencilfishes (Characiforme, Lebiasinidae): First Insights from a Molecular Cytogenetic Perspective. *Genes* 11, 91. doi:10.3390/genes11010091

Sember, A., Nguyen, P., Perez, M. F., Altmanová, M., Ráb, P., and Cioffi, M. d. B. (2021). Multiple Sex Chromosomes in Teleost Fishes from a Cytogenetic Perspective: State of the Art and Future Challenges. *Phil. Trans. R. Soc. B* 376, 20200098. doi:10.1098/rstb.2020.0098

Souza, J., Guimarães, E., Pinheiro-Figliuolo, V., Cioffi, M. B., Bertollo, L. A. C., and Feldberg, E. (2021). Chromosomal Analysis of Ctenolucius Hujeta Valenciennes, 1850 (Characiformes): A New Piece in the Chromosomal Evolution of the Ctenoluciidae. *Cytogenet. Genome Res.* 161, 195–202. doi:10.1159/000515456

Sumner, A. T. (1972). A Simple Technique for Demonstrating Centromeric Heterochromatin. *Exp. Cel Res.* 75, 304–306. doi:10.1016/0014-4827(72)90558-7

Symonová, R., Majtánová, Z., Sember, A., Staaks, G. B., Bohlen, J., Freyhof, J., et al. (2013). Genome Differentiation in a Species Pair of Coregonine Fishes: an Extremely Rapid Speciation Driven by Stress-Activated Retrotransposons Mediating Extensive Ribosomal DNA Multiplications. *BMC Evol. Biol.* 13, 42–11. doi:10.1186/1471-2148-13-42

Terencio, M. L., Schneider, C. H., Gross, M. C., Nogaroto, V., de Almeida, M. C., Artoni, R. F., et al. (2012). Repetitive Sequences Associated with Differentiation of W Chromosome in Semaprochilodus Taeniurus. *Genetica* 140, 505–512. doi:10.1007/s10709-013-9699-4

Toma, G. A., de Moraes, R. L. R., Sassi, F. d. M. C., Bertollo, L. A. C., de Oliveira, E. A., Rab, P., et al. (2019). Cytogenetics of the Small-Sized Fish, Copeina Guttata (Characiformes, Lebiasinidae): Novel Insights into the Karyotype Differentiation of the Family. *PLoS One* 14, e0226746. doi:10.1371/journal.pone.0226746

Weitzman, M., and Weitzman, S. H. (2003). Family Lebiasinidae. *Check List Freshw. fishes South Cent. Am. Porto Alegre, Edipucrs* 729p, 241–250.

Yano, C. F., Bertollo, L. A. C., and Cioffi, M. D. B. (2017). "Fish-FISH: Molecular Cytogenetics in Fish Species," in *Fluorescence* In Situ

Hybridization (FISH). Berlin, Germany: Springer, 429–443. doi:10.1007/978-3-662-52959-1_44

Yano, C. F., Bertollo, L. A. C., Molina, W. F., Liehr, T., and de Bello Cioffi, M. (2014). Genomic Organization of Repetitive DNAs and its Implications for Male Karyotype and the Neo-Y Chromosome Differentiation in Erythrinus Erythrinus (Characiformes, Erythrinidae). *Comp Cytogenet.* 8, 139–151. doi:10.3897/compcytogen.v8i2.7597

Zwick, M. S., Hanson, R. E., Islam-Faridi, M. N., Stelly, D. M., Wing, R. A., Price, H. J., et al. (1997). A Rapid Procedure for the Isolation of C0t-1 DNA from Plants. *Genome* 40, 138–142. doi:10.1139/g97-020

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 25 October 2021
doi: 10.3389/fgene.2021.768056

Germline-Restricted Chromosome (GRC) in Female and Male Meiosis of the Great Tit (Parus major, Linnaeus, 1758)

Anna Torgasheva[1,2†], *Lyubov Malinovskaya*[1,2†], *Kira Zadesenets*[1], *Elena Shnaider*[3], *Nikolai Rubtsov*[1,2] *and Pavel Borodin*[1,2*]

[1]Institute of Cytology and Genetics, Russian Academy of Sciences, Siberian Branch, Novosibirsk, Russia, [2]Novosibirsk State University, Novosibirsk, Russia, [3]Bird of Prey Rehabilitation Centre, Novosibirsk, Russia

All songbirds studied so far have a germline-restricted chromosome (GRC), which is present in the germ cells and absent in the somatic cells. It shows a wide variation in size, morphology, and genetic content between the songbird species. In this paper, we analyzed GRC behavior in female and male meiosis of the great tit, using immunolocalization of meiotic proteins and FISH with GRC-derived DNA probes. We found that, despite dozens of million years of independent evolution, the great tit GRC displays a striking similarity with the GRCs of two species of martins and two species of estrildid finches examined earlier. It was usually present in two copies in females forming recombining bivalent and in one copy in males forming a condensed heterochromatic body with dotted-like axial elements of the synaptonemal complex. We observed mosaicism for the GRC copy number in the female and male great tit. This indicates that one of the GRC copies might be passively lost during premeiotic germ cell divisions. After the meiotic prophase, the GRC was ejected from most male germ cells. The reverse and interspecies FISH with GRC-specific microdissected DNA probes indicates that GRCs of the great tit, pale martin, and zebra finch differ substantially in their genetic content despite similarities in the meiotic behavior.

Keywords: avian chromosomes, programmed DNA elimination, recombination, synaptonemal complex, MLH1, SYCP3, crossing over

OPEN ACCESS

Edited by:
Ricardo Utsunomia, Federal Rural University of Rio de Janeiro, Brazil

Reviewed by:
María Inés Pigozzi, CONICET-Universidad de Buenos Aires, Argentina
Adél Sepsi, Hungarian Academy of Sciences (MTA), Hungary

***Correspondence:**
Pavel Borodin
borodin@bionet.nsc.ru

†These authors share first authorship

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *31 August 2021*
Accepted: *28 September 2021*
Published: *25 October 2021*

Citation:
Torgasheva A, Malinovskaya L, Zadesenets K, Shnaider E, Rubtsov N and Borodin P (2021) Germline-Restricted Chromosome (GRC) in Female and Male Meiosis of the Great Tit (Parus major, Linnaeus, 1758). Front. Genet. 12:768056. doi: 10.3389/fgene.2021.768056

INTRODUCTION

Programmed DNA elimination has been observed in many species of different taxa (Wang and Davis, 2014). One of the most recent examples is the germline-restricted chromosome (GRC) of the songbirds, which is present in the germline and absent in the somatic cells (Pigozzi and Solari, 1998, Pigozzi and Solari, 2005; del Priore et al., 2014; Kinsella et al., 2019; Torgasheva et al., 2019; Malinovskaya et al., 2020). In the male germ cells, it is usually present in one copy. It is heterochromatic, highly enriched in histone H3 trimethylated at lysine 9, and ejected from the nuclei after the meiotic divisions (del Priore et al., 2014). In the female germline, the GRC is usually present in two copies, which pair and recombine with each other and are transmitted to the progeny (Pigozzi and Solari, 2005; Torgasheva et al., 2019; Malinovskaya et al., 2020).

There is a variation in GRC copy number in three species examined: zebra finch, sand martin, and pale martin. Most oocytes of zebra finches and martins contained two copies of GRC, but some

specimens had one copy (Pigozzi and Solari, 2005; Torgasheva et al., 2019; Malinovskaya et al., 2020). Male pale martins demonstrated mosaicism for the number of GRCs in the spermatocytes. Most cells contained one copy, but the cells with two and three copies were also detected (Malinovskaya et al., 2020). More species have to be examined to estimate an abundance and possible causes of the GRC polymorphism and mosaicism.

The GRC shows a wide variation in size, morphology, and genetic content between the songbird species. In most species, it is a large macrochromosome. In other species, it is a microchromosome (Torgasheva et al., 2019). Cross-species fluorescent *in situ* hybridization (FISH) with GRC-derived DNA probes revealed low, if any, homology between GRCs of distantly related species (Torgasheva et al., 2019). The intraspecies variation in the number of GRCs in the germ cells and its interspecies variation in size and genetic content are intriguing because GRC appears to be an indispensable component of the germ cells. Detailed analysis of zebra finch GRC revealed that it contains dozens of genes actively transcribed in the germ cells. Some of them show signs of positive selection (Biederman et al., 2018; Kinsella et al., 2019).

In this paper, using immunolocalization of several meiotic proteins, we examined GRC behavior in female and male meiosis of the great tit (*Parus major* Linnaeus, 1758) and compared it with that of two estrildid finches (zebra and Bengalese) (Pigozzi and Solari, 1998, Pigozzi and Solari, 2005; del Priore et al., 2014) and two martins (sand and pale) (Malinovskaya et al., 2020). We also estimated a homology between the GRC of these species using cross-species FISH.

MATERIALS AND METHODS

Experimental Model

We examined seven adult males and seven female nestlings collected in two mixed forest parks in Novosibirsk (54,50N; 83,05E and 55.09N, 82.95E). The males were captured with bird nets at the beginning of the breeding season. Nestling females on days 3–6 after hatching were collected from the nests.

Conventional Chromosome Spreading and Staining

Mitotic chromosome spreads were prepared from short-term bone marrow cell cultures incubated in Dulbecco's modified Eagle's medium (ThermoFisher Scientific, cat# 41965062) with 10 μg/ml colchicine (Sigma, cat# C9754) for 2 h at 37°C. The cells were swollen in 0.56% KCl, fixed in fresh Carnoy's solution (methanol:glacial acetic acid, 3:1). The cell suspension was dropped on clean, cold, wet microscope slides and spread by air-drying. Chromosomes were stained with DAPI dissolved in Vectashield antifade solution (Vector Laboratories, cat# H-1200-10, United States).

Meiotic chromosome spreads were prepared from a suspension of testicular cells of adult males treated with hypotonic solution (0.88% KCl) for 3 h at 37°C and then with Carnoy's solution. The cell suspension was dropped onto clean, cold, wet coverslips (60 × 24 mm, 0.17 mm thick), and dried. Chromosomes were stained with 0.1% Giemsa solution (Biovitrum, cat# 20-043\L).

Synaptonemal Complex Spreading and Immunostaining

Chromosome spreads for SC analysis were prepared from testes and ovaries by the drying down method (Peters et al., 1997). Testes and ovaries were dissected and placed for 30–60 min in an extraction buffer containing 30 mM Tris (Helicon, cat# 77-86-1, Russia), 50 mM sucrose (Sigma, cat# S7903-1KG), 17 mM trisodium citrate dehydrate (Chimmed, cat# A1227436-500.0, Russia), 5 mM EDTA (Panreac&AppliChem, cat# A5097), and pH 8.2. Then, small pieces of testis or the whole ovary were macerated in 40 μl of 100 mM of sucrose, pH 8.2, on a glass slide. A fine suspension was made, and 20 μl of the suspension was gently dropped at the slide moistened by 1% paraformaldehyde (Sigma-Aldrich, cat# 158127) solution, pH 9.2, containing 0.15% Triton X-100 (Sigma, cat# T8787). The slides were dried for 2 h, washed in 0.4% Kodak Photo-Flo 200 (Kodak, cat# 742057), and dried at room temperature.

Immunostaining was performed according to Anderson et al. (1999) using rabbit polyclonal anti-SYCP3 (1:500; Abcam, cat# ab15093), mouse monoclonal anti-MLH1 (1:50; Abcam, cat# ab14206), human anticentromere (ACA) (1:100; Antibodies Inc., cat# 15–234), and rabbit polyclonal anti-H3K9me3 (1:50; Abcam, cat# ab8898) primary antibodies. The secondary antibodies used were Cy3-conjugated goat antirabbit (1:500; Jackson ImmunoResearch, cat# 111-165-144), FITC-conjugated goat antimouse (1:50; Jackson ImmunoResearch, cat# 115-095-003), and AMCA-conjugated donkey antihuman (1:100; Jackson ImmunoResearch, cat# 709-155-149). Antibodies were diluted in PBT [3% bovine serum albumin and 0.05% Tween 20 in phosphate-buffered saline (PBS)]. Slides were incubated in a solution of 10% PBT for 40 min to reduce the unspecific binding of the antibodies. Primary antibody incubations were performed overnight in a humid chamber at 37°C and secondary antibody incubations for 1 h at 37°C. After antibody incubations, slides were washed three times in PBST (PBS with 0.1% Tween 20) for 10 min. Slides were mounted in Vectashield antifade mounting medium (Vector Laboratories, cat# H-1000-10, United States).

Generation of the Microdissected DNA Probe and FISH

The DNA probe for the great tit GRC was prepared by microdissection of 15 round Giemsa-positive bodies located near spermatocytes and spermatids at the conventionally prepared meiotic chromosome spreads. According to Pigozzi and Solari (1998) they contain GRC ejected from the germ cells. DNA isolated from the microdissected material was amplified with the GenomePlex Single Cell Whole Genome Amplification

Kit (Sigma-Aldrich, cat# WGA4) according to the manufacturer's protocol (Zadesenets et al., 2017). The obtained PCR products were labeled with Flu-dUTP (Genetyx, cat# N-801000, Russia) in additional PCR cycles. DNA probes for zebra finch and pale martin GRCs were prepared as described earlier (Torgasheva et al., 2019). C_0t-1 DNA isolation from pale martin was performed as described earlier (Trifonov et al., 2017).

FISH with the DNA probes on the SC spreads were performed according to a standard protocol with salmon sperm DNA (Ambion, cat# AM9680, United States) as a DNA carrier (Trifonov et al., 2017). Chromosomes were counterstained with DAPI dissolved in Vectashield antifade solution.

Microscopic Analysis

Images of DAPI-stained metaphase chromosomes and SC spreads after immunostaining and FISH were captured using a CCD camera installed on an Axioplan 2 compound microscope (Carl Zeiss, Germany) equipped with filter cubes #49, #10, and #15 (ZEISS, Germany) using ISIS4 (METASystems GmbH, Germany).

Chromosome Measurements and Generation of Recombination Maps of GRCs

Centromeres were identified by ACA foci. MLH1 signals were only scored if they were localized on SCs. The length of the SC was measured in micrometers, and the positions of MLH1 foci in relation to the centromere were recorded using MicroMeasure 3.3 (Reeves, 2001). SCs of GRC and macrochromosomes were identified by their relative lengths and centromeric indices. STATISTICA 6.0 software package (StatSoft, Tulsa, OK, United States) was used for descriptive statistics. All results were expressed as mean ± SD.

RESULTS

Pachytene Karyotype of the Great Tit

A diploid chromosome number (2n) in the somatic cells of the great tit was 80 and corresponded to that previously described (Nanda et al., 2011). Macrochromosome 1 was metacentric; 2 was subacrocentric; 3, 4, 5, and 6 were submetacentric; and all other chromosomes but five were telocentric, forming a row gradually decreasing in length. Z chromosome was a submetacentric macrochromosome of intermediate size. The W chromosome was a large microchromosome. Pachytene karyotype contained six macrobivalents, 33 microbivalents, the sex bivalent (ZZ in males, ZW in females), and an additional bivalent or univalent of a large acrocentric chromosome, which was not present on the bone marrow spreads (**Figure 1**). We identify this chromosome as a GRC.

Mosaicism for GRC Copy Number in Females

We detected mosaicism for the number of GRC copies in pachytene oocytes (**Supplementary Table 1**). Three out of seven females contained two copies of GRC in all examined cells at the pachytene stage (**Figure 1B**). Four females were mosaic for GRC copy number. Most of their oocytes contained two GRCs. The proportion of cells with one GRC varied from 2 to 26% (**Supplementary Table 1**). The GRC bivalents appeared as normal autosomal bivalents with at least one MLH1 focus (**Figure 1B**) and were distinguished as the only acrocentrics. The GRC univalents were distinguished from the bivalents by a lack of MLH1 signals and less intense labeling with antibodies to SYCP3 (**Figure 1C**). They were significantly longer than the bivalents (19.4 ± 3.0 and 14.1 ± 5.4 μm, respectively; Mann–Whitney test, $p = 0.02$).

Synapsis and Recombination of GRC in Females

GRC bivalents differed substantially in average number and distribution of MLH1 foci from the macrobivalents of comparable size (SC1: 17.2 ± 3.7 μm and SC2: 14.7 ± 3.1 μm). Most of the GRC bivalents contained one MLH1 signal. It was always located near the centromere (**Supplementary Image 1**). The bivalents with two or three MLH1 signals were rare (4.0 and 0.4%, respectively). The average number of MLH1 signals per GRC bivalent was 1.05 ± 0.2. The macrobivalents 1 and 2 contained a significantly higher number of MLH1 foci (3.5 ± 0.9 and 2.7 ± 0.8, respectively, Mann–Whitney test, $p = 0.00$), which showed a rather even distribution with peaks near the bivalent ends (**Supplementary Figure S1**).

Mosaicism for GRC Copy Number in Males

In total, we analyzed 612 spermatocytes from seven males of the great tit. In all analyzed cells, GRC occurred as a condensed body extensively labeled by antibodies to the centromere proteins. SYCP3 signal was only observed near the proximal end of GRC or its both ends as single or double dots or short lines (**Figure 1D**).

To estimate the copy number variation of GRC at different stages of spermatogenesis, we used antibodies against histone H3 trimethylated at lysine 9 (H3K9me3). It is a marker of heterochromatic transcriptionally repressed regions (Nicetto and Zaret, 2019). The zebra finch GRC is enriched in H3K9me3 compared with the basic chromosome set during prophase I and after the elimination (del Priore et al., 2014).

Most male germline cells of the great tit contain one strong H3K9me3 signal marking the GRC (**Figure 2A**). In 15 cells out of the 411 examined (3.6%) we detected two H3K9me3 signals (**Figure 2B**). Spermatids and spermatozoa usually do not show the H3K9me3 signals. Near some of these cells, we detected condensed chromatin bodies with strong H3K9me3 signal (**Figure 2C**). Apparently, they were GRCs ejected from the cells. Surprisingly, we found the H3K9me3 signal in a few spermatozoa (3 out of 880 cells examined) (**Figure 2D**).

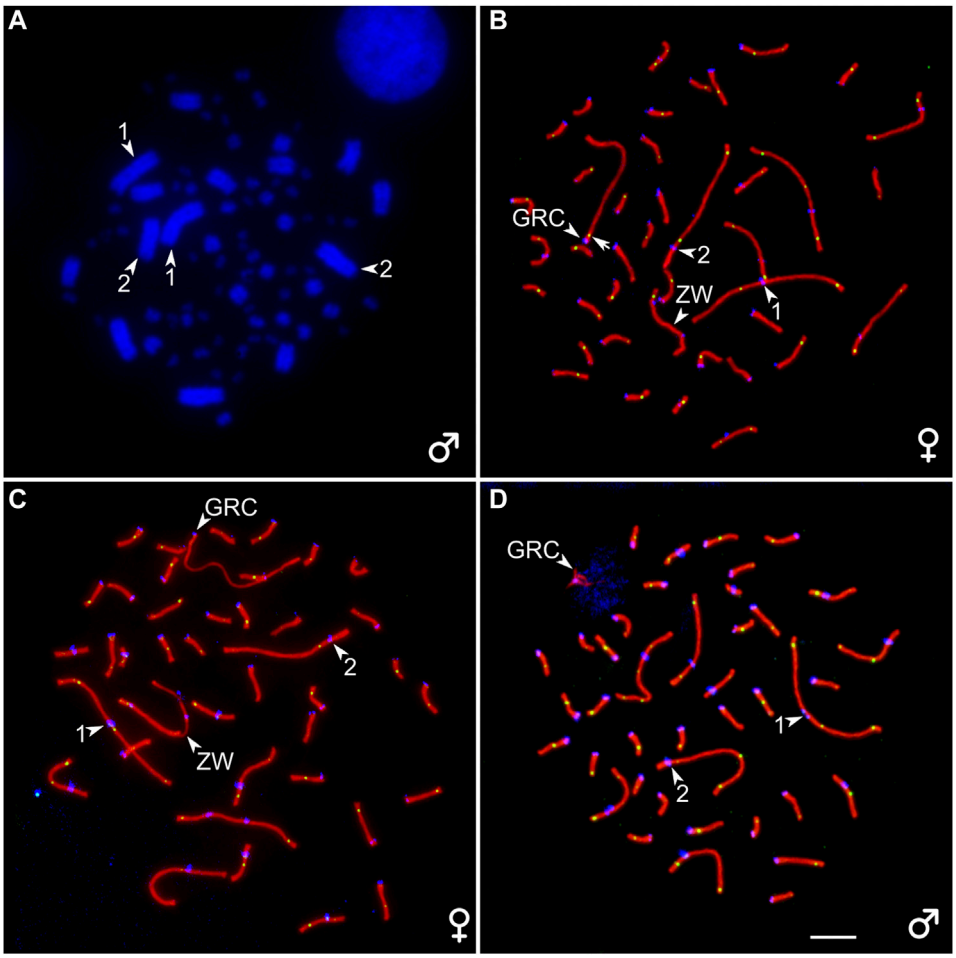

FIGURE 1 | (A) DAPI-stained bone marrow cell of the male great tit. **(B–D)** Pachytene oocytes **(B, C)** and spermatocyte **(D)** of the great tit with two **(B)** and one **(C, D)** copies of GRC after immunostaining with antibodies to SYCP3 (red), centromere proteins (blue), and MLH1 (green). The arrowheads point to centromeres of the two largest macrobivalents and ZW bivalent (identified by heteromorphic SC and unaligned centromeres) and GRC. The arrow points to the MLH1 signal at GRC bivalent **(B)**. Bar–5 μm.

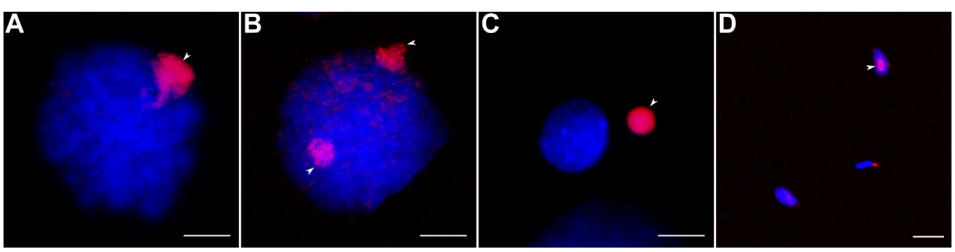

FIGURE 2 | Male germ cells after H3K9me3 labeling (red) and DAPI staining (blue). **(A, B)** Cells with one **(A)** and two **(B)** GRC copies. **(C)** Post-meiotic cell and eliminated GRC. **(D)** Spermatozoa with and without GRC. Arrowheads point to GRCs. Bar—10 μm.

FISH With the GRC-Specific DNA Probe

To estimate a homology between GRC and the chromosomes of the basic set of the great tit, we performed reverse FISH with the GRC DNA probe. It produced a strong specific signal on the condensed GRC in pachytene spermatocytes (**Figure 3A**). It also faintly labeled different regions of chromosomes of the basic set. In pachytene oocytes, FISH with suppression of repeated sequences using C_0t-1 DNA produced a strong specific signal on the whole GRC except short pericentromeric region (**Figure 3B**), where most MLH1 foci were located

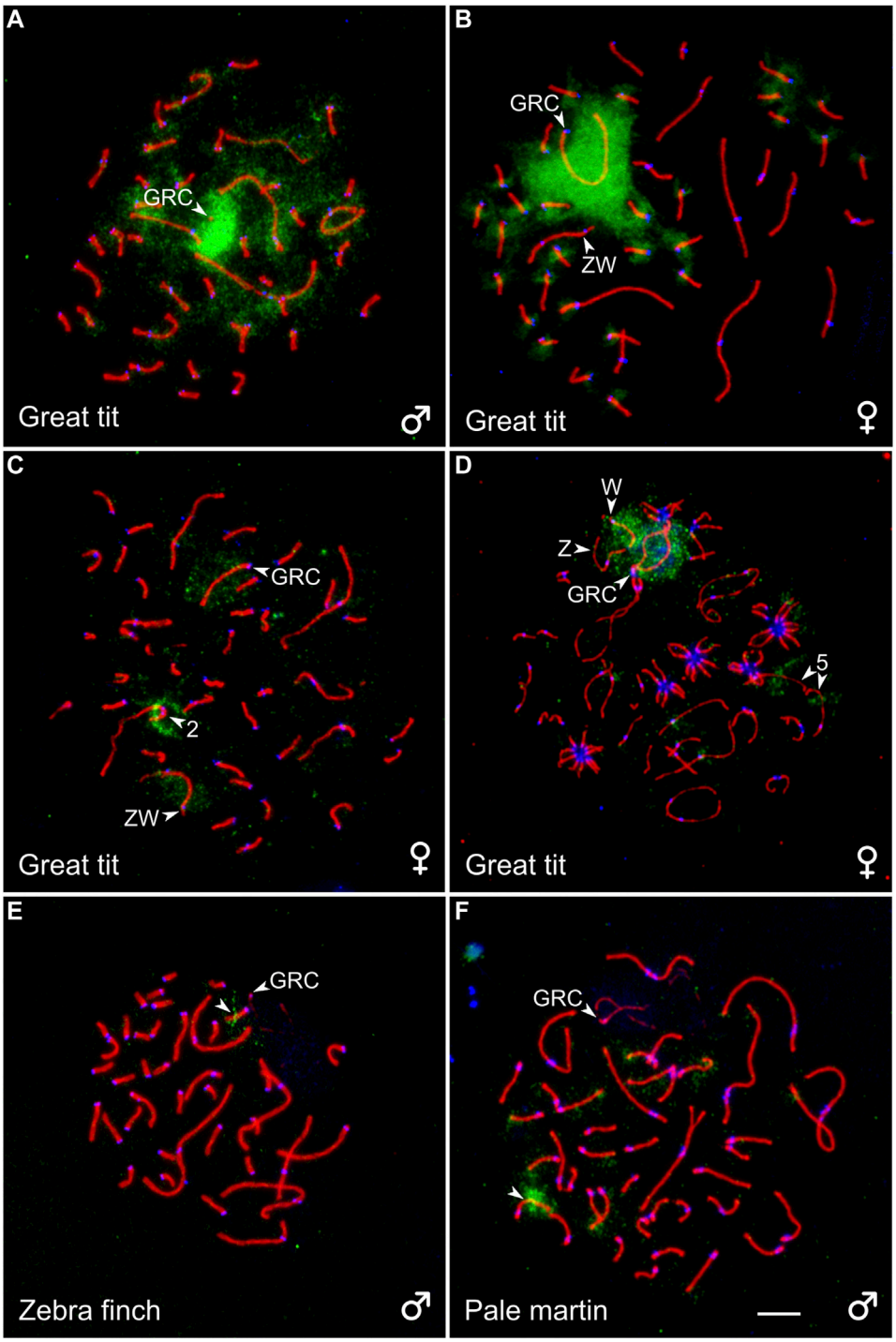

FIGURE 3 | SC spreads of the great tit **(A–D)**, zebra finch **(E)**, and pale martin **(F)** after reverse FISH with the great tit GRC probe (green) without **(A)** and with **(B)** suppression of repetitive sequences using C_0t-1 DNA of the great tit and cross-species FISH with DNA probes (green) derived from GRC of the zebra finch **(C)**, pale martin **(D)**, and great tit **(E, F)**. Spreads were immunolabeled with antibodies against SYCP3 (red) and centromere proteins (blue). The arrowheads point to macrobivalents 2 and 5, sex bivalent ZW, GRC and microbivalents with hybridization signals. Bar—5 μm.

(**Supplementary Image 1**). It also labeled pericentromeric regions of most microchromosomes (**Figure 3B**). We did not detect GRC probe signals in post-meiotic cells.

To estimate a homology between GRC of different species, we carried out cross-species FISH of the zebra finch and pale martin GRC DNA probes with the great tit oocytes and the great tit GRC

DNA probe with the zebra finch and pale martin spermatocytes (**Figures 3C–F**). The zebra finch and pale martin GRC probes produced a clear hybridization signal at the GRC of the great tit. They also faintly labeled its W chromosome (**Figures 3C,D**). Additionally, the zebra finch GRC DNA probe produced a distinct signal at the short arm of the SC2 (**Figure 3C**) and the pale martin GRC DNA probe in the middle of the long arm of the SC5 (**Figure 3D**). The great tit GRC probe gave a strong signal at the middle of one zebra finch microbivalent and the distal half of one pale martin microbivalent (**Figures 3E,F**). We also found the signals of the great tit GRC probe at telomeres of some macro- and microbivalents of pale martin. We did not detect signals of great tit GRC probe at zebra finch and pale martin GRCs.

DISCUSSION

The estimated times of divergence of the great tit from estrildid finches and martins are approximately the same: 38 MYA CI: (34–42 MYA) and 40 MYA CI: (37–43 MYA), respectively (Claramunt and Cracraft, 2015). Our results of cross-species FISH with GRC DNA probes of these species indicate that GRCs of each pair of these three phylogenetically equidistant species of songbirds still share some common (probably repetitive) sequences. In reverse FISH, the great tit GRC probe labeled pericentromeric regions of microchromosomes. Torgasheva et al. (2019) observed a similar effect in the reverse FISH experiment in the pale martin. This indicates GRCs of these species are enriched with repeated sequences characteristic to pericentromeric regions. However low intensity of the cross-species hybridization signals and lack of hybridization signal of great tit GRC probe at zebra finch and pale martin GRCs suggest that GRCs of the zebra finch, pale martin, and great tit have already undergone substantial genetic divergence. The distribution of GRC probe FISH signals on chromosomes of the basic set also confirms significant differences in genetic content between GRCs of these species.

Yet, despite dozens of million years of independent evolution and a substantial divergence in the genetic content, GRCs of the great tit, estrildid finches, and martins are very similar in their morphology and meiotic behavior. They are large acrocentric macrochromosomes of approximately the same size (Pigozzi and Solari, 1998; del Priore et al., 2014; Malinovskaya et al., 2020). The great tit, estrildid finches, and martins show the same sexual dimorphism in the GRC copy number. Most males examined had one GRC in spermatocytes; all females had two GRCs in the majority of their oocytes (Pigozzi and Solari, 2005; Malinovskaya et al., 2020).

Recombination in the GRC bivalents of the great tit, zebra finch, and pale and sand martins is strongly suppressed everywhere beyond the chromosome ends. The only difference is that the GRC bivalents of the great tit usually contain a single recombination nodule located in their pericentromeric region, and the GRCs of the finches and martins have two nodules at their both ends (Pigozzi and Solari, 2005; Malinovskaya et al., 2020). Malinovskaya et al. (2020) suggest that the polarized recombination pattern along the GRC bivalents in the female songbirds could facilitate GRC non-disjunction during MI. It is shown that in human oocytes chiasmata located too close to centromere are responsible for a high rate of non-disjunction in female meiosis (Hassold and Hunt, 2001).

In all species examined, we observed sexual dimorphism in the appearance of GRC univalents. In female meiosis, it appears as a normal lateral element of SC, and in males its lateral element is usually thin and often fragmented (Pigozzi and Solari, 1998; del Priore et al., 2014; Torgasheva et al., 2019; Malinovskaya et al., 2020). The GRC univalents in great tit males show weak or no polymerization of the lateral elements at all. This might indicate a greater degree of GRC heterochromatinization in this species.

The frequency of polymorphism and mosaicism for GRC copy number in songbirds is difficult to estimate because it would demand large samples of birds and germ cells. The data obtained to date indicate that the mosaicism is rather frequent in the martin males, rare in both sexes of the great tit, and has not been detected yet in the estrildid finches. However, the mere existence of the polymorphism and mosaicism for GRC elucidates two important features of GRC.

Polymorphism indicates that at least one GRC copy is indispensable for the germ cell survival until the MI in males and to term in females because no cells without GRC were observed in any species examined (Torgasheva et al., 2019). At the same time, the cells with one and two copies apparently have the same chances to survive. Mosaicism in males indicates that the GRC is actively ejected from spermatocytes after the meiotic prophase, but any additional copy of GRC can be passively lost during the germ cell mitotic divisions. Mosaicism in females indicates a possibility of the same passive loss of the second GRC during premeiotic germ cell divisions. This is consistent with the model of GRC transmission proposed earlier (Malinovskaya et al., 2020).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Care and Use Committee of the Institute of Cytology and Genetics SB RAS (protocol # 45/2 of January 10, 2019 and #85 of June 15, 2021).

AUTHOR CONTRIBUTIONS

Conceptualization by PB, AT; methodology by NR; formal analysis by LM, AT; investigation by LM, KZ, and AT; resources by ES, KZ, and NR; writing and original draft preparation by LM, AT; writing, review and editing by AT, PB; visualization by LM, AT; supervision by PB; project administration by AT.

FUNDING

The study was funded by Russian Foundation of Basic Researches, grant # 19-34-90118 and by Ministry of Science

and Higher Education of the Russian Federation, grants # 0259-2021-0011 and # 2019-0546 (FSUS-2020-0040). Microscopy was carried out at the Core Facility for Microscopy of Biologic Objects, SB RAS, Novosibirsk, Russia (regulation no. 3054).

REFERENCES

Anderson, L. K., Reeves, A., Webb, L. M., and Ashley, T. (1999). Distribution of Crossing over on Mouse Synaptonemal Complexes Using Immunofluorescent Localization of MLH1 Protein. *Genetics* 151, 1569–1579. doi:10.1093/genetics/151.4.1569

Biederman, M. K., Nelson, M. M., Asalone, K. C., Pedersen, A. L., Saldanha, C. J., and Bracht, J. R. (2018). Discovery of the First Germline-Restricted Gene by Subtractive Transcriptomic Analysis in the Zebra Finch, Taeniopygia guttata. *Curr. Biol.* 28, 1620–1627. doi:10.1016/j.cub.2018.03.067

Claramunt, S., and Cracraft, J. (2015). A New Time Tree Reveals Earth History's Imprint on the Evolution of Modern Birds. *Sci. Adv.* 1, e1501005. doi:10.1126/sciadv.1501005

del Priore, L., and Pigozzi, M. I. (2014). Histone Modifications Related to Chromosome Silencing and Elimination during Male Meiosis in Bengalese Finch. *Chromosoma* 123, 293–302. doi:10.1007/s00412-014-0451-3

Hassold, T., and Hunt, P. (2001). To Err (Meiotically) Is Human: the Genesis of Human Aneuploidy. *Nat. Rev. Genet.* 2, 280–291. doi:10.1038/35066065

Kinsella, C. M., Ruiz-Ruano, F. J., Dion-Côté, A. M., Charles, A. J., Gossmann, T. I., Cabrero, J., et al. (2019). Programmed DNA Elimination of Germline Development Genes in Songbirds. *Nat. Commun.* 10, 5468. doi:10.1038/s41467-019-13427-4

Malinovskaya, L. P., Zadesenets, K. S., Karamysheva, T. V., Akberdina, E. A., Kizilova, E. A., Romanenko, M. V., et al. (2020). Germline-restricted Chromosome (GRC) in the Sand martin and the Pale martin (Hirundinidae, Aves): Synapsis, Recombination and Copy Number Variation. *Sci. Rep.* 10, 1058. doi:10.1038/s41598-020-58032-4

Nanda, I., Benisch, P., Fetting, D., Haaf, T., and Schmid, M. (2011). Synteny Conservation of Chicken Macrochromosomes 1-10 in Different Avian Lineages Revealed by Cross-Species Chromosome Painting. *Cytogenet. Genome. Res.* 132, 165–181. doi:10.1159/000322358

Nicetto, D., and Zaret, K. S. (2019). Role of H3K9me3 Heterochromatin in Cell Identity Establishment and Maintenance. *Curr. Opin. Genet. Dev.* 55, 1–10. doi:10.1016/j.gde.2019.04.013

Peters, A. H. F. M., Plug, A. W., vanVugt, M. J., and de Boer, P. (1997). A Drying-Down Technique for the Spreading of Mammalian Meiocytes from the Male and Female Germline. *Chromosom. Res.* 5, 66–68. doi:10.1023/a:1018445520117

Pigozzi, M. I., and Solari, A. J. (1998). Germ Cell Restriction and Regular Transmission of an Accessory Chromosome that Mimics a Sex Body in the Zebra Finch, Taeniopygia guttata. *Chromosom. Res.* 6, 105–113. doi:10.1023/A:1009234912307

Pigozzi, M. I., and Solari, A. J. (2005). The Germ-Line-Restricted Chromosome in the Zebra Finch: Recombination in Females and Elimination in Males. *Chromosoma* 114, 403–409. doi:10.1007/s00412-005-0025-5

Reeves, A. (2001). MicroMeasure: A New Computer Program for the Collection and Analysis of Cytogenetic Data. *Genome*, 44(3):439-443. doi:10.1139/g01-037

Torgasheva, A. A., Malinovskaya, L. P., Zadesenets, K. S., Karamysheva, T. V., Kizilova, E. A., Akberdina, E. A., et al. (2019). Germline-restricted Chromosome (GRC) Is Widespread Among Songbirds. *Proc. Natl. Acad. Sci. USA* 116, 17373–11650. doi:10.1073/pnas.1817373116

Trifonov, V. A., Vorobieva, N. V., Serdyukova, N. A., and Rens, W. (2017). "FISH with and without COT1 DNA," in *In Fluorescence* in Situ *Hybridization (FISH) - Application Guide*. Editor T. Liehr (Heidelberg, Germany: Springer Berlin), 123–133. doi:10.1007/978-3-662-52959-1_11

Wang, J., and Davis, R. E. (2014). Programmed DNA Elimination in Multicellular Organisms. *Curr. Opin. Genet. Dev.* 27, 26–34. doi:10.1016/j.gde.2014.03.012

Zadesenets, K. S., Schärer, L., and Rubtsov, N. B. (2017). New Insights into the Karyotype Evolution of the Free-Living Flatworm Macrostomum Lignano (Platyhelminthes, Turbellaria). *Sci. Rep.* 7, 6066. doi:10.1038/s41598-017-06498-0

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.768056/full#supplementary-material

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 September 2021
doi: 10.3389/fgene.2021.726211

Genome Size Doubling Arises From the Differential Repetitive DNA Dynamics in the Genus *Heloniopsis* (Melanthiaceae)

Jaume Pellicer[1,2], Pol Fernández[1], Michael F. Fay[2,3], Ester Michálková[4] and Ilia J. Leitch[2]*

[1] Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Barcelona, Spain, [2] Royal Botanic Gardens, Kew, Richmond, United Kingdom, [3] School of Plant Biology, University of Western Australia, Crawley, WA, Australia, [4] Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czechia

Plant genomes are highly diverse in size and repetitive DNA composition. In the absence of polyploidy, the dynamics of repetitive elements, which make up the bulk of the genome in many species, are the main drivers underpinning changes in genome size and the overall evolution of the genomic landscape. The advent of high-throughput sequencing technologies has enabled investigation of genome evolutionary dynamics beyond model plants to provide exciting new insights in species across the biodiversity of life. Here we analyze the evolution of repetitive DNA in two closely related species of *Heloniopsis* (Melanthiaceae), which despite having the same chromosome number differ nearly twofold in genome size [i.e., *H. umbellata* (1C = 4,680 Mb), and *H. koreana* (1C = 2,480 Mb)]. Low-coverage genome skimming and the RepeatExplorer2 pipeline were used to identify the main repeat families responsible for the significant differences in genome sizes. Patterns of repeat evolution were found to correlate with genome size with the main classes of transposable elements identified being twice as abundant in the larger genome of *H. umbellata* compared with *H. koreana*. In addition, among the satellite DNA families recovered, a single shared satellite (HeloSAT) was shown to have contributed significantly to the genome expansion of *H. umbellata*. Evolutionary changes in repetitive DNA composition and genome size indicate that the differences in genome size between these species have been underpinned by the activity of several distinct repeat lineages.

Keywords: C-value, DNA repeats, chromosome, transposable elements, satellite DNA

OPEN ACCESS

Edited by:
Magdalena Vaio, Universidad de la República, Uruguay

Reviewed by:
Tony Heitkam, Technische Universität Dresden, Germany
Ales Kovarik, Academy of Sciences of the Czech Republic (ASCR), Czechia

****Correspondence:***
Jaume Pellicer
jaume.pellicer@ibb.csic.es

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *16 June 2021*
Accepted: *19 August 2021*
Published: *06 September 2021*

Citation:
Pellicer J, Fernández P, Fay MF, Michálková E and Leitch IJ (2021) Genome Size Doubling Arises From the Differential Repetitive DNA Dynamics in the Genus Heloniopsis (Melanthiaceae). Front. Genet. 12:726211. doi: 10.3389/fgene.2021.726211

INTRODUCTION

Plant genomes are dynamic and can expand in size through a variety of processes such as the proliferation of repetitive elements (including transposable elements and tandem repeats), and whole genome duplications (Wang D. et al., 2021). In parallel, regulatory mechanisms (i.e., epigenetic modifications) can act to prevent repetitive sequences from uncontrolled expansion and these, together with various recombination-based processes which may eliminate DNA, can result in genome downsizing (Devos et al., 2002; Hawkins et al., 2009; Schubert and Vu, 2016; Vu et al., 2017; Wang X. et al., 2021). It is now widely recognized that it is the relative activity of each of these opposing evolutionary forces driving genome expansion or contraction that has underpinned the generation of the outstanding diversity of genome sizes in plants, especially in

angiosperms which vary c. 2,400-fold (Pellicer et al., 2018). Even at the genus level, genome size can vary by orders of magnitude, such as for example in *Cuscuta*, in which nuclear DNA content ranges c. 102-fold (Neumann et al., 2021). At the family level, Melanthiaceae stand out within monocots as being among the most diverse, with genome sizes varying > 230-fold (Pellicer et al., 2014). This is even more remarkable given that this family is made up of just c. 180 species, and similar levels of diversity have only been reported in much larger groups, such as the eudicot family Santalaceae (c. 1,000 species, 1C range: c. 395-fold).[1] The main driver underpinning the extensive genome size diversity in Melanthiaceae lies in a striking genome expansion that occurred during the diversification of tribe Parideae. This event is estimated to have taken place c. 57-31 million years ago (Kim et al., 2019), and resulted in the emergence of some of the largest genomes known to date (Pellicer et al., 2014). Besides, a thorough analysis at lower taxonomic levels beyond Parideae revealed an almost doubling of the nuclear DNA content between species in *Heloniopsis*, a small genus made up of six species, with 1C-values ranging from 2,480 Mb in *H. koreana* to 4,680 Mb in *H. umbellata* (ratio = c. 1.90). This raises the question as to what are the key mechanisms responsible for underpinning such genome size differences among closely related taxa?

Although polyploidy is known to be frequent in many angiosperms and has indeed been reported in some genera of Melanthiaceae (e.g., *Paris*, *Trillium* and *Veratrum*), extant species of *Heloniopsis* share a chromosome number of 2n = 34 (Pellicer et al., 2014). The authors consistently found the same chromosome number across the tribe Heloniadeae, despite chromosome-based modeling approaches inferring with high probability that ancient polyploid events coupled with chromosome losses to have happened during the evolution of the tribe. Should this reconstructed scenario hold true, then it would imply that potential chromosomal reorganizations had not resulted in changes in the overall chromosome number or ploidy level, given the relatively stable karyotype features reported for several species in the genus (Kokubugata et al., 2004).

Considering the evolutionary past of the tribe, the observed differences in genome size can most be likely attributed to the differential activity of repetitive DNA sequences and the associated recombination-based mechanisms in charge of their removal. Repetitive DNA in plants includes both dispersed mobile elements and tandem repeats (Bennetzen and Wang, 2014). DNA transposon and retrotransposon dynamics involve cut-and-paste and copy-and-paste insertion mechanisms, respectively, to spread across the genome, and are recognized as dispersed mobile elements. Of these, long terminal repeat (LTR) retrotransposons are widely known to monopolize a substantial fraction of plant genomes, and comprise several superfamilies, with Ty1/copia and Ty3/gypsy elements being the most common in plants (Wicker et al., 2007). Indeed, in many cases, their dominance leaves a secondary role for tandem repeats in shaping plant genome evolutionary dynamics. This is the case, in for example, *Hesperis* (Brassicaceaee), where most of the genome size variation observed is driven by the activity of a diverse array of LTR families (Hloušková et al., 2019).

As our understanding of how genomes are structured and function increases, it is becoming apparent that at the lower end of the genome size spectrum, changes in size are often seen to be driven by the activity of just a few lineages of transposable elements (Hawkins et al., 2006; Piegu et al., 2006; Macas et al., 2015), whereas plants with larger genomes (i.e., > c. 10 Gb/1C) have most likely arisen through the accumulation of elements over long periods of time, given their more heterogeneous composition (Nystedt et al., 2013; Kelly et al., 2015; Novák et al., 2020a). Despite the above-mentioned critical role of transposable elements in shaping many plant genomes, Ågren et al. (2015) emphasized the need to also recognize the importance of short simple sequence repeats (including simple repeats, satellite and low complexity DNA) in contributing to the evolution of genome size in some plant species, using the evening primroses (*Oenothera* species) to illustrate this. Altogether, research aiming to uncover what variables influence the dynamics of repetitive DNA sequences in non-model plants, why they accumulate in some lineages and not others, and what are the key sequences involved, is urgently needed to continue to further our understanding of the origins of the staggering genome size diversity across eukaryotes in general, and in plants in particular.

To contribute to these goals, we have carried out a comparative study of two species of the genus *Heloniopsis* with contrasting genome sizes. We have used next generation sequencing to characterize and assess the abundance of different types of DNA repeats and their role in contributing to changes in the composition and size of both genomes. We also explore whether the differences in genome size between these two species are underpinned by (i) differences in the amounts of just a few repetitive elements, as observed in other species with small genomes, and (ii), evaluate the differences across the repetitive landscape composition between both species.

MATERIALS AND METHODS

Plant Material and DNA Sequencing

Details for provenance and vouchers of *H. koreana* and *H. umbellata* can be found in Pellicer et al. (2014). These species were selected as they show the largest difference in genome size between the six species which comprise this genus, with *H. umbellata* (4,680 Mb/1C) having nearly double the DNA amount compared with *H. koreana* (2,480 Mb Mb/1C), and represent species belonging to the two main clades of this small genus. Genomic DNA extraction was carried out using the 2x CTAB method with minor modifications (Doyle and Doyle, 1987) followed by a CsCl/ethidium bromide density gradient and dialysis. The DNA products were run on an 1% agarose gel and quality control assessed using a Qubit 3 fluorometer (Thermo Fisher Scientific). Paired-end shotgun libraries with an average insert size of 500 bp were prepared and sequenced by Beijing Genomics Institute (BGI, Shenzhen, China) on an Illumina HiSeq 2000 (Illumina, San Diego, CA, United States) generating 100 nucleotide reads (0.15 × genome coverage).

[1] https://cvalues.science.kew.org

The quality of sequencing data was assessed using FastQC[2] and reads were pre-processed using the FASTX-Toolkit.[3] Sequence reads were filtered using a threshold quality score of 20 over the full length of the read. Reads of organellar origin were filtered using custom databases of monocot plastid and mitochondrial genomes (all available from NCBI at the time of analysis) using the standalone version of BLAST (v2.2.16; Altschul et al., 1997). Reads with significant hits to either the plastid or mitochondrial databases were then filtered using a custom Perl script (supplied by Laura J. Kelly). The remaining reads were thus considered to be of nuclear origin.

Graph-Based Clustering in RepeatExplorer 2

Repeat identification by similarity-based clustering of Illumina paired-end reads was carried out using the Repeat Explorer 2 pipeline (Novák et al., 2013, 2020b), a GALAXY-based server for characterisation of repetitive elements.[4] FASTQ reads were converted to FASTA format and interlaced prior to the clustering analysis. A preliminary round of clustering was performed with the original datasets [*H. koreana* = 3,127,826 reads (0.11×) and *H. umbellata* = 5,968,792 reads (0.11×)] to determine the maximum number of reads for each species to include in the final analysis. This employed the default settings (90% similarity over 55% of the read length, and cluster size threshold = 0.01%). After the initial screening, each set of reads was randomly down-sampled according to their genome size to represent reads comprising 1.5% of the genome of each species (i.e., genome proportion = 0.015×, *H. koreana* = 410,000 reads and *H. umbellata* = 784,993 reads). Automated repeat classification was based on connection-based clustering via paired-end reads and BLAST (n, x) similarity searches to REXdb (Neumann et al., 2019), a comprehensive database of conserved protein domains in retrotransposons. Output directories were individually examined for a final manual annotation and quantification of clusters and connections to superclusters. In addition, a comparative clustering analysis was carried out using a combined dataset of 1,015,000 reads (each species at a genome proportion equal to 0.013×). A four-letter prefix identity code was added to each sample dataset and used as the input to Repeat Explorer as described above. Repeat annotation of shared clusters between the two species was done following the same parameters as for the individual analyses.

Genome Dynamics and Relative Abundance of Particular Repetitive Elements

The different repeat families in the two *Heloniopsis* species were recorded using the annotation output files from the Repeat Explorer analysis and summarized accordingly. Baseline statistics including, genome proportion (in percentage), abundance (Mb/1C), ratios of transposable elements and correlations between the main families of DNA repeats identified were calculated using R (R Core Team, 2019). A pairwise scatterplot of the main repeat element classes identified was constructed by comparing the number of shared reads between the two species based on McCann et al. (2018), and using the function *ggplot* built in the ggplot2 package (Wickham, 2016). The number of shared reads per cluster were obtained from the output files of the comparative analysis in Repeat Explorer. The slope of the line in the scatterplot represents the genome size ratio between the two species, thus any deviation from the line indicates biases in the contribution of a given element to the genome size of one species, compared with the other. Note that due to the large amounts of satellite DNA in the genome of *H. umbellata* compared with *H. koreana* this repeat type was not included in the scatterplot to enable a better visualization of the remaining data (but included in subsequent statistical regression analyses). Further linear model regression analyses of shared read clusters from the comparative Repeat Explorer analysis were carried out using the function *lm* in R *Stats* package (R Core Team, 2019).

To compare repeat abundances with changes in genome size, ancestral 1C-values in tribe Heloniadeae were reconstructed using maximum likelihood (ML) under a Brownian motion model using functions *ace* and *fastAnc* of the library *Phytools* (Revell, 2012). Genome size data available for extant species in Pellicer et al. (2014) and the phylogenetic tree from Kim et al. (2016) were used for the reconstruction. Following Macas et al. (2015), we also assessed the abundance of solo-LTRs, a product of ectopic unequal homologous recombination between LTRs of the same element type, in the two most abundant retrotransposon lineages (i.e., Ty1/copia-Angela and Ty3/gypsy-Tekay). Whilst not being conclusive, this approach provides an insight into the activity of one of the mechanisms by which LTR retrotransposons can be deleted from the genome (Cossu et al., 2017), and so can be used as a proxy to evaluate the potential impact of this process on genome size. Briefly, the method uses short Illumina reads to calculate an Rsf value, which is the ratio between the number of solo-LTRs to full-length elements for a particular repeat type. Larger Rsf values can indicate a higher impact of unequal homologous recombination. The analysis consists of five-steps: (1) Identification of the LTR-3′end and 5′-UTR junctions from read assemblies produced by the Repeat Explorer pipeline. (2) Extraction of 30 nt sequence tags which are used to create BLAST databases for the LTR-3′, 5′-UTR and a combined LTR-3′ + 5′-UTR (60 nt) regions. (3) BLAST all read sequences to the tag databases for the LTR-3′. (4) Blast all hits from the previous step against the 5′-UTR database. Finally, (5) BLAST hits from the previous step against the combined LTR-3′ + 5′-UTR database. These steps result in sets of reads representing LTR-3′end/5′-UTR junctions (LU) and LTR-3′end only (Lx). The Rsf ratio is then calculated using the formula: Rsf = (Lx—LU)/LU.

Chromosome Preparations and Mapping of DNA Satellite HeloSAT by Fluorescence *in situ* Hybridisation (FISH)

Roots were collected from the same accessions used for genome size estimations and sequencing based on Pellicer et al. (2014).

[2]http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[3]http://hannonlab.cshl.edu/fastx_toolkit/

[4]https://repeatexplorer-elixir.cerit-sc.cz/galaxy/

Briefly, roots were pre-treated in a saturated solution of 1-bromonaphthalene at 20°C for 24 h. Samples were then transferred to ice-cold 90% acetic acid for 10 min and stored in 70% ethanol at −20°C. Protoplast preparation was based on Kato et al. (2011). Roots were washed 3 × in ice-cold 1× citric buffer (50 mM sodium citrate, 50 mM EDTA, pH 5.5), then the tips were excised and macerated in 200 μL tubes containing 20 μL of enzymatic solution containing 4% cellulase Onozuka R-10 (Duchefa, Haarlem, The Netherlands) and 1% pectolyase from *Aspergillus niger* (Merck, Darmstadt, Germany) in 1× citric buffer pH 5.5 for 45–48 min at 37°C and transferred to ice. Digested roots were subsequently washed three times in ice cold 70% ethanol. Finally, 30 μL of ice-cold glacial acetic acid was added and mixed before dropping 4 μL of the protoplast suspension onto a microscope slide in a humid chamber until dry.

A non-denaturing and formamide-free fluorescent *in situ* hybridisation (FISH) protocol based on Cuadrado et al. (2009) and Mian (2019) was applied. A 26 bp oligo probe of HeloSAT was synthesized and labeled with FITC based on the output RE cluster monomers obtained (**Supplementary Online Resource 1**). The probe was evaluated to avoid self-hybridisation (i.e., dimerization and hairpins) with Oligo Calc.[5] Oligo probes are single-stranded, therefore they do not need denaturation prior to hybridisation. The hybridisation mixture was simply prepared by diluting 2 μL of the 5′ end-labeled HeloSAT oligo (1 pmol/μL, Eurofins) in 1 × SSC pH 7.0 in a final volume of 15 μL. Hybridisation was carried out for 1 h at 37°C in a humid chamber. A post hybridization stringency wash was performed by transferring the slides to 1 × SSC 0.1% Triton X-100 buffer at 37°C, for long enough to allow coverslips to fall away from the slides (c. 5 min). The slides were then dehydrated in an ethanol series of 70, 90, and 100%, air-dried and subsequently counterstained with DAPI (Vectashield, Vector Laboratories, Burlingame, CA, United States). Preparations were examined using a Zeiss Axio Imager.Z2 fitted with an Axiocam 506 mono camera. Images were processed with Zeiss Zen 2.6 (blue edition) software (Zeiss).

RESULTS

Repeat Content in *H. koreana* and *H. umbellata*

Details on the number of reads analyzed for each species and their genomic coverages are given in **Table 1**. A minimum coverage of 0.01% was required to classify a given cluster as repetitive DNA (i.e., a medium to high abundance repeat). The proportion of the genome estimated to be comprised of repetitive DNA sequences varied from 56.32% in *H. koreana* to 73.03% in *H. umbellata* (**Table 2**). Annotation and classification of the most abundant repeat clusters is presented in **Table 2** and shown graphically in **Figure 1**. The clusters that failed to match any known elements from the REXdb where left as unclassified (i.e., 5.06% in *H. koreana* and 7.19% *H. umbellata*, **Table 2**). Overall, most of the identified repeats in the analysis of each species independently were more abundant in the larger genome of *H. umbellata*, with the exception of three specific lineages (i.e., Ty1/copia-Angela, Ty3/gypsy-Athila and LINE), which each had a higher genome proportion (in %) in the genome of *H. koreana* (**Table 2**). The repetitive landscape of both species was dominated by long terminal repeat (LTR) retrotransposons, which accounted for c. 60–78% of all identifiable DNA repeats. Among them, Ty1/copia-like elements were abundant, accounting for 29.04% (i.e., 714.45 Mb) and 24.27% (i.e., 1,134.43 Mb) of the genome in *H. koreana* and *H. umbellata*, respectively. Of these, Angela elements played a significant role in shaping these genomes, with genome proportions reaching 26.92% (i.e., 662.23 Mb) and 21.06% (i.e., 985.70 Mb) in *H. koreana* and *H. umbellata*, respectively. Other Ty1/copia lineages were present, but with a much lower contribution to each genome (i.e., <2%, **Table 2** and **Figure 1**). Ty3/gypsy-like elements were also present in both species and accounted for a lower but still significant proportion of the genomes compared with Ty1/copia-like elements [i.e., 15.12% (i.e., 373.62 Mb) and 19.61% (917.79 Mb) of the repetitive fraction in *H. koreana* and *H. umbellata,* respectively]. The abundance of tandem repeats, namely satellite DNA, differed considerably between the species. While four different major satellite clusters were found in *H. koreana* (2.33%), in *H. umbellata*, only two satellite clusters were recovered, but these accounted for 14.24% of the genome (one of them having a genome proportion of 10.20%).

Genome Size and Comparative Repeat Dynamics in *Heloniopsis*

Ancestral 1C-values (Anc1C) reconstruction in Heloniadeae is depicted in **Figure 2**. The most recent common ancestor (MRCA) of *Heloniopsis* was reconstructed to have an Anc1C = 3,022 Mb (**Figure 2**, clade 2). Since then, contrasting genome size dynamics have been inferred in the two main clades of the genus (**Figure 2**, clades 3 and 4), illustrating that ups and downs in genome size have taken place during the evolution of the genus. The comparative clustering involved analyzing 1.05 million reads (0.013 × GP/1C per species). The genomic proportions (in %) of the shared top 20 superclusters (grouped by repeat classification) are depicted in **Figure 3**. The ratios observed between the genome proportions of these shared repeats in each species are illustrated in **Figure 4A**, and show that a few repeat types (e.g., Ty1/copia-Angela, Ty1/copia-TAR, Ty3/gypsy-Tekay) occur in similar genomic proportions in the two species and hence indicating that they are (nearly) twice as abundant in total copy number in *H. umbellata* compared with *H. koreana*. Nevertheless, other repeat types deviate from this ratio, and hence comprise a higher (e.g., Ty1/copia-Ale) or lower (e.g., Ty3/gypsy-Athila) genome proportions in *H. koreana* compared with *H. umbellata*. At the sequence read level, by comparing the total number of reads from each species in each repeat cluster, this trend is also illustrated in **Figure 4B**, where many reads in shared clusters were biased toward contributing to the genome of *H. umbellata* (especially clusters containing over 2,500 reads).

Table 3 shows the regressions between the abundance of different repeat types based on the genome sizes of the two

[5]http://biotools.nubic.northwestern.edu/OligoCalc.html

TABLE 1 | Estimation of genome size and sequencing of the two *Heloniopsis* species studied.

Species	Genome size 1C (Mbp)	Chrom number (2n)	Sequencing run number	Individual clustering		Comparative clustering
				No. reads	Coverage (×1C)	No. reads (0.013×)
H. koreana	2,480	34	SRR15208643	410,000	0.015×	349,706
H. umbellata	4,680	34	SRR15208642	784,994	0.015×	665,294

Heloniopsis species studied. Significant strong relationships were found based on the genome sizes and the two major lineages of retrotransposons which have the highest impact on the genome composition of these two species [i.e., Ty1/copia-Angela (R^2 = 0.76, p = 1.32e^{-09}) and Ty3/gypsy-Retand (R^2 = 0.96, p = 4.28e^{-08})]. A positive correlation was also found when all the LTR-retrotransposon elements were analyzed together (R^2 = 0.81, p = 2.01e^{-16}). However, when all DNA repeats were analyzed as a whole, DNA satellites were seen to have a significant impact on the regression, as shown by the improvement of the correlation coefficient when DNA satellites were excluded from the analysis (R^2 = 0.26, p = 3.49e^{-11} versus R^2 = 0.88, p = 2.24e^{-16}, respectively).

To explore whether differences in the amount of unequal homologous recombination between the LTR sequences could be contributing to the differences in genome size between the two *Heloniopsis* species, we identified the LTR-3′ end and 5′ untranslated regions from the Ty1/copia-Angela and Ty3/gypsy-Retand elements, which were the two most abundant transposable elements (**Table 2**). Despite the relatively high Rsf values estimated for both repeats in both species, and the caution that needs to be paid when interpreting these data as evidence for recombination using this approach, the values obtained for *H. koreana* (Rsf-Angela: 11.41, Rsf-Retand: 4.22), were higher than those in *H. umbellata* (Rsf-Angela: 7.94, Rsf-Retand: 3.88).

TABLE 2 | Repeat composition inferred in the studied *Heloniopsis* species.

Repeat type	Lineage	Genome proportion (%)	
		H. koreana	*H. umbellata*
Retrotransposon			
Ty1/copia			
	All	**29.04**	**24.27**
	Angela	26.92	21.06
	Ale	0.85	1.63
	TAR	0.87	0.90
	Tork	0.29	0.43
	Ikeros	0.10	0.23
Ty3/gypsy			
	All	**15.12**	**19.61**
	Retand	6.04	8.80
	Tekay	2.18	4.80
	Athila	5.07	3.52
	Tat/Ogre	1.84	2.23
	CRM	0.06	0.25
Other repeats			
	LINE	0.61	0.30
	Pararetrovirus	–	0.01
DNA transposons			
	All	**3.85**	**7.02**
	Enspm/CACTA	2.96	6.26
	hAT	0.83	0.51
	MuDR	0.06	0.17
	Helitron	–	0.08
Tandem repeats			
	rDNA	0.24	0.36
	Satellite	2.33	14.24
Unclassified		5.06	7.19
Total repeats		**56.32**	**73.03**
Low/single copy		43.67	26.97

Bold values are refer to the cumulative overall proportions of elements belonging to the same group.

Identification and Characterisation of DNA Satellites

The clustering analysis identified four distinct types of satellite DNA, two of which were specific to the *H. koreana* genome while the other two were present in both species (**Supplementary Online Resource 1**). The abundance of the two shared DNA satellites varied between species, particularly in *H. umbellata* where their combined abundance was over six times higher (i.e., 14.24%) than in *H. koreana* (2.33%, **Figure 4A**). The abundance of just one of these satellites in particular, hereafter named HeloSAT, highlights the contrasting evolutionary dynamics between these two closely related species. A single cluster of HeloSAT accounted for 10.20% of the genome in *H. umbellata*, while comprising just 0.83% in *H. koreana*.

To further investigate the presence of this satellite from a comparative viewpoint, HeloSAT was physically mapped onto the chromosomes and interphase nuclei of both species using fluorescent *in situ* hybridization (FISH) (**Supplementary Online Resource 2**). Results from FISH corroborated those from the clustering approach, with HeloSAT signals being more abundant and spread across the chromosomes of *H. umbellata* (**Supplementary Online Resource 2A**) than in *H. koreana* (**Supplementary Online Resource 2C**). The size, fluorescence intensity and number of signals in the latter were lower, which was even more evident when analyzing interphase nuclei from both species (**Supplementary Online Resource 2B,D**).

DISCUSSION

Diversity and Dynamics of Repetitive Elements in *Heloniopsis*

In this work we provide the first insights into genome evolutionary dynamics in the genus *Heloniopsis*, by combining

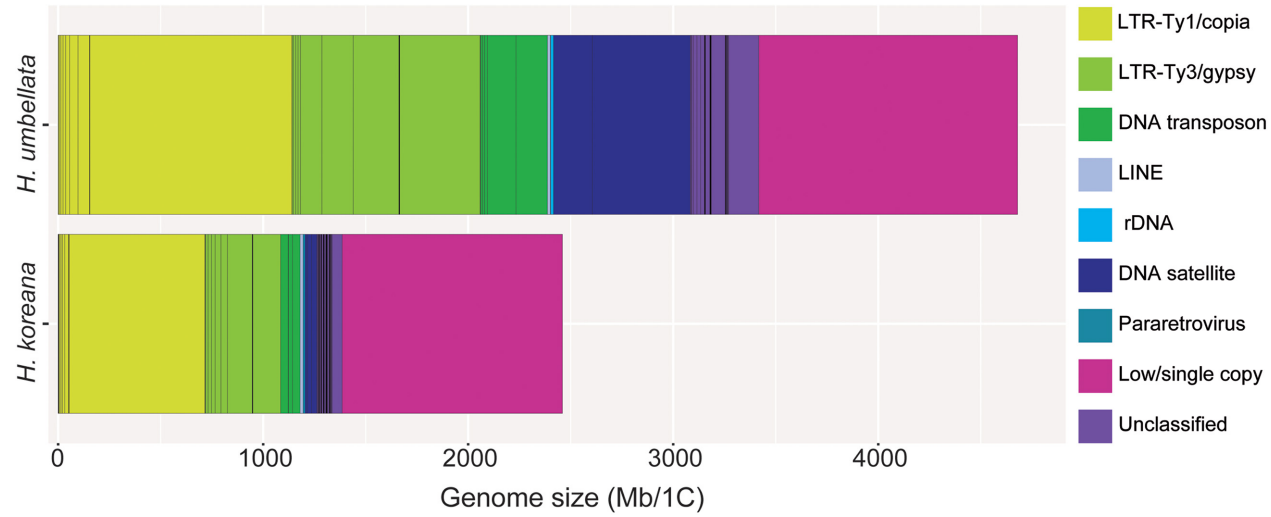

FIGURE 1 | Genomic composition of *Heloniopsis koreana* (2,480 Mb/1C) and *H. umbellata* (4,680 Mb/1C). Estimates of the genomic abundance (in Mb/1C) of different repeats are colored by repeat class. The size of the unclassified (pale purple) and low/single copy fraction (pink) of each genome is also shown.

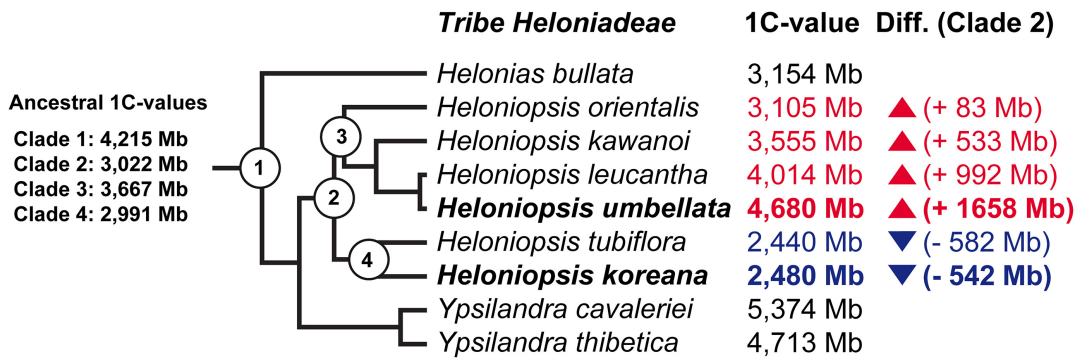

FIGURE 2 | Summary of the ancestral genome size reconstruction in tribe Heloniadeae. Branch tips include the extant 1C-values taken from Pellicer et al. (2014) with the increase (red) or reduction (blue) in genome size indicated based on the 1C-value reconstructed for the most recent common ancestor of the genus (i.e., Clade 2).

high throughput sequence data and cytogenetics. Genome size evolution in the genus may be considered to be bi-directional based on our ancestral state reconstruction analysis, which showed opposite evolutionary trajectories in the two clades (**Figure 2**). Our analysis showed that the most recent common ancestor of *Heloniopsis* likely had an Anc1C of 3,022 Mb, indicating that during the evolution of *H. umbellata* its genome has expanded by 1,658 Mb. Such a trend is in striking contrast to that observed in *H. koreana*, in which we inferred a genome reduction of 542 Mb with respect to the MCRA of the genus. Based on the study by Kim et al. (2019), genome size divergence is estimated to have taken place within a c. 10 Mya period, possibly as far back as the Miocene, when the genus is estimated to have started to diverge. The observed genome sizes, however, do not preclude the possibility that additional shifts also took place during the evolution of the genus, thus our analyses should therefore be seen as just one potential evolutionary scenario based on genome size data from extant species. Certainly, shifts in genome size during the evolutionary history of plants have been reported in many plant lineages (e.g., Lysak et al., 2009; Pellicer et al., 2013; Vallès et al., 2013; Kelly et al., 2015), and Melanthiaceae are no exception. Furthermore, the fact that both species show relatively high proportions of solo-LTRs (i.e., Rsf values) indicate that despite recombination likely affecting *H. koreana* more significantly than *H. umbellata*, based on the higher Rsf values in the two most abundant repeats (i.e., Ty1/copia Angela and Ty3/gypsy-Retand), both species appear to have reduced the abundance of these major repeat types contributing to the overall genome size of these species, and this may have led to an overall genomic contraction if such recombination processes have been sufficiently active to overcome the impact of repeat amplification.

In the absence of polyploidy, the dynamics of transposable elements (mainly LTR-retrotransposons belonging to Ty1/copia and Ty3/gypsy superfamilies) underpin most changes in genome

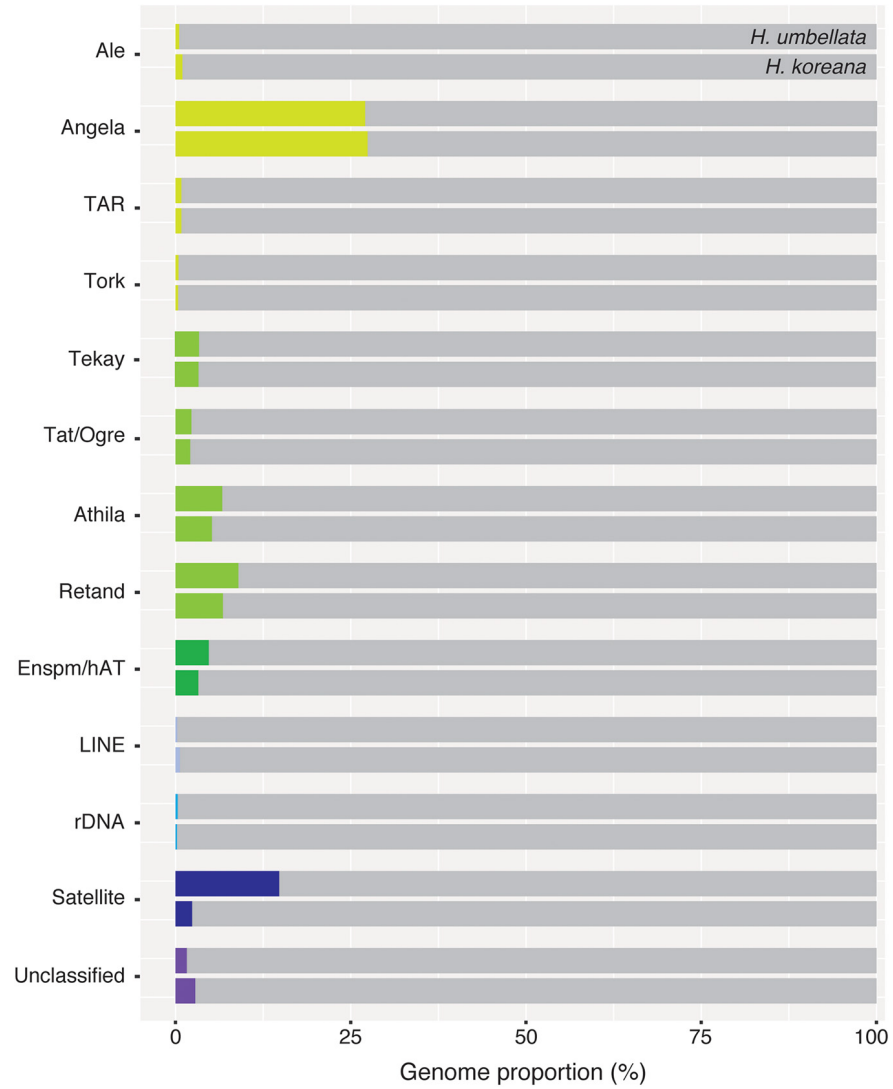

FIGURE 3 | Genomic proportions (in %) of shared repetitive elements between *Heloniopsis koreana* and *H. umbellata*.

size given their significant contribution to the genomic landscape in plants (e.g., McCann et al., 2020). The repetitive fraction of the *Heloniopsis* genomes explored here include several transposable elements and tandem repeats, and the genome size differences between the two species studied can be explained, at least partially, by their repetitive content, given the significant overall correlation between their abundance and genome size variation (**Table 3**). The individual clustering analysis revealed that 56.3% of the genome of *H. koreana* is highly repetitive, whereas in *H. umbellata*, with a larger genome, the repetitive fraction reached 73% of its genome. Such proportions fall within the ranges previously reported for seed plants with similar genome size to those investigated here (Novák et al., 2020a), with the differences in the repetitive landscape observed here providing support for the contrasting genome sizes between the two species analyzed. Indeed, the larger genome of *H. umbellata* containing a larger proportion of repeats, many of which occupied a similar genome proportion to *H. koreana* indicate that their copy number has nearly doubled (**Figure 4**).

The detailed characterisation and identification of repetitive DNA content in the two *Heloniopsis* species studied highlighted the relative impact of the Ty1/copia-Angela elements in their genomes, with proportions ranging from c. 21 to 26% (**Table 2**). Among the Ty1/copia lineages that have been identified in plants (Neumann et al., 2019), Angela elements have been reported to be abundant in other genomes, with similar proportions as found here (e.g., *Passiflora*, *Thinopyrum*; Divashuk et al., 2020; Sader et al., 2021). Such proportions are, nonetheless, lower than for other LTR-retrotransposon elements reported in some plant genomes of comparable size to those of *Heloniopsis*. For example, Tekay/Del elements which belong to the Ty3/gypsy lineage were reported to account for c. 67 and 97% of the repetitive landscape in *Capsicum anuum*

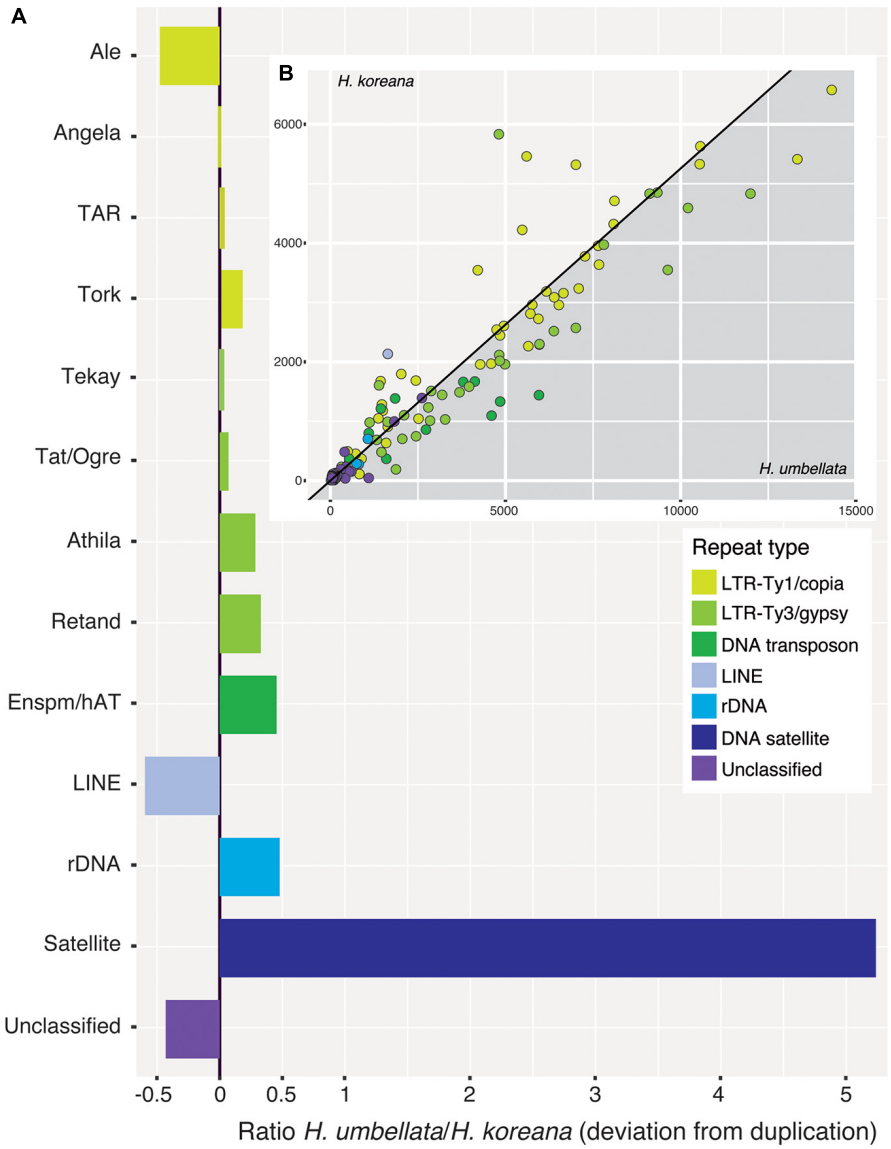

FIGURE 4 | (A) Ratios of incidence of shared repeat clusters from the comparative analysis (*Heloniopsis umbellata*/*H. koreana*). A ratio of zero indicates clusters present in the same genomic proportion in both species. **(B)** (inset) Pairwise scatterplot of the number of reads from each species in shared repeat clusters from the comparative analysis (excluding satellite DNA). The slope of the line is equal to the ratio of the genome sizes of the two species (i.e., 1.9). Dots falling along the line are present in the same genomic proportions in the two species.

and *C. chinense,* respectively (de Assis et al., 2020), although they only accounted for 15% of the repetitive genome of the closely related *C. baccatum* (of similar genome size), illustrating that even within a genus, contrasting evolutionary dynamics can give rise to distinctive repeat profiles. Overall, the analyses indicate that it is often the combined activity of a diverse array of repeat lineages which contribute to differences in genome size observed between species rather than the differential rates of amplification/deletion of just one or few transposable element families. Despite this, our observations on the clustering analysis also suggest that transpositional bursts might have occurred, as in the case of Ty1/copia-Angela, where large superclusters were recovered (**Figure 1**). This would result in higher levels of homogeneity between sequence copies, a pattern in direct contrast to evidence from much larger genomes, such as in those of *Fritillaria* in which even the smallest genome analyzed (in *F. davidii,* 33,525 Mb/1C) is over seven times larger than *H. umbellata* (Kelly et al., 2015). The repeat composition of these immense genomes is highly heterogeneous, indicative of long-term amplification processes combined with low rates of deletion, resulting in a wealth of relatively low−abundance repeat−derived DNAs. Furthermore, retrotransposition can occur at different rates, even between closely related species, making it sometimes difficult to interpret repetitive DNA composition and dynamics in relation to genome size because of the challenges of uncovering the

TABLE 3 | Linear model regressions of repeat contents with genome size variation among *H. koreana* and *H. umbellata*.

Repeats (*H. koreana*/ *H. umbellate)*	R^2	*p*-value
All LTR-retrotransposons	0.818	$2.01e^{-16}$
Ty1/copia-Angela	0.762	$1.32e^{-09}$
Ty3/gypsy-Retand	0.969	$4.28e^{-08}$
All repeats		
(including DNA satellites)	0.259	$3.49e^{-11}$
(excluding DNA satellites)	0.875	$2.24e^{-16}$

signatures of recombination-based mechanisms from short-read sequence data (e.g., Macas et al., 2015). Indeed, an analysis of genome diversity in *Anacyclus* (Asteraceae), revealed that changes in genome size were more significantly underpinned by chromosomal restructuring than by differential dynamics of a reduced set of high-copy-number transposable element families (Vitales et al., 2020).

The Impact of Satellite DNA in Shaping Genome Evolution of *Heloniopsis*

Despite LTR-retrotansposons being the most abundant repeat types uncovered in many plant genomes, the analysis of tandem repeats (i.e., satellite DNA) has also revealed a great diversity in terms of sequence composition, organization and genomic abundance across different land plant species (Garrido-Ramos, 2015). In *Heloniopsis*, compared with other types of repeats identified (see above), satellite DNAs are not the major genomic component. Nevertheless, differences in their abundance have contributed to the differences in genome size observed between the two species analyzed. This is shown by the contrasting genome proportions of the most abundant satellite identified called HeloSAT. Thus, although HeloSAT accounted for up to c. 477 Mb (i.e., 10.20%) of the *H. umbellata* genome, its genome proportion in *H. koreana* was just c. 0.83%. To further explore this satellite repeat, its overall physical organization along the chromosomes of the two *Heloniopsis* species was determined using FISH. As **Supplementary Online Resource 2** shows, there were more hybridisation signals visible on the chromosomes and interphase nuclei of *H. umbellata* than of *H. koreana*. Although FISH is not a fully quantitative technique, the results support the contrasting genome proportions of this satellite in *H. umbellata* compared with *H. koreana* estimated using Repeat Explorer. The satellite appeared to be more widely distributed across the genome of *H. umbellata*, than that of *H. koreana*, with hybridisation signals present on most chromosomes. Nevertheless, despite genome size and satellite size correlating to some extent in *Heloniopsis*, such a trend is not the rule across all plant lineages studied to date. For example, closely related *Paphiopedilum* species with very similar genome sizes were shown to contain divergent satellite elements which differed considerably in abundance between closely related species (Lee et al., 2018). The data provided support to the suggestion that satellite DNAs often evolve rapidly and differ considerably in abundance even in related species with little correlation with genome size (Macas et al., 2010). In addition, recent research in *Passiflora* (1C range = 207.34–2,621.04 Mb) reported an unusually large number of satellite repeats in the species with the smallest genome, albeit at lower frequencies, leading the authors to propose that, in most species, tandem repeats have only a limited impact on the overall genome size of *Passiflora* (Sader et al., 2021).

The number and types of satellite DNAs present in plant genomes can be highly variable. For instance, in *Cuscuta*, > 113 putative DNA satellites were recovered, with relatively substantial genome proportions up to c. 15–18%, and comprising several gigabasepairs in some taxa with relatively large genomes (i.e., 3,400 Mbp/1C) (Neumann et al., 2021). Similarly, in *Vicia peregrina,* 51 satellites were identified (Macas et al., 2015), whereas in *Luzula elegans* 37 satellites were reported (Heckmann et al., 2013). In contrast, other species including *Heloniopsis* have been shown to have a much lower diversity of satellite types, and a higher incidence of species-specific satellites (e.g., Macas et al., 2011; Lee et al., 2018; Mata-Sucre et al., 2020). This is the case, for example in *Fritillaria affinis*—in which only one satellite—FriSAT—was identified, although it accounted for c. 11% of its genome, and it was almost absent in closely related species (Kelly et al., 2015), indicative of rapidly evolving DNA clusters with strong phylogenetic signal.

CONCLUSION

Novel data characterizing the repetitive DNA landscape in *Heloniopsis* have been presented using genome skimming data from short read high throughput sequencing. Although polyploidy and the differential activity of repetitive DNAs have been shown to be major drivers of genome size evolution in plants, even in some closely related species, differences in genome size may evolve through contrasting repeat dynamics alone. Our analysis of the repetitive genome of two *Heloniopsis* species illustrates the latter, as the nearly twofold difference in genome size between species has arisen without any change in chromosome number. The detailed characterisation and comparative analysis of the repetitive DNA content of *H. umbellata* and *H. koreana* show that their genomes are dominated by LTR-retrotransposons, with the larger genome of *H. umbellata* mainly being determined by the increased abundance of the same LTR-retrotransposon elements already present in *H. koreana* rather than the amplification of new repeat types. Few satellite DNAs were recovered, but the characterisation of HeloSAT and its abundance in the genome of *Heloniopsis*, especially *H. umbellata*, provides support for the relevance of satellite DNA in shaping genome size evolution in some plant species.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, SRR15208642; https://www.ncbi.nlm.nih.gov/, SRR15208643.

AUTHOR CONTRIBUTIONS

JP, MFF, and IJL conceived and designed the study. JP, PF, and EM conducted the laboratory work and analyses of data. JP and IJL wrote the manuscript and all authors reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

The authors acknowledge financial support of the Agencia Estatal de Investigación—Ministerio de Ciencia e Innovación (Spanish Government: PID2019-108173GA-I00). JP benefited from a Ramón y Cajal contract (Ministerio de Ciencia e Innovación, Gobierno de España—Ref: RYC-2017-2274). PF and EM received Erasmus + grants from the European Union.

ACKNOWLEDGMENTS

We thank Wendy Zomlefer for discussions on Melanthiaceae in the early stages of this project, Laura J. Kelly for providing the Perl script used to filter plastid DNA, Michael Chester for his technical advice in cytogenetics, and Richard Kernick and Kit Strange (RBG, Kew) for taking care of the plant collections.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.726211/full#supplementary-material

Supplementary Online Resource 1 | Top four-most abundant monomers of HeloSAT.

Supplementary Online Resource 2 | Physical mapping of the satellite DNA (HeloSAT) on chromosomes and interphase nuclei of *Heloniopsis umbellata* **(A,B)** and *H. koreana* **(C,D)**.

REFERENCES

Ågren, J. A., Greiner, S., Johnson, M. T. J., and Wright, S. I. (2015). No evidence that sex and transposable elements drive genome size variation in evening primroses. *Evolution (N. Y.)* 69, 1053–1062. doi: 10.1111/evo.12627

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

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

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

Cuadrado, Á, Golczyk, H., and Jouve, N. (2009). A novel, simple and rapid nondenaturing FISH (ND-FISH) technique for the detection of plant telomeres. Potential used and possible target structures detected. *Chromosom. Res.* 17:755. doi: 10.1007/s10577-009-9060-z

de Assis, R., Baba, V. Y., Cintra, L. A., Gonçalves, L. S. A., Rodrigues, R., and Vanzela, A. L. L. (2020). Genome relationships and LTR-retrotransposon diversity in three cultivated *Capsicum* L. (Solanaceae) species. *BMC Genomics* 21:237. doi: 10.1186/s12864-020-6618-9

Devos, K. M., Brown, J. K. M., and Bennetzen, J. (2002). Genome size seduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 12, 1075–1079. doi: 10.1101/gr.132102

Divashuk, M. G., Karlov, G. I., and Kroupin, P. Y. (2020). Copy number variation of transposable elements in *Thinopyrum* intermedium and its diploid relative species. *Plants* 9:15. doi: 10.3390/plants9010015

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

Garrido-Ramos, M. A. (2015). Satellite DNA in plants: more than just rubbish. *Cytogenet. Genome Res.* 146, 153–170. doi: 10.1159/000437008

Hawkins, J. S., Kim, H., Nason, J. D., Wing, R. A., and Wendel, J. F. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* 16, 1251–1261. doi: 10.1101/gr.5282906

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

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

Hloušková, P., Mandáková, T., Pouch, M., Trávníček, P., and Lysak, M. A. (2019). The large genome size variation in the *Hesperis* clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. *Ann. Bot.* 124, 103–120. doi: 10.1093/aob/mcz036

Kato, A., Lamb, J. C., Albert, P. S., Danilova, T., Han, F., Gao, Z., et al. (2011). "Chromosome Painting for Plant Biotechnology," in *Plant Chromosome Engineering: Methods and Protocols*, ed. J. A. Birchler (New York, NY: Springer), 67–96. doi: 10.1007/978-1-61737-957-4_4

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

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

Kim, S.-C., Kim, J. S., Chase, M. W., Fay, M. F., and Kim, J.-H. (2016). Molecular phylogenetic relationships of Melanthiaceae (Liliales) based on plastid DNA sequences. *Bot. J. Linn. Soc.* 181, 567–584. doi: 10.1111/boj.12405

Kokubugata, G., Peng, C. I., and Yokota, M. (2004). Comparison of karyotypes among three *Heloniopsis* species from Ryuku Archipelago and Taiwan. *Ann. Tsukuba Bot. Gard.* 23, 13–16.

Lee, Y.-I., Yap, J. W., Izan, S., Leitch, I. J., Fay, M. F., Lee, Y.-C., et al. (2018). Satellite DNA in *Paphiopedilum* subgenus *Parvisepalum* as revealed by high-throughput sequencing and fluorescent in situ hybridization. *BMC Genomics* 19:578. doi: 10.1186/s12864-018-4956-7

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., Kejnovský, E., Neumann, P., Novák, P., Koblížková, A., and Vyskot, B. (2011). Next generation sequencing-based analysis of repetitive DNA in the model dioceous plant *Silene latifolia*. *PLoS One* 6:e27335. doi: 10.1371/journal.pone.0027335

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. *Bioinformatics* 26, 2101–2108.

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

Mata-Sucre, Y., Sader, M., Van-Lume, B., Gagnon, E., Pedrosa-Harand, A., Leitch, I. J., et al. (2020). How diverse is heterochromatin in the *Caesalpinia* group? Cytogenomic characterization of *Erythrostemon hughesii* Gagnon & G.P. Lewis (Leguminosae: Caesalpinioideae). *Planta* 252:49. doi: 10.1007/s00425-020-03453-8

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 saisies (*Melampodium sect. Melampodium*, Asteraceae). *Syst. Biol.* 67, 1010–1024. doi: 10.1093/sysbio/syy024

McCann, J., Macas, J., Novák, P., Stuessy, T. F., Villaseñor, J. L., 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

Mian, S. (2019). *The Impact of Genomic Structural Variation on Meiotic Pairing and Segregation in Beta vulgaris subsp. maritima*. Ph. D. thesis. London: Queen Mary University of London, doi: 10.34885/61

Neumann, P., Novák, P., Hoštáková, 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. *Mob. DNA* 10:1. doi: 10.1186/s13100-018-0144-1

Neumann, P., Oliveira, L., Čížková, J., Jang, T.-S., Klemme, S., Novák, P., et al. (2021). Impact of parasitic lifestyle and different types of centromere organization on chromosome and genome evolution in the plant genus *Cuscuta*. *New Phytol.* 229, 2365–2377. doi: 10.1111/nph.17003

Novák, P., Guignard, M. S., Neumann, P., Kelly, L. J., Mlinarec, J., Koblížková, A., et al. (2020a). Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nat. Plants.* 6, 1325–1329. doi: 10.1038/s41477-020-00785-x

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

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

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211

Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes (Basel).* 9:88. doi: 10.3390/genes9020088

Pellicer, J., Kelly, L. J., Leitch, I. J., Zomlefer, W. B., and Fay, M. F. (2014). A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. *New Phytol.* 201, 1484–1497. doi: 10.1111/nph.12617

Pellicer, J., Kelly, L. J., Magdalena, C., and Leitch, I. J. (2013). Insights into the dynamics of genome size and chromosome evolution in the early diverging angiosperm lineage Nymphaeales (water lilies). *Genome* 56, 437–449. doi: 10.1139/gen-2013-0039

Piegu, B., Guyot, R., Picault, N., Roulin, A., Sanyal, A., Kim, H., et al. (2006). Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* 16, 1262–1269. doi: 10.1101/gr.5290206

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

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

Sader, M., Vaio, M., Cauz-Santos, L. A., Dornelas, M. C., Vieira, M. L. C., Melo, N., et al. (2021). Large vs small genomes in *Passiflora*: the influence of the mobilome and the satellitome. *Planta* 253:86. doi: 10.1007/s00425-021-03598-0

Schubert, I., and Vu, G. T. H. (2016). Genome stability and evolution: attempting a holistic view. *Trends Plant Sci.* 21, 749–757. doi: 10.1016/j.tplants.2016.06.003

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

Vitales, D., Álvarez, I., Garcia, S., Hidalgo, O., Nieto Feliner, G., Pellicer, J., et al. (2020). Genome size variation at constant chromosome number is not correlated with repetitive DNA dynamism in *Anacyclus* (Asteraceae). *Ann. Bot.* 125, 611–623. doi: 10.1093/aob/mcz183

Vu, G. T. H., Cao, H. X., Reiss, B., and Schubert, I. (2017). Deletion-bias in DNA double-strand break repair differentially contributes to plant genome shrinkage. *New Phytol.* 214, 1712–1721. doi: 10.1111/nph.14490

Wang, D., Zheng, Z., Li, Y., Hu, H., Wang, Z., Du, X., et al. (2021). Which factors contribute most to genome size variation within angiosperms? *Ecol. Evol.* 11, 2660–2668. doi: 10.1002/ece3.7222

Wang, X., Morton, J., Pellicer, J., Leitch, I. J., and Leitch, A. R. (2021). Genome downsizing after polyploidy: mechanisms, rates and selection pressures. *Plant J.* doi: 10.1111/tpj.15363 [Epub ahead of print].

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

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, 2nd Edn. New York, NY: Springer-Verlag.

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 AK declared a past co-authorship with the authors JP, IJL to the handling editor.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

CORRECTION
published: 03 November 2021
doi: 10.3389/fgene.2021.799661

Corrigendum: Genome Size Doubling Arises From the Differential Repetitive DNA Dynamics in the Genus *Heloniopsis* (Melanthiaceae)

Jaume Pellicer[1,2]*, *Pol Fernández*[1], *Michael F. Fay*[2,3], *Ester Michálková*[4] *and Ilia J. Leitch*[2]

[1]*Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Barcelona, Spain,* [2]*Royal Botanic Gardens, Kew, Richmond, United Kingdom,* [3]*School of Plant Biology, University of Western Australia, Crawley, WA, Australia,* [4]*Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czechia*

Keywords: C-value, DNA repeats, chromosome, transposable elements, satellite DNA

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Jaume Pellicer
jaume.pellicer@ibb.csic.es

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *21 October 2021*
Accepted: *22 October 2021*
Published: *03 November 2021*

Citation:
Pellicer J, Fernández P, Fay MF, Michálková E and Leitch IJ (2021) Corrigendum: Genome Size Doubling Arises From the Differential Repetitive DNA Dynamics in the Genus Heloniopsis *(Melanthiaceae). Front. Genet. 12:799661. doi: 10.3389/fgene.2021.799661*

A Corrigendum on

Genome Size Doubling Arises From the Differential Repetitive DNA Dynamics in the Genus *Heloniopsis* (Melanthiaceae)

by Pellicer J., Fernández P., Fay M. F., Michálková E. and Leitch I. J. (2021). Front. Genet. 12:726211. doi: 10.3389/fgene.2021.726211

In the published article, there was an error in **affiliation 4**. Instead of "**Faculty of Sciences, University of Masaryk, Brno, Czechia**", it should be "**Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czechia**".

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 December 2021
doi: 10.3389/fgene.2021.768987

Comparison of Karyotypes in Two Hybridizing Passerine Species: Conserved Chromosomal Structure but Divergence in Centromeric Repeats

Manon Poignet[1], Martina Johnson Pokorná[1,2,3], Marie Altmanová[2,3], Zuzana Majtánová[3], Dmitry Dedukh[3], Tomáš Albrecht[1,4], Jiří Reif[5,6], Tomasz S. Osiejuk[7] and Radka Reifová[1]**

[1]Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic, [2]Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic, [3]Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Liběchov, Czech Republic, [4]Institute of Vertebrate Biology, Czech Academy of Sciences, Brno, Czech Republic, [5]Institute for Environmental Studies, Faculty of Science, Charles University, Prague, Czech Republic, [6]Department of Zoology and Laboratory of Ornithology, Faculty of Science, Palacký University, Olomouc, Czech Republic, [7]Department of Behavioural Ecology, Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

OPEN ACCESS

Edited by:
Magdalena Vaio, Universidad de la República, Uruguay

Reviewed by:
Rafael Kretschmer, Federal University of São Carlos, Brazil
Edivaldo Herculano Correa De Oliveira, Evandro Chagas Institute, Brazil

***Correspondence:**
Manon Poignet
manon.poignet@gmail.com
Radka Reifová
radka.reifova@natur.cuni.cz

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *01 September 2021*
Accepted: *10 November 2021*
Published: *06 December 2021*

Citation:
Poignet M, Johnson Pokorná M, Altmanová M, Majtánová Z, Dedukh D, Albrecht T, Reif J, Osiejuk TS and Reifová R (2021) Comparison of Karyotypes in Two Hybridizing Passerine Species: Conserved Chromosomal Structure but Divergence in Centromeric Repeats. Front. Genet. 12:768987. doi: 10.3389/fgene.2021.768987

Changes in chromosomal structure involving chromosomal rearrangements or copy number variation of specific sequences can play an important role in speciation. Here, we explored the chromosomal structure of two hybridizing passerine species; the common nightingale (*Luscinia megarhynchos*) and the thrush nightingale (*Luscinia luscinia*), using conventional cytogenetic approaches, immunostaining of meiotic chromosomes, fluorescence *in situ* hybridization as well as comparative genomic hybridization (CGH). We found that the two nightingale species show conserved karyotypes with the same diploid chromosome number of 2n = 84. In addition to standard chromosomes, both species possessed a small germline restricted chromosome of similar size as a microchromosome. Just a few subtle changes in chromosome morphology were observed between the species, suggesting that only a limited number of chromosomal rearrangements occurred after the species divergence. The interspecific CGH experiment suggested that the two nightingale species might have diverged in centromeric repetitive sequences in most macro- and microchromosomes. In addition, some chromosomes showed changes in copy number of centromeric repeats between the species. The observation of very similar karyotypes in the two nightingale species is consistent with a generally slow rate of karyotype evolution in birds. The divergence of centromeric sequences between the two species could theoretically cause meiotic drive or reduced fertility in interspecific hybrids. Nevertheless, further studies are needed to evaluate the potential role of chromosomal structural variations in nightingale speciation.

Keywords: chromosomal structure, karyotype evolution, comparative genomic hybridization, rDNA, centromere, GRC, birds, *Luscinia*

INTRODUCTION

Despite an increasing number of sequenced avian genomes (Jarvis et al., 2014; Zhang et al., 2014; Feng et al., 2020), we still know relatively little about the organization of the genomes into chromosomes and to what degree the chromosomal structure (i.e., number, size and collinearity of chromosomes) varies among species. It has been proposed that changes in chromosomal structure, including chromosomal translocations, inversions and copy number variations, may play an important role in the origin of reproductive isolation between species (White, 1978; Rieseberg, 2001; Wellenreuther et al., 2019; Zhang et al., 2021). For example, chromosomal translocations may cause problems with chromosome pairing, recombination and segregation during meiosis, which can lead to hybrid sterility (White, 1978; King, 1993; Homolka et al., 2007). Structural changes, such as inversions, may facilitate speciation by reducing the recombination rate within the structural variant, which may help to maintain the species-specific traits in the face of gene flow (Rieseberg, 2001; Ortíz-Barrientos et al., 2002; Butlin, 2005; Hoffmann and Rieseberg, 2008). Finally, copy number variations may serve as a source of adaptive phenotypic variation (Perry et al., 2007; Zhou et al., 2011; Iskow et al., 2012; Bruders et al., 2020; Minias et al., 2020) and, in the case of copy number variation of the centromeric repeats, they can affect chromosome segregation during meiotic division (Akera et al., 2019), which may, in the extreme case, cause sterility in hybrids (Hurst and Pomiankowski, 1991; Phadnis and Orr, 2009; Zhang et al., 2015). Despite the assumed importance of structural variants in speciation, there are still relatively few studies comparing chromosomal structure between closely related species in the early stages of divergence (Hooper and Price, 2015, 2017; Hooper et al., 2019; Weissensteiner et al., 2020).

Among terrestrial vertebrates, birds have relatively stable karyotypes, usually composed of approximately 40 pairs of chromosomes, which include around 10 macrochromosomes and 30 mostly indistinguishable microchromosomes (Christidis, 1990; Pichugin et al., 2001; Masabanda et al., 2004; Griffin et al., 2007; Ellegren, 2010; Nanda et al., 2011; Degrandi et al., 2020a). In addition to size, macrochromosomes and microchromosomes differ in their GC content, gene density, recombination rate and substitution rate (Auer et al., 1987; Rodionov et al., 1992; Smith et al., 2000; Burt, 2002; Axelsson et al., 2005). Although the number and size of chromosomes is quite conserved in birds, indicating that interchromosomal rearrangements are rare in this group, intrachromosomal rearrangements such as inversions can occur relatively frequently (Aslam et al., 2010; Völker et al., 2010; Ellegren, 2013; Hooper and Price, 2017; Rodrigues et al., 2017). There is also evidence for relatively frequent copy number variations among birds (Skinner et al., 2014). All birds possess a ZW sex determination system (ZZ for male; ZW for female) with a large Z chromosome and usually smaller heterochromatic W chromosome (dos Santos et al., 2015; Schartl et al., 2016; Barcellos et al., 2019). In addition, it has been revealed that passerines possess an additional chromosome in their germ cells, the so-called germ-line restricted chromosome (GRC) (Pigozzi and Solari, 1998; Kinsella et al., 2019; Torgasheva et al., 2019). This chromosome is eliminated from the somatic cells during early development; being maintained only in the germline. In some passerines the GRC represents a big macrochromosome, while in others, a small microchromosome (Torgasheva et al., 2019). However, the size of this chromosome has only been characterized in 16 species so far and it is not clear how often it differs among closely related species (Torgasheva et al., 2019).

To date, somatic karyotypes of approximately 1,000 avian species (i.e., 10% of all bird species) have been described using mostly classical cytogenetic techniques such as G- and C-banding and Giemsa staining (reviewed in Degrandi et al., 2020a). Such techniques enable rough estimation of the diploid chromosome number as well as the detection of large chromosomal translocations or inversions. However, distinguishing smaller-scale chromosomal rearrangements and counting the number of microchromosomes is usually challenging. Moreover, karyotypes from somatic cells do not allow for the detection of the GRC. The development of molecular cytogenetic methods, such as fluorescence *in situ* hybridization (FISH) and whole chromosome probes (Griffin et al., 1999), made more detailed cross-species comparisons of chromosomal structure possible, but have so far only been applied to relatively few avian species, with chicken probes mostly being used as a reference (reviewed in Kretschmer et al., 2018; Degrandi et al., 2020a). In addition, immunostaining of the synapsed chromosomes during meiosis provides a useful approach for detection of the GRC and comparing the chromosomal structure among species (Hale et al., 1988; Torgasheva et al., 2019).

Based on the FISH technique Kallioniemi et al. (1992) developed a new fine scale molecular cytogenetic method called comparative genomic hybridization (CGH). This method allows for the detection of unbalanced chromosomal rearrangements (i.e., duplications, deletions, and copy number variation) between two sources of DNA. Originally the method was designed to detect chromosomal changes in tumor cells compared to normal cells (Kallioniemi et al., 1992). Later, it was used for sex chromosome detection using male and female DNA (e.g., Koubová et al., 2014; Pokorná et al., 2014). Finally, an interspecific design was developed to detect chromosomal rearrangements between species (Bi and Bogart, 2006; Symonová et al., 2015; de Oliveira et al., 2019). In most CGH studies done in birds, the chicken genome has been used as a reference with a microarray-based CGH platform (array-CGH) (Skinner et al., 2009, 2014; Völker et al., 2010). To our knowledge, no interspecific CGH comparisons have been performed to detect copy number variation between closely related bird species.

In this study, we compared the karyotypes of two closely related passerines species, the common nightingale (*Luscinia megarhynchos*) and the thrush nightingale (*Luscinia luscinia*), that diverged ~1.8 Mya (Storchová et al., 2010) and currently hybridize in a secondary contact zone spanning Central and Eastern Europe (Reifová et al., 2011a). These species are separated by incomplete reproductive isolation, which is mainly caused by female-limited hybrid sterility (Reifová et al., 2011b; Mořkovský et al., 2018) and partial ecological divergence in sympatry (Reif

et al., 2018; Sottas et al., 2018, 2020). In addition, divergence in sperm morphology might contribute to postmating prezygotic isolation (Albrecht et al., 2019). Of these species, the common nightingale's karyotype has been previously described using classical cytogenetic analysis of the somatic metaphases (Bozhko, 1971). However, the karyotype of the thrush nightingale has yet to be determined.

Here we performed a cytogenetic analysis of the nightingale karyotypes to test whether changes in chromosomal structure might be linked to reproductive isolation between the species. To do so, we applied conventional and molecular cytogenetics methods to mitotic and meiotic spreads. These methods included C-banding, immunofluorescence staining of synapsed pachytene chromosomes, physical mapping of telomeric and 18S rDNA probes using FISH, and finally CGH.

MATERIALS AND METHODS

Sampling Procedure

The sampling of the two nightingale species was carried out in allopatric regions (where only one of the two species occurs) to avoid possible sampling of interspecific hybrids. The common nightingale was sampled in South-western Poland by the Odra river, near the town Brzeg Dolny (N 51.2602°, E 16.7440°). The thrush nightingale was sampled in North-eastern Poland by the Narew river, near the town Łomża (N 53.1621°, E 22.1246°). In total, we sampled four common nightingales (one male, three females) and two thrush nightingales (one male, one female) for mitotic spreads and two males of each species for meiotic spreads. The birds were euthanized by a standard cervical dislocation and the tibia and testes were immediately dissected for the preparation of mitotic and meiotic chromosomal spreads, respectively. In addition, we collected a blood sample from the brachial vein from one female of each species. The blood sample was later used for DNA isolation and preparation of species-specific DNA probes for the interspecific comparative genomic hybridization (CGH) experiment. All individuals were sampled in May 2019, during the breeding season, and were captured using mist nets or collapsible traps. The work was approved by the General Directorate for Environmental Protection, Poland (permission no. DZP-WG.6401.03.123.2017.dl.3).

Mitotic Chromosome Preparation and C-Banding

Bone marrow from the tibias of each bird was flushed out using a syringe needle with D-MEM medium (Sigma Aldrich) and cultivated in 5 ml of D-MEM medium (Sigma Aldrich) with 75 µl of colcemid solution (Roche) for 40 min at 37°C. After that, the cells were hypotonized in pre-warmed 0.075 M KCl solution for 25 min at 37°C. Finally, cells were washed four times with fixative solution (methanol:acetic acid, 3:1) and then stored at −20°C prior to use.

Chromosomal spreading was done using the air-drying technique followed by conventional Giemsa staining (5% Giemsa in 0.07 M phosphate buffer, pH 7.4). The C-banding method was applied for visualization of constitutive heterochromatin according to Sumner (1972). More specifically slides with chromosomal spreads were aged at 60°C for 1 h then successively soaked in 0.2 N HCl for 20 min at room temperature then in 5% $Ba(OH)_2$ solution for 4–5 min at 45°C and subsequently in 2× SSC for 1 h at 60°C, with intermediate washes in distilled water. Finally, metaphases were mounted with 4′,6-diamidino-2-phenylindole (DAPI) in mounting medium Vectashield (Vector laboratories).

Meiotic Chromosome Preparation and Immunostaining

Synaptonemal complex (SC) spreads were prepared from the testes of reproductively active males following Peters et al. (1997). Briefly, the left testis was cut into two pieces and placed in hypotonic solution (30 mM Tris, 50 mM sucrose, 17 mM trisodium citrate dehydrate and 5 mM EDTA; pH 8.2) for 50 min. The testis tissue was then disaggregated in 200 µl of 100 mM sucrose and the resulting cell suspension was applied in 40 µl drops and spread onto a slide previously treated with 1% PFA and 0.15% of Triton X100 (Sigma Aldrich). All slides were placed in a humid chamber for 90 min and washed for 2 min in 1× PBS. Slides were directly used for immunostaining.

Immunostaining was performed according to Moens et al. (1987) using the following primary antibodies: rabbit polyclonal anti-SYCP3 antibody (ab15093, Abcam) recognizing the lateral elements of the synaptonemal complex (dilution 1:200), and human anticentromere serum (CREST, 15-234, Antibodies Incorporated) binding kinetochores (dilution 1:50). The corresponding secondary antibodies were Alexa-594-conjugated goat anti-Rabbit IgG (H + L) (A32740, Invitrogen; dilution 1:200) and Alexa-488-conjugated goat anti-Human IgG (H + L) (A-11013, Invitrogen; dilution 1:200). Primary and secondary antibodies were diluted in PBT (3% BSA and 0.05% Tween 20 in 1× PBS) and incubated in a humid chamber for 90 min. Slides were washed three times in 1× PBS and dehydrated through an ethanol row (50, 70 and 96%, 3 min each). Finally, all slides were dried and stained with DAPI in mounting medium Vectashield (Vector laboratories).

Fluorescence *In Situ* Hybridization (FISH) With Telomeric and 18S rRNA Probes

Telomeric repeat probe $(TTAGGG)_n$ was applied to meiotic and mitotic spreads using FISH. In both experiments, the telomeric repeat sequences were detected using a commercial kit probe directly labelled with Cy3 (DAKO). We followed the manufacturer's instructions, with the hybridization extended to 1.5 h.

The distribution of 18S rDNA genes was analyzed on mitotic spreads using FISH. The 18S rDNA probe was generated by PCR amplification and nick-translation labelling according to the protocol of Cioffi et al. (2009). The template genomic DNA originated from a reptile species, slow-worm (*Anguis fragilis*), and the PCR product was 1,456 bp in length (sequence is provided in **Supplementary Table 1**). The probe showed high

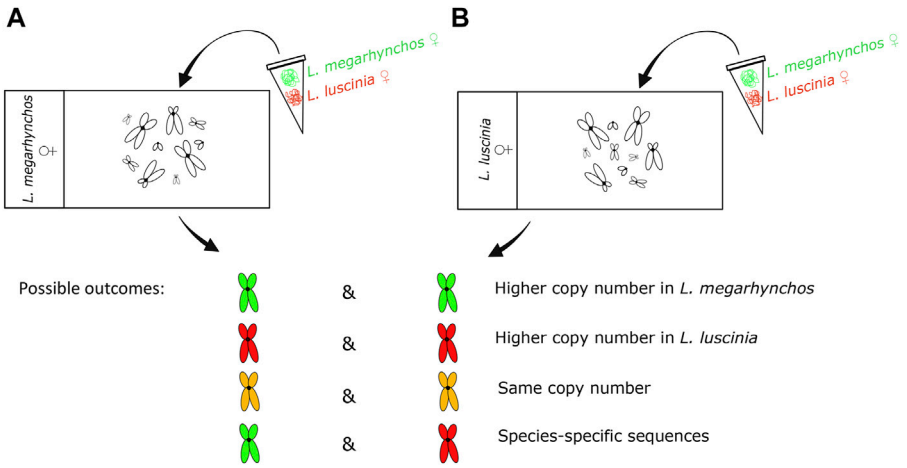

FIGURE 1 | Design and possible outcomes of the interspecific comparative genomic hybridization (CGH) experiment. The genomic probe of the common nightingale (*L. megarhynchos*) (stained green) and the thrush nightingale (*L. luscinia*) (stained red) were hybridized on the metaphase of common nightingale **(A)** and thrush nightingale **(B)**. Whereas in the first three outcomes the same repetitive sequences exist in both species and the color reflects differences in their copy number, the last outcome points to the existence of species-specific repetitive sequences.

sequence similarity with the rDNA of several bird species, *Gallus gallus* (97.73%), *Hirundo rustica* and *Motacilla alba* species (both 97.11%), and thus was considered similar enough to detect the distribution of 18S rDNA clusters in the nightingale species. Slides were aged for 1 h at 60°C, then treated with RNase for 1 h at 37°C and washed three times for 5 min in 2× SSC. Chromosomes were treated with pepsin for 3 min at 37°C and then fixed for 10 min in 1% formaldehyde solution. Slides were dehydrated in an ethanol row (70, 85 and 96%, 3 min each) and air-dried. The chromosomes were denatured in 75% formamide/2× SSC at 76°C for 3 min followed by dehydration in an ethanol row. Meanwhile, the probe was denatured at 80°C for 6 min and chilled on ice for 10 min prior to the hybridization. The probe-chromosome hybridization was performed overnight at 37°C. Post-hybridization washes were performed three times for 5 min in 50% formamide/2× SSC at 37°C and washed twice for 5 min in 2× SSC and finally for 5 min in 4× SSC/0.05% Tween 20 (Sigma Aldrich). Slides were first incubated for 30 min at 37°C with 4× SSC/5% blocking reagent (Roche), then the probe signal was detected by 4× SSC/5% blocking reagent mixed with fluorescein conjugated avidin (Vector laboratories) for 30 min at 37°C, followed by three washes in 4× SSC/0.05% Tween 20 for 5 min. Slides were then incubated with biotinylated anti-avidin (Vector laboratories) for 30 min at 37°C, followed by a second round of fluorescein conjugated avidin treatment for signal amplification. Slides were finally washed in 4× SSC/0.05% Tween 20 twice for 5 min, dehydrated through an ethanol row and air-dried. Chromosomes were stained with DAPI in mounting medium, Vectashield (Vector laboratories).

Comparative Genomic Hybridization (CGH)

The CGH experiment was performed with (i) common nightingale and (ii) thrush nightingale metaphase chromosomes. In both cases, equal concentrations of DNA probes from the common nightingale and the thrush nightingale were hybridized to the chromosomes (**Figure 1**) following the procedure described in Symonová et al. (2015) with slight modification. The DNA probe was labelled by biotin (detected by streptoavidin-FITC, green) in the common nightingale and digoxigenin (detected by antidigoxigenine-rhodopsine, red) in the thrush nightingale. In both experimental designs, the green signal suggests a higher copy number of a particular repetitive sequence in the common nightingale, while the red signal indicates a higher copy number in the thrush nightingale. An intermediate yellow/orange signal suggests the same copy number in both species. Finally, a green signal in one design, while red in the other, indicates the presence of species-specific sequence (**Figure 1**).

Genomic DNA for the preparation of probes was extracted from blood samples using DNeasy Blood and Tissue Kit (Qiagen). The probes were prepared by nick translation (Abbott Laboratories) according to the manufacturer's protocol and labelled with biotin-dUTP (Roche) and digoxigenin-dUTP (Roche). The nick translation took place at 15°C for 2 h. From each sample, 1 µg of DNA was co-precipitated overnight at −20°C with an additional 5 µl of salmon sperm DNA (10 mg/ml, Sigma Aldrich), 3 µl of 3 M sodium acetate (pH 5.2) and 2.5 volume of 96% ethanol. After precipitation, the dry pellets were resuspended in 11 µl of hybridization buffer for each slide (50% formamide in 2× SSC, 10% dextran sulfate, 10% sodium dodecyl sulfate and 1× Denhardt's buffer, pH 7.0), denatured at 86°C for 6 min and then chilled on ice for 10 min prior to hybridization.

Metaphase slides were prepared by treatment with RNase and pepsin before being fixed with 1% formaldehyde, dehydrated through an ethanol row (70, 85 and 96%, 3 min each) and air-dried. Chromosomes were then denatured in 75% formamide/2× SSC at 76°C for 3 min and dehydrated again in an ice cold ethanol raw (70, 80 and 96%, 3 min each). Finally, 11 µl of the probe mix

was applied to each slide and hybridization took place at 37°C for 48 h. The same probe mix was applied to metaphases of both nightingale species in the two CGH designs.

Post-hybridization washes were performed two and three times in 50% formamide/2× SSC and 1× SSC at 44°C, respectively. Each slide was incubated with 100 µl of 4× SSC/5% blocking reagent (Roche) at 37°C for 30 min and then with 100 µl of detection mix containing 4× SSC/5% blocking reagent, 2 µl of streptavidin-FITC (Vector Laboratories) and 1 µl of anti-digoxigenin-rhodamine (Roche) at 37°C for 1 h. The slides were subsequently washed in 4× SSC/0.01% Tween 20 (Sigma Aldrich) at 44°C, dehydrated through an ethanol row (70, 85 and 96%, 3 min each) and air-dried. Finally, the chromosomes were counterstained with DAPI in mounting medium Vectashield (Vector laboratories).

Microscopy and Image Processing

Mitotic spreads were captured with an Axio Imager Z2 microscope (Zeiss) equipped with the automatic Metafer-MSearch scanning platform and a CoolCube 1 b/w digital camera (MetaSystems). Meiotic spreads were analyzed using an Olympus BX53 fluorescence microscope (Olympus) equipped with a DP30BW digital camera (Olympus). Ikaros karyotyping software (Metasystems) was used to remove the background from the metaphase images and to arrange the karyotypes. The colors of the C-banded metaphase images were inverted. All color images were captured in black and white, and later pseudocolored and superimposed using Adobe Photoshop software (version CC 2017).

A total of 18 and 17 metaphases from bone marrow were analyzed for the common nightingale and the thrush nightingale, respectively. The W chromosome was detected by C-banding due to its heterochromatic character. The Z chromosome was identified by comparing the female (ZW) and male (ZZ) metaphases. The size of each bivalent and its arm ratio was measured using the LEVAN plugin in the program ImageJ (Schindelin et al., 2012). Depending on the position of the centromere, we distinguished for each macrochromosome whether it was telocentric, acrocentric, submetacentric or metacentric chromosomes (Levan, 1964). For microchromosomes, the telocentric/acrocentric categories and submetacentric/metacentric categories were merged as they were difficult to distinguish clearly. The Z chromosome was identified among bivalents based on its relative size to the other macrochromosomes identified on the mitotic spreads. Chromosomes were measured in 15 pachytene cells in each species.

The metaphase chromosomes with applied CGH were analyzed using Photoshop (version CC 2017). For each CGH design, three cells were analyzed and compared. The centromeric red and green signals of the nine largest macrochromosomes and the sex chromosomes were measured using the histogram color tools. Each metaphase was measured three times to reduce the technical error associated with the signal measurement. The color ratio was calculated using the median value of both colors, after their normalization using the total red and green signal color.

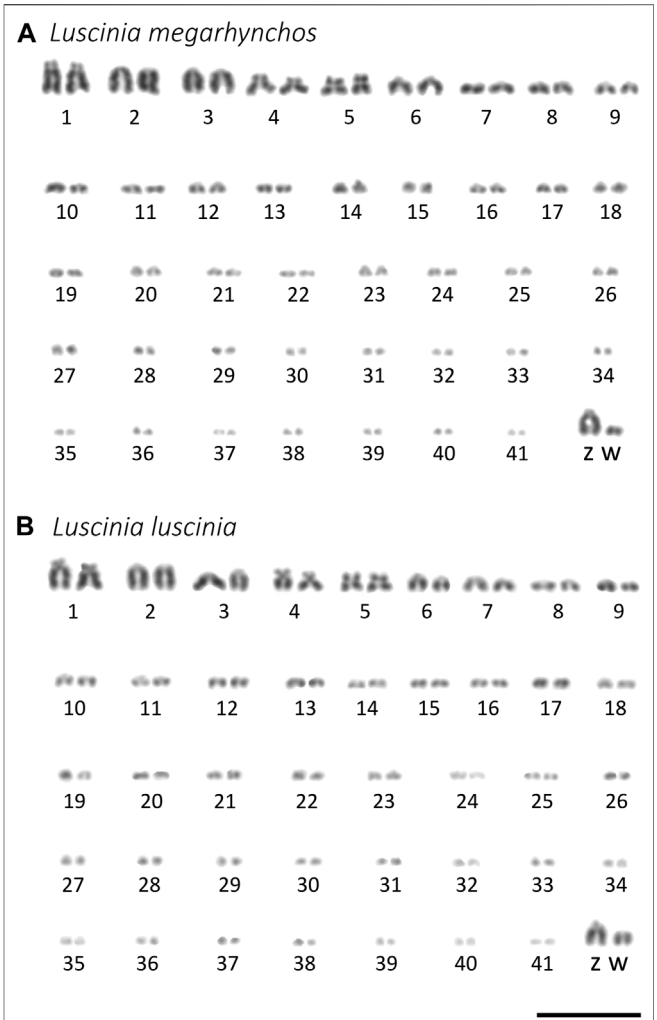

FIGURE 2 | Karyotypes of the common nightingale (*L. megarhynchos*) (**A**) and the thrush nightingale (*L. luscinia*) (**B**) females arranged after Giemsa staining. W chromosome was detected using C-banding. Scale bar = 10 µm.

RESULTS

Mitotic and Meiotic Karyotypes

Both species showed a more or less continuous decrease in chromosome size without a clear distinction between macrochromosomes and microchromosomes (**Figure 2**). We categorized the 10 largest chromosome pairs including the sex chromosomes to be macrochromosomes with the remaining chromosomes considered to be microchromosomes. Because the mitotic chromosomes were not always well spread, it was difficult to estimate the diploid chromosome number from metaphase spreads. This was especially true with respect to the number of microchromosomes. We thus calculated the diploid chromosome number for both species based on immunostained meiotic spreads (Hale et al., 1988; del Priore and Pigozzi, 2020). Both the common nightingale and the thrush nightingale consistently displayed 42 bivalents, establishing a diploid chromosome number of 84 for each species (**Figure 3**). In

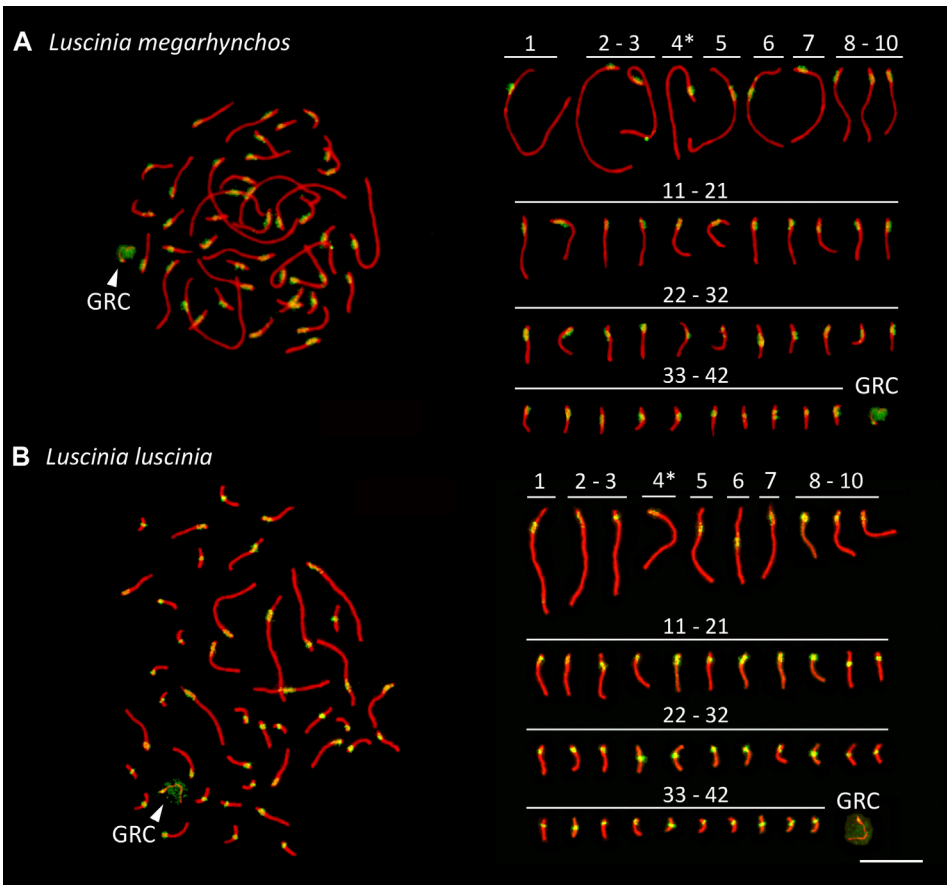

FIGURE 3 | Synaptonemal complex spreads made from testes of the common nightingale (*L. megarhynchos*) **(A)** and the thrush nightingale (*L. luscinia*) **(B)**, immunostained with antibodies against the lateral elements of the synaptonemal complex, SYCP3 (red) and against centromere proteins (green). The presumed Z chromosome bivalents are indicated with an asterisk and the germline restricted chromosome (GRC) with an arrowhead. Scale bar = 10 μm.

addition to these bivalents, both species displayed an extra univalent chromosome in male germ cells, corresponding to the GRC (**Figure 3**). The GRC was stained weaker by anti-SYCP3 antibody and showed a CREST signal not only in the centromere, but along the whole chromosome, as has been described previously in other passerine species (Torgasheva et al., 2019).

Staining of centromeres in meiotic chromosomes by the CREST antibody allowed us to estimate the arm ratio for each chromosome and compare chromosome morphology between species in a more precise way than was possible with mitotic chromosomes. The ten largest chromosomes had the same morphology between the species, suggesting that no chromosomal rearrangements that would have changed the position of the centromere occurred on these chromosomes. In both species, the largest chromosome, SC1, was identified as acrocentric, SC2 to SC4 as telocentric, SC5 as submetacentric, SC6 as metacentric and SC7 to SC9 as telocentric. However, SC10 was telocentric in the common nightingale, but acrocentric in the thrush nightingale (**Figure 3**; **Supplementary Table 2**), indicating that some rearrangements might have occurred on this chromosome.

Based on the comparison of male and female mitotic spreads, we identified the Z chromosome as the fourth largest chromosome in both species and in both species it was telocentric. The W chromosome was also telocentric, with a size between the 10th and 11th chromosome in the common nightingale and between the nineth and 10th chromosome in the thrush nightingale (**Figure 4**).

The morphology of the microchromosomes slightly differed between the two species with 17 acrocentric/telocentric and 15 submetacentric/metacentric microchromosomes in the common nightingale and 19 acrocentric/telocentric and 13 submetacentric/metacentric microchromosomes in the thrush nightingale (**Figure 3**; **Supplementary Table 2**). The GRC was present in both nightingale species, with its size corresponding to a microchromosome.

Distribution of Heterochromatin, 18S rDNA Genes and Telomeric Repeats in the Two Species

The distribution of the constitutive heterochromatin revealed by C-banding displayed the same pattern in the two nightingale

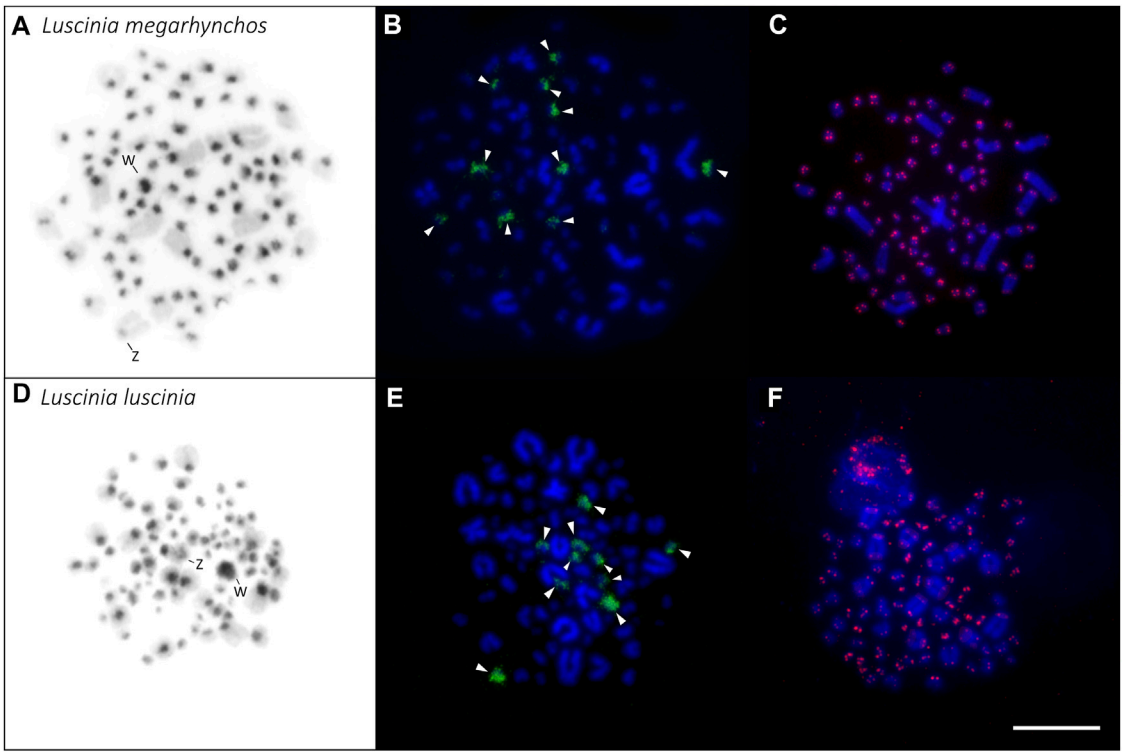

FIGURE 4 | Distribution of heterochromatin **(A,D)**, 18S rDNA clusters **(B,E)** and telomeric repeats **(C,F)** in the karyotype of the two nightingale species. C-banding in the common nightingale (*L. megarhynchos*) **(A)** and the thrush nightingale (*L. luscinia*) **(D)** female karyotypes. Sex chromosomes are indicated in both karyotypes. rDNA clusters (green) in the common nightingale **(B)** and the thrush nightingale **(E)**. Arrowheads point to 10 microchromosomes displaying a rDNA signal. Telomeric repeat sequences $(TTAGGG)_n$ (red) in the common nightingale **(C)** and the thrush nightingale **(F)**. Scale bar = 10 μm.

species. C-banding signals mainly occurred in the centromeric regions of macrochromosomes and microchromosomes, but sometimes the signal covered the entire microchromosome. The W chromosome displayed a large C-banding signal in both species, but the signal was slightly larger in the thrush nightingale than in the common nightingale (**Figure 4**). In both species, the Z chromosome had a small heterochromatic band in the centromeric region (**Figure 4**).

18S rDNA clusters were consistently located on 10 microchomosomes in both species (**Figure 4**). The same number and distribution of the 18S rDNA clusters suggests that no rearrangements that would include rDNA genes had occurred between the two species.

The telomeric motif $(TTAGGG)_n$ was detected at the terminal regions of all chromosomes. No interstitial telomeric signal was detected (**Figure 4**; **Supplementary Figure 1**). This can be seen in both the mitotic (**Figure 4**) and meiotic spreads (**Supplementary Figure 1**).

Divergence of Centromeric Repeats Between the Two Nightingale Species

In both interspecific CGH experimental designs (i.e., with the common nightingale and the thrush nightingale metaphases, **Figure 1**), the distribution pattern of the probe signal was similar to that of the heterochromatin from the C-banding experiment, meaning that the probe signal was brightest in the centromeric regions of the macrochromosomes and microchromosomes (**Figures 4**, **5**). In some microchromosomes, the whole chromosome appeared to be generating signal, however, due to the small size of chromosomes and the signal strength of the probes, this might still only represent centromeric binding.

Interestingly, the centromeric regions of the nine largest autosomes were mostly green (common nightingale probe) in the CGH with common nightingale metaphases and red (thrush nightingale probe) in the CGH with thrush nightingale metaphases, suggesting sequence divergence of repetitive elements in the centromeric regions. The exceptions were the first and fifth chromosome pairs, which showed an increased red signal in both CGH designs, indicating a higher copy number of centromeric repetitive elements in the thrush nightingale genome. The fourth pair produced variable signals across the three metaphases, making the results difficult to interpret (**Figures 5**, **6**; **Supplementary Table 3**).

The whole W chromosome displayed a higher red signal in both interspecific CGH designs, indicating a higher number of repetitive elements in the thrush nightingale genome. Contrastingly, the probe signal in the centromeric region of the Z chromosome was greener in both the common

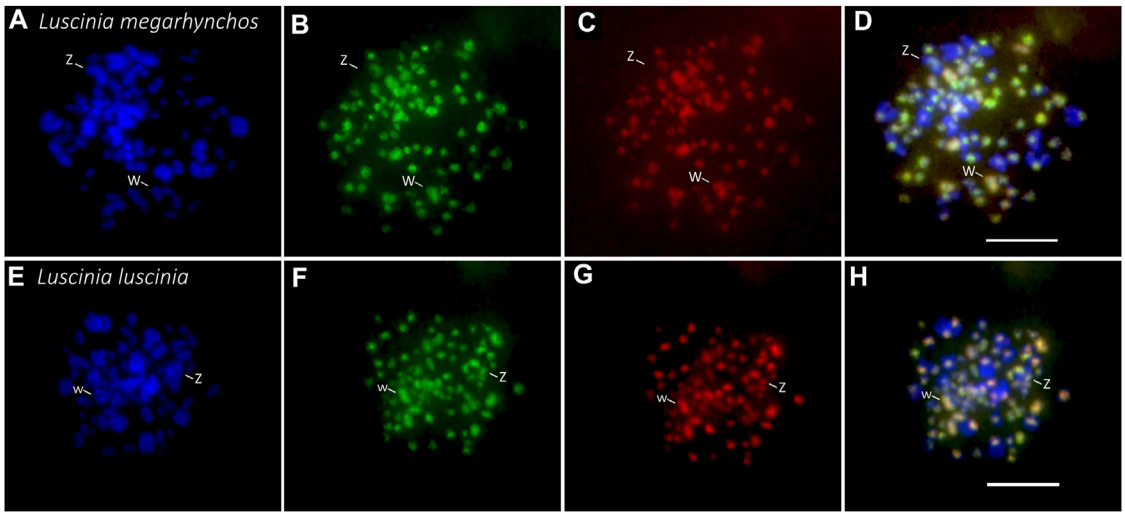

FIGURE 5 | Interspecific comparative genomic hybridization (CGH) in two nightingale species. Female probes of the common nightingale (*L. megarhynchos*) and the thrush nightingale (*L. luscinia*) were labelled by streptavidin-FITC (green) and anti-digoxigenin-rhodamine (red), respectively, and hybridized on common nightingale (**A–D**), and thrush nightingale (**E–H**) metaphase spreads. The first column displays DAPI images (blue) (**A,E**); the second column displays metaphases with the common nightingale DNA probe signal (green) (**B,F**); the third column displays metaphases with the thrush nightingale DNA probe signal (red) (**C,G**); the fourth column displays the merged colors of both genomic DNA probes and DAPI staining (**D,H**). Scale bar = 10 μm.

nightingale and the thrush nightingale experimental designs (**Figures 5**, **6**; **Supplementary Table 3**). Thus, the Z chromosome seems to have a higher copy number of centromeric repetitive elements in the common nightingale compared to the thrush nightingale.

The centromeric regions of microchromosomes were mostly green in the CGH with the common nightingale metaphases and red in the CGH with the thrush nightingale metaphases, suggesting sequence divergence of centromeric repeats on most microchromosomes (**Figure 5**).

DISCUSSION

In this study, we compared the chromosomal structure in two closely related passerine species, the common nightingale and the thrush nightingale, that show partial reproductive isolation caused mainly by hybrid female sterility and ecological differentiation (Storchová et al., 2010; Reifová et al., 2011b; Mořkovský et al., 2018; Reif et al., 2018; Sottas et al., 2018). We found that the two species have the same diploid chromosome number 2n = 84 and both possess a micro GRC in the germ cells. However, a few subtle changes in chromosome morphology imply that some chromosomal rearrangements might have occurred between the species. Interestingly, the interspecific CGH experiment suggests that the two nightingale species might have diverged in centromeric repetitive sequences on most chromosomes. Some chromosomes showed changes in copy number of centromeric repeats between the species.

Changes in chromosomal structure are assumed to play an important role in the origin of reproductive isolation. They can for example impair meiosis in hybrids leading to hybrid sterility, or suppress recombination linking together species-specific combinations of alleles, which may help to maintain species differentiation in the face of gene flow (Rieseberg, 2001; Ortíz-Barrientos et al., 2002; Butlin, 2005). The two nightingale species have very similar karyotypes, with 10 macrochromosomes (including the sex chromosomes) and 32 microchromosomes. This observed diploid chromosome number is consistent with the previously described chromosome number for the common nightingale (Bozhko, 1971). The diploid chromosome number in other species of the family Muscicapidae varies between 2n = 64 and 2n = 86 (Udagawa, 1955; Bulatova and Panov, 1973; Mittal and Satija, 1978; Bulatova, 1981; Degrandi et al., 2020a). Thus, although some large-scale chromosomal rearrangements occurred between more distantly related species of the Muscicapidae family, the closely related nightingale species seem to have the same chromosome number, which is consistent with the generally slow evolution of bird karyotypes (Christidis, 1990; Pichugin et al., 2001; Masabanda et al., 2004; Griffin et al., 2007; Ellegren, 2010; Nanda et al., 2011).

The distribution of constitutive heterochromatin blocks observed in the nightingale species is typical for passerine birds (e.g., Kretschmer et al., 2014; Barcellos et al., 2019). The larger heterochromatin block on the W in the thrush nightingale suggests that this species might have accumulated more repetitive sequences on this chromosome compared to the common nightingale. This was supported by the CGH experiment demonstrating that the W of the thrush nightingale show a higher copy number of repetitive sequences than the common nightingale. Together these results suggest the relatively fast evolution of the W repetitive content, which might

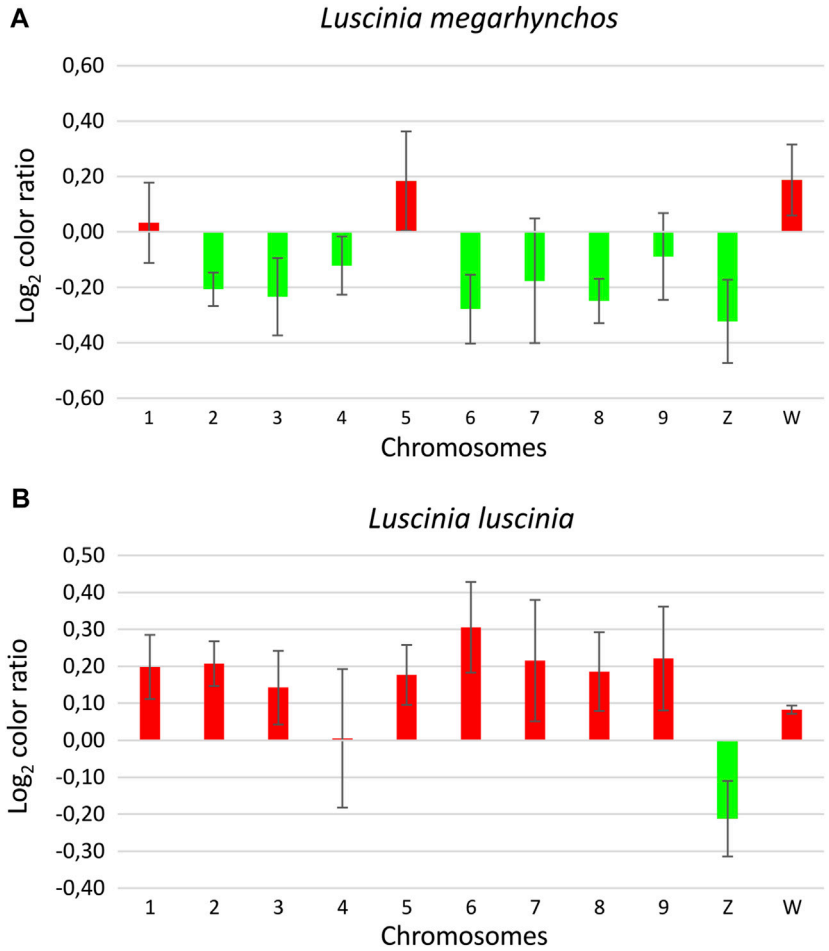

FIGURE 6 | Ratio of the green and red signal intensity at centromeric regions from the interspecific comparative genomic hybridization (CGH) experiment with the common nightingale (*L. megarhynchos*) metaphases **(A)** and the thrush nightingale (*L. luscinia*) metaphases **(B)**. Log_2 color ratio is shown for nine macrochromosomes and the sex chromosomes. Values lower than zero represent higher signal of common nightingale probes (green) and values higher than zero represent higher signal of thrush nightingale probes (red). Bar charts are based on the Log_2 ratio with error bars representing the standard error.

theoretically contribute to reproductive isolation between the species (Peona et al., 2021).

The rDNA clusters are considered as hotspots of chromosomal breakage due to their repetitive nature as well as their intense transcriptome activity (Huang et al., 2008; Cazaux et al., 2011). In birds, a large variation in the number of chromosome pairs bearing the 18S rDNA cluster is observed, ranging from one to six or seven pairs, with the majority of species displaying only one chromosome pair with rDNA cluster (Degrandi et al., 2020b). Both nightingale species showed rDNA FISH signal on five microchromosome pairs, suggesting that no rDNA associated chromosomal changes have occurred between these two species. Interestingly, five chromosome pairs bearing rDNA clusters is the highest number found in passerines so far (Degrandi et al., 2020b). In the closest related species, where rDNA has been cytogenetically localized, *Turdus rufiventris* and *Turdus albicolis*, belonging to the Turdidae family, only three and two microchromosome pairs, respectively, bear rDNA (Kretschmer et al., 2014). Such differences in the number of rDNA clusters might result from chromosome translocations, transpositions and duplications mediated by transposable elements or ectopic recombination (Nguyen et al., 2010; Teixeira et al., 2021).

Telomeric tandem repeats $(TTAGGG)_n$ are normally found at the end of chromosomes but can sometimes be present also inside the chromosomes. Such interstitial telomere sites (ITSs) may result from chromosome translocation or fusions (Nanda and Schmid, 1994; Nanda et al., 2002), although not all chromosome fusions lead to ITSs (de Oliveira et al., 2005; Nishida et al., 2008). In birds, a high number of ITSs have been found in Ratites and Galliformes (Nanda and Schmid, 1994; Nanda et al., 2002), however, in passerines only a few or no ITSs have been identified (Nanda et al., 2002; Derjusheva et al., 2004). No ITSs were detected in either of the nightingale species, providing more evidence for a conserved chromosomal structure in the two species.

It has been shown that the GRC, an extra chromosome occurring in the germline of songbirds (Pigozzi and Solari, 1998, 2005), is highly variable in its size among species (Torgasheva et al., 2019; Malinovskaya et al., 2020). Torgasheva et al. (2019) compared the size of this chromosome in 16 passerine species belonging to nine families and showed that in 10 of them the GRC is a big macrochromosome, while in six it is a small microchromosome. Our results showed that both nightingale species had a small GRC, comparable in size to a microchromosome.

Despite the same chromosome number in the two nightingale species, we observed a few small changes in the centromere position on one macrochromosome and several microchromosomes, suggesting that some intrachromosomal rearrangements might have occurred between these two species. More detailed analyses of nightingale karyotypes and their genomic sequence will be needed, however, to confirm the existence of structural variants between the two species and determine their size and content. We should also note that our cytogenetic approach cannot detect smaller chromosomal rearrangements, which do not change the position of the centromere, result in ITSs or change the number of rDNA clusters. Analysis of high-quality chromosome-level genome assemblies of the two species could shed more light on the possible smaller-scale structural changes between the species.

Another interesting difference in the chromosome structure of the two nightingale species was revealed by the interspecific CGH experiment. This experiment suggested that some chromosomes have different copy numbers of centromeric repeats between the two species. In addition, many macrochromosomes and microchromosomes displayed higher conspecific signals in the CGH experiment, suggesting that the two species have diverged in their centromeric repeat sequences. Our observation is consistent with other studies in birds (Ellegren et al., 2012), as well as other taxa (Haaf and Willard, 1997; Bensasson et al., 2008; Pertile et al., 2009; de Sassi et al., 2021; Oliveira et al., 2021), showing fast evolution of centromeric sequences. For example, comparison of the whole genome sequences of two closely related species of *Ficedula* flycatchers, which also belong to the Muscicapidae family, revealed that the centromeres were among the most differentiated regions of the genome between the species (Ellegren et al., 2012).

The rapid divergence of the centromeric sequences or their copy number, between species is assumed to be the result of centromere-associated female meiotic drive, where some centromeric sequences can bias their transmission to the egg, leaving others to end up in the polar bodies (Henikoff, 2001; Pardo-Manuel de Villena and Sapienza, 2001). This can lead to a swift fixation of particular centromeric sequences in the population and a fast divergence of centromeric repeats, or their copy number, between the species. However, distorting the transmission ratio can sometimes be harmful to the organism, for example, if it is linked to the sex chromosomes and leads to a sex ratio distortion. In such cases, it is often associated with the evolution of drive suppressors (McLaughlin and Malik, 2017). Interestingly, while most large autosomes showed species-specific sequences in our CGH experiment, sex chromosomes showed differences in the copy number of centromeric repeats, but not species-specific sequences. This suggests that the evolution of centromeric repeats on the sex chromosomes might be constrained by the sex ratio effect of sex chromosome linked meiotic drive.

It has been demonstrated that divergence of centromeric sequences between species may lead to female meiotic drive in interspecific hybrids (Chmátal et al., 2014; Akera et al., 2019; Knief et al., 2020). Theoretically, the divergence of centromeres could, in an extreme case, also cause the sterility of female hybrids and thus contribute to reproductive isolation between species (Hurst and Pomiankowski, 1991; Phadnis and Orr, 2009; Zhang et al., 2015). In nightingales, consistent with Haldane's rule (Haldane, 1922), F_1 hybrid females are sterile, while F_1 males are fertile (Reifová et al., 2011b; Mořkovský et al., 2018). It is thus possible that divergence in centromeric sequences between the two nightingale species could contribute to female-limited hybrid sterility. Further studies of centromere composition in the two nightingale species should be done to explore this possibility.

In conclusion, although the two nightingale species have very similar karyotypes, it is possible that a small number of chromosomal rearrangements have occurred between them and may contribute to reproductive isolation between the species. Interestingly, the two species appear to differ in their centromeric sequences. Such divergence could cause female meiotic drive or female sterility in interspecific hybrids. Further studies are, however, needed to confirm the presence of structural variants and diverged centromeric repeats in the two nightingale species and to examine their potential role in the nightingales' speciation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The animal study was approved by the General Directorate for Environmental Protection, Poland (permission no. DZP-WG.6401.03.123.2017.dl.3).

AUTHOR CONTRIBUTIONS

RR, MJP, and MP conceived the research, JR, TA, MP, and TO captured birds in the field, TA, RR, and MP performed the dissections, MP, MJP, MA, DD, and ZM performed the molecular cytogenetic analyses, MP, MJP, MA, and RR analyzed and interpreted the data, MP, RR, and MJP drafted the manuscript with contribution of all co-authors. All authors have read and agreed on the final version of the manuscript.

FUNDING

This research was funded by the Grant Agency of Charles University (grant 1169420 to MP), the Czech Science

Foundation (grants 18–14325S, 20-23794S to RR and TA), and the Charles University grant PRIMUS/19/SCI/008 to RR. MA was supported by Charles University Research Centre program (204069).

ACKNOWLEDGMENTS

We are grateful to Pavel Kverek and Lucie Baránková for help with the sample collection. We would like also to express our gratitude to Stephen Schlebusch, Manuelita Sotelo, Camille Sottas, Jakub Rídl and Michail Rovatsos as well as two reviewers for their comments and suggestions on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.768987/full#supplementary-material

REFERENCES

Akera, T., Trimm, E., and Lampson, M. A. (2019). Molecular Strategies of Meiotic Cheating by Selfish Centromeres. *Cell* 178, 1132–1144. doi:10.1016/j.cell.2019.07.001

Albrecht, T., Opletalová, K., Reif, J., Janoušek, V., Janoušek, L., Cramer, E. R. A., et al. (2019). Sperm Divergence in a Passerine Contact Zone: Indication of Reinforcement at the Gametic Level. *Evolution* 73, 202–213. doi:10.1111/evo.13677

Aslam, M. L., Bastiaansen, J. W., Crooijmans, R. P., Vereijken, A., Megens, H.-J., and Groenen, M. A. (2010). A SNP Based Linkage Map of the turkey Genome Reveals Multiple Intrachromosomal Rearrangements between the Turkey and Chicken Genomes. *BMC Genomics* 11, 647. doi:10.1186/1471-2164-11-647

Auer, H., Mayr, B., Lambrou, M., and Schleger, W. (1987). An Extended Chicken Karyotype, Including the NOR Chromosome. *Cytogenet. Cel Genet.* 45, 218–221. doi:10.1159/000132457

Axelsson, E., Webster, M. T., Smith, N. G. C., Burt, D. W., and Ellegren, H. (2005). Comparison of the Chicken and turkey Genomes Reveals a Higher Rate of Nucleotide Divergence on Microchromosomes Than Macrochromosomes. *Genome Res.* 15, 120–125. doi:10.1101/gr.3021305

Barcellos, S. A., Kretschmer, R., de Souza, M. S., Costa, A. L., Degrandi, T. M., dos Santos, M. S., et al. (2019). Karyotype Evolution and Distinct Evolutionary History of the W Chromosomes in Swallows (Aves, Passeriformes). *Cytogenet. Genome Res.* 158, 98–105. doi:10.1159/000500621

Bensasson, D., Zarowiecki, M., Burt, A., and Koufopanou, V. (2008). Rapid Evolution of Yeast Centromeres in the Absence of Drive. *Genetics* 178, 2161–2167. doi:10.1534/genetics.107.083980

Bi, K., and Bogart, J. P. (2006). Identification of Intergenomic Recombinations in Unisexual Salamanders of the Genus *Ambystoma* by Genomic *In Situ* Hybridization (GISH). *Cytogenet. Genome Res.* 112, 307–312. doi:10.1159/000089885

Bozhko, S. I. (1971). Karyotypes of Two Bird Species, Nightingale (*Luscinia megarhynchos* Brehm) and Song Trush (*Turdus philomelos* Brehm). *Acta Biologica Debrecina* 9, 131–136.

Bruders, R., Van Hollebeke, H., Osborne, E. J., Kronenberg, Z., Maclary, E., Yandell, M., et al. (2020). A Copy Number Variant Is Associated with a Spectrum of Pigmentation Patterns in the Rock Pigeon (*Columba livia*). *Plos Genet.* 16, e1008274. doi:10.1371/journal.pgen.1008274

Bulatova, N. S. (1981). *A Comparative Karyological Study of Passerine Birds*, 15. Praha: Academia Nakladatelstvi Československé akademie věd, 1–44.

Bulatova, N. S., and Panov, E. N. (1973). Comparative Analysis of Karyotypes of 18 Species Family Turdidae (Aves). *Caryologia* 26, 229–244. doi:10.1080/00087114.1973.10796539

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

Butlin, R. K. (2005). Recombination and Speciation. *Mol. Ecol.* 14, 2621–2635. doi:10.1111/j.1365-294X.2005.02617.x

Cazaux, B., Catalan, J., Veyrunes, F., Douzery, E. J., and Britton-Davidian, J. (2011). Are Ribosomal DNA Clusters Rearrangement Hotspots? A Case Study in the Genus *Mus* (Rodentia, Muridae). *BMC Evol. Biol.* 11, 124. doi:10.1186/1471-2148-11-124

Chmátal, L., Gabriel, S. I., Mitsainas, G. P., Martínez-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

Christidis, L. (1990). *Animal Cytogenetics/Vol. 4, Chordata. 3. B, Aves/by Les Christidis*. Editor B. John (Berlin: Gebrüder Bornträger).

Cioffi, M. B., Martins, C., Centofante, L., Jacobina, U., and Bertollo, L. a. C. (2009). Chromosomal Variability Among Allopatric Populations of Erythrinidae Fish *Hoplias malabaricus*: Mapping of Three Classes of Repetitive DNAs. *Cytogenet. Genome Res.* 125, 132–141. doi:10.1159/000227838

de M. C. Sassi, F., Perez, M. F., Oliveira, V. C. S., Deon, G. A., de Souza, F. H. S., Ferreira, P. H. N., et al. (2021). High Genetic Diversity Despite Conserved Karyotype Organization in the Giant Trahiras from Genus *Hoplias* (Characiformes, Erythrinidae). *Genes* 12, 252. doi:10.3390/genes12020252

de Oliveira, E. A., Bertollo, L. A. C., Rab, P., Ezaz, T., Yano, C. F., Hatanaka, T., et al. (2019). Cytogenetics, Genomics and Biodiversity of the South American and African Arapaimidae Fish Family (Teleostei, Osteoglossiformes). *PLoS ONE* 14, e0214225. doi:10.1371/journal.pone.0214225

de Oliveira, E. H. C., Habermann, F. A., Lacerda, O., Sbalqueiro, I. J., Wienberg, J., and Müller, S. (2005). Chromosome Reshuffling in Birds of Prey: the Karyotype of the World's Largest eagle (Harpy eagle, *Harpia harpyja*) Compared to that of the Chicken (*Gallus gallus*). *Chromosoma* 114, 338–343. doi:10.1007/s00412-005-0009-5

Degrandi, T. M., Barcellos, S. A., Costa, A. L., Garnero, A. D. V., Hass, I., and Gunski, R. J. (2020a). Introducing the Bird Chromosome Database: An Overview of Cytogenetic Studies in Birds. *Cytogenet. Genome Res.* 160, 199–205. doi:10.1159/000507768

Degrandi, T. M., Gunski, R. J., Garnero, A. d. V., Oliveira, E. H. C. d., Kretschmer, R., Souza, M. S. d., et al. (2020b). The Distribution of 45S rDNA Sites in Bird Chromosomes Suggests Multiple Evolutionary Histories. *Genet. Mol. Biol.* 43, e20180331. doi:10.1590/1678-4685-GMB-2018-0331

del Priore, L., and Pigozzi, M. I. (2020). MLH1 Focus Mapping in the guinea Fowl (*Numida meleagris*) Give Insights into the Crossover Landscapes in Birds. *PLoS ONE* 15, e0240245. doi:10.1371/journal.pone.0240245

Derjusheva, S., Kurganova, A., Habermann, F., and Gaginskaya, E. (2004). High Chromosome Conservation Detected by Comparative Chromosome Painting in Chicken, Pigeon and Passerine Birds. *Chromosome Res.* 12, 715–723. doi:10.1023/B:CHRO.0000045779.50641.00

dos Santos, M. d. S., Kretschmer, R., Silva, F. A. O., Ledesma, M. A., O'Brien, P. C. M., Ferguson-Smith, M. A., et al. (2015). Intrachromosomal Rearrangements in Two Representatives of the Genus *Saltator* (Thraupidae, Passeriformes) and the Occurrence of Heteromorphic Z Chromosomes. *Genetica* 143, 535–543. doi:10.1007/s10709-015-9851-4

Ellegren, H. (2010). Evolutionary Stasis: the Stable Chromosomes of Birds. *Trends Ecol. Evol.* 25, 283–291. doi:10.1016/j.tree.2009.12.004

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., et al. (2012). The Genomic Landscape of Species Divergence in *Ficedula* Flycatchers. *Nature* 491, 756–760. doi:10.1038/nature11584

Ellegren, H. (2013). The Evolutionary Genomics of Birds. *Annu. Rev. Ecol. Evol. Syst.* 44, 239–259. doi:10.1146/annurev-ecolsys-110411-160327

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., et al. (2020). Dense Sampling of Bird Diversity Increases Power of Comparative Genomics. *Nature* 587, 252–257. doi:10.1038/s41586-020-2873-9

Griffin, D. K., Haberman, F., Masabanda, J., O'Brien, P., Bagga, M., Sazanov, A., et al. (1999). Micro- and Macrochromosome Paints Generated by Flow

Cytometry and Microdissection: Tools for Mapping the Chicken Genome. *Cytogenet. Genome Res.* 87, 278–281. doi:10.1159/000015449

Griffin, D. K., Robertson, L. B. W., Tempest, H. G., and Skinner, B. M. (2007). The Evolution of the Avian Genome as Revealed by Comparative Molecular Cytogenetics. *Cytogenet. Genome Res.* 117, 64–77. doi:10.1159/000103166

Haaf, T., and Willard, H. F. (1997). Chromosome-specific α-satellite DNA from the Centromere of Chimpanzee Chromosome 4. *Chromosoma* 106, 226–232. doi:10.1007/s004120050243

Haldane, J. B. S. (1922). Sex Ratio and Unisexual Sterility in Hybrid Animals. *Journ. Gen.* 12, 101–109. doi:10.1007/BF02983075

Hale, D. W., Ryder, E. J., Sudman, P. D., and Greenbaum, I. F. (1988). Application of Synaptonemal Complex Techniques for Determination of Diploid Number and Chromosomal Morphology in Birds. *The Auk* 4, 776–779.

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

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

Homolka, D., Ivanek, R., Capkova, J., Jansa, P., and Forejt, J. (2007). Chromosomal Rearrangement Interferes with Meiotic X Chromosome Inactivation. *Genome Res.* 17, 1431–1437. doi:10.1101/gr.6520107

Hooper, D. M., Griffith, S. C., and Price, T. D. (2019). Sex Chromosome Inversions Enforce Reproductive Isolation across an Avian Hybrid Zone. *Mol. Ecol.* 28, 1246–1262. doi:10.1111/mec.14874

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

Hooper, D. M., and Price, T. D. (2015). Rates of Karyotypic Evolution in Estrildid Finches Differ between Island and Continental Clades. *Evolution* 69, 890–903. doi:10.1111/evo.12633

Huang, J., Ma, L., Yang, F., Fei, S.-z., and Li, L. (2008). 45S rDNA Regions Are Chromosome Fragile Sites Expressed as Gaps *In Vitro* on Metaphase Chromosomes of Root-Tip Meristematic Cells in *Lolium* Spp. *PLOS ONE* 3, e2167. doi:10.1371/journal.pone.0002167

Hurst, L. D., and Pomiankowski, A. (1991). Causes of Sex Ratio Bias May Account for Unisexual Sterility in Hybrids: a New Explanation of Haldane's Rule and Related Phenomena. *Genetics* 128, 841–858. doi:10.1093/genetics/128.4.841

Iskow, R. C., Gokcumen, O., Abyzov, A., Malukiewicz, J., Zhu, Q., Sukumar, A. T., et al. (2012). Regulatory Element Copy Number Differences Shape Primate Expression Profiles. *Proc. Natl. Acad. Sci.* 109, 12656–12661. doi:10.1073/pnas.1205199109

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., et al. (2014). Whole-genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds. *Science* 346, 1320–1331. doi:10.1126/science.1253451

Kallioniemi, A., Kallioniemi, O.-P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., et al. (1992). Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors. *Science* 258, 818–821. doi:10.1126/science.1359641

King, M. (1993). *Species Evolution: The Role of Chromosome Change*. Cambridge: Cambridge University Press.

Kinsella, C. M., Ruiz-Ruano, F. J., Dion-Côté, A.-M., Charles, A. J., Gossmann, T. I., Cabrero, J., et al. (2019). Programmed DNA Elimination of Germline Development Genes in Songbirds. *Nat. Commun.* 10, 5468. doi:10.1038/s41467-019-13427-4

Knief, U., Forstmeier, W., Pei, Y., Wolf, J., and Kempenaers, B. (2020). A Test for Meiotic Drive in Hybrids between Australian and Timor Zebra Finches. *Ecol. Evol.* 10, 13464–13475. doi:10.1002/ece3.6951

Koubová, M., Pokorná, M. J., Rovatsos, M., Farkačová, K., Altmanová, M., and Kratochvíl, L. (2014). Sex Determination in Madagascar Geckos of the Genus *Paroedura* (Squamata: Gekkonidae): Are Differentiated Sex Chromosomes Indeed So Evolutionary Stable? *Chromosome Res.* 22, 441–452. doi:10.1007/s10577-014-9430-z

Kretschmer, R., Ferguson-Smith, M., and De Oliveira, E. (2018). Karyotype Evolution in Birds: From Conventional Staining to Chromosome Painting. *Genes* 9, 181. doi:10.3390/genes9040181

Kretschmer, R., Gunski, R. J., Garnero, A. D. V., Furo, I. d. O., O'Brien, P. C. M., Ferguson-Smith, M. A., et al. (2014). Molecular Cytogenetic Characterization of Multiple Intrachromosomal Rearrangements in Two Representatives of the Genus *Turdus* (Turdidae, Passeriformes). *PLoS ONE* 9, e103338. doi:10.1371/journal.pone.0103338

Levan, A. (1964). Nomenclature for Centromeric Position on Chromosomes. *Hereditas* 52, 201–220.

Malinovskaya, L. P., Zadesenets, K. S., Karamysheva, T. V., Akberdina, E. A., Kizilova, E. A., Romanenko, M. V., et al. (2020). Germline-restricted Chromosome (GRC) in the Sand Martin and the Pale Martin (Hirundinidae, Aves): Synapsis, Recombination and Copy Number Variation. *Sci. Rep.* 10, 1058. doi:10.1038/s41598-020-58032-4

Masabanda, J. S., Burt, D. W., O'Brien, P. C. M., 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

McLaughlin, R. N., and Malik, H. S. (2017). Genetic Conflicts: the Usual Suspects and beyond. *J. Exp. Biol.* 220, 6–17. doi:10.1242/jeb.148148

Minias, P., Gutiérrez, J. S., and Dunn, P. O. (2020). Avian Major Histocompatibility Complex Copy Number Variation Is Associated with Helminth Richness. *Biol. Lett.* 16, 20200194. doi:10.1098/rsbl.2020.0194

Mittal, O. P., and Satija, K. (1978). On the Somatic Chromosomes of *Saxicoloides Fulicata Combaiensis* (Latham). *Proc. 65th Ind. Sci. Cong. Ahmedabad* 3, 243.

Moens, P. B., Heyting, C., Dietrich, A. J., van Raamsdonk, W., and Chen, Q. (1987). Synaptonemal Complex Antigen Location and Conservation. *J. Cel Biol.* 105, 93–103. doi:10.1083/jcb.105.1.93

Mořkovský, L., Janoušek, V., Reif, J., Rídl, J., Pačes, J., Choleva, L., et al. (2018). Genomic Islands of Differentiation in Two Songbird Species Reveal Candidate Genes for Hybrid Female Sterility. *Mol. Ecol.* 27, 949–958. doi:10.1111/mec.14479

Nanda, I., Benisch, P., Fetting, D., Haaf, T., and Schmid, M. (2011). Synteny Conservation of Chicken Macrochromosomes 1-10 in Different Avian Lineages Revealed by Cross-Species Chromosome Painting. *Cytogenet. Genome Res.* 132, 165–181. doi:10.1159/000322358

Nanda, I., and Schmid, M. (1994). Localization of the Telomeric (TTAGGG)n Sequence in Chicken (*Gallus domesticus*) Chromosomes. *Cytogenet. Cel Genet* 65, 190–193. doi:10.1159/000133630

Nanda, I., Schrama, D., Feichtinger, W., Haaf, T., Schartl, M., and Schmid, M. (2002). Distribution of Telomeric (TTAGGG)n Sequences in Avian Chromosomes. *Chromosoma* 111, 215–227. doi:10.1007/s00412-002-0206-4

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

Nishida, C., Ishijima, J., Kosaka, A., Tanabe, H., Habermann, F. A., Griffin, D. K., et al. (2008). Characterization of Chromosome Structures of Falconinae (Falconidae, Falconiformes, Aves) by Chromosome Painting and Delineation of Chromosome Rearrangements during Their Differentiation. *Chromosome Res.* 16, 171–181. doi:10.1007/s10577-007-1210-6

Oliveira, V. C. S., Altmanová, M., Viana, P. F., Ezaz, T., Bertollo, L. A. C., Ráb, P., et al. (2021). Revisiting the Karyotypes of Alligators and Caimans (Crocodylia, Alligatoridae) after a Half-century Delay: Bridging the gap in the Chromosomal Evolution of Reptiles. *Cells* 10, 1397. doi:10.3390/cells10061397

Ortíz-Barrientos, D., Reiland, J., Hey, J., and Noor, M. A. F. (2002). Recombination and the Divergence of Hybridizing Species. *Genet. Mate Choice: Sex. Selection Sex. Isolation* 116, 167–178. doi:10.1007/978-94-010-0265-3_2

Pardo-Manuel de Villena, F., and Sapienza, C. (2001). Nonrandom Segregation during Meiosis: the Unfairness of Females. *Mamm. Genome* 12, 331–339. doi:10.1007/s003350040003

Peona, V., Palacios-Gimenez, O. M., Blommaert, J., Liu, J., Haryoko, T., Jønsson, K. A., et al. (2021). The Avian W Chromosome Is a Refugium for Endogenous Retroviruses with Likely Effects on Female-Biased Mutational Load and Genetic Incompatibilities. *Phil. Trans. R. Soc. B.* 376, 20200186. doi:10.1098/rstb.2020.0186

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., et al. (2007). Diet and the Evolution of Human Amylase Gene Copy Number Variation. *Nat. Genet.* 39, 1256–1260. doi:10.1038/ng2123

Pertile, M. D., Graham, A. N., Choo, K. H. A., and Kalitsis, P. (2009). Rapid Evolution of Mouse Y Centromere Repeat DNA Belies Recent Sequence Stability. *Genome Res.* 19, 2202–2213. doi:10.1101/gr.092080.109

Peters, A. H. F. M., Plug, A. W., Van Vugt, M. J., and Boer, P. d. (1997). A Drying-Down Technique for the Spreading of Mammalian Meiocytes from the Male and Female Germline. *Chromosome Res.* 5 (1), 66–68. doi:10.1023/a:1018445520117

Phadnis, N., and Orr, H. A. (2009). A Single Gene Causes Both Male Sterility and Segregation Distortion in *Drosophila* Hybrids. *Science* 323, 376–379. doi:10.1126/science.1163934

Pichugin, A. M., Galkina, S. A., Potekhin, A. A., Punina, E. O., Rautian, M. S., and Rodionov, A. V. (2001). Estimation of the Minimal Size of Chicken *Gallus gallus domesticus* Microchromosomes via Pulsed-Field Electrophoresis. *Russ. J. Genet.* 37, 535–538. doi:10.1023/A:1016622816552

Pigozzi, M. I., and Solari, A. J. (1998). Germ Cell Restriction and Regular Transmission of an Accessory Chromosome that Mimics a Sex Body in the Zebra Finch, *Taeniopygia guttata*. *Chromosome Res.* 6, 105–113. doi:10.1023/A:1009234912307

Pigozzi, M. I., and Solari, A. J. (2005). The Germ-Line-Restricted Chromosome in the Zebra Finch: Recombination in Females and Elimination in Males. *Chromosoma* 114, 403–409. doi:10.1007/s00412-005-0025-5

Pokorná, M., Rens, W., Rovatsos, M., and Kratochvíl, L. (2014). A ZZ/ZW Sex Chromosome System in the Thick-Tailed Gecko (*Underwoodisaurus milii*; Squamata: Gekkota: Carphodactylidae), a Member of the Ancient Gecko Lineage. *Cytogenet. Genome Res.* 142, 190–196. doi:10.1159/000358847

Reif, J., Reifová, R., Skoracka, A., and Kuczyński, L. (2018). Competition-driven Niche Segregation on a Landscape Scale: Evidence for Escaping from Syntopy towards Allotopy in Two Coexisting Sibling Passerine Species. *J. Anim. Ecol.* 87, 774–789. doi:10.1111/1365-2656.12808

Reifová, R., Kverek, P., and Reif, J. (2011b). The First Record of a Female Hybrid between the Common Nightingale (*Luscinia megarhynchos*) and the Thrush Nightingale (*Luscinia luscinia*) in Nature. *J. Ornithol* 152, 1063–1068. doi:10.1007/s10336-011-0700-7

Reifová, R., Reif, J., Antczak, M., and Nachman, M. W. (2011a). Ecological Character Displacement in the Face of Gene Flow: Evidence from Two Species of Nightingales. *BMC Evol. Biol.* 11, 138. doi:10.1186/1471-2148-11-138

Rieseberg, L. H. (2001). Chromosomal Rearrangements and Speciation. *Trends Ecol. Evol.* 16, 351–358. doi:10.1016/S0169-5347(01)02187-5

Rodionov, A. V., Myakoshina, Y. A., Chelysheva, L. A., Solovei, I. V., and Gaginskaya, E. R. (1992). Chiasmata in the Lampbrush Chromosomes of *Gallus gallus domesticus*: the Cytogenetic Study of Recombination Frequency and Linkage Map Lengths. *Genetika* 28, 53–63.

Rodrigues, B. S., Kretschmer, R., Gunski, R. J., Garnero, A. D. V., O'Brien, P. C. M., Ferguson-Smith, M., et al. (2017). Chromosome Painting in Tyrant Flycatchers Confirms a Set of Inversions Shared by Oscines and Suboscines (Aves, Passeriformes). *Cytogenet. Genome Res.* 153, 205–212. doi:10.1159/000486975

Schartl, M., Schmid, M., and Nanda, I. (2016). Dynamics of Vertebrate Sex Chromosome Evolution: from Equal Size to Giants and Dwarfs. *Chromosoma* 125, 553–571. doi:10.1007/s00412-015-0569-y

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an Open-Source Platform for Biological-Image Analysis. *Nat. Methods* 9, 676–682. doi:10.1038/nmeth.2019

Skinner, B. M., Al Mutery, A., Smith, D., Völker, M., Hojjat, N., Raja, S., et al. (2014). Global Patterns of Apparent Copy Number Variation in Birds Revealed by Cross-Species Comparative Genomic Hybridization. *Chromosome Res.* 22, 59–70. doi:10.1007/s10577-014-9405-0

Skinner, B. M., Robertson, L. B., Tempest, H. G., Langley, E. J., Ioannou, D., Fowler, K. E., et al. (2009). Comparative Genomics in Chicken and Pekin Duck Using FISH Mapping and Microarray Analysis. *BMC Genomics* 10, 357. doi:10.1186/1471-2164-10-357

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

Sottas, C., Reif, J., Kreisinger, J., Schmiedová, L., Sam, K., Osiejuk, T. S., et al. (2020). Tracing the Early Steps of Competition-Driven Eco-Morphological Divergence in Two Sister Species of Passerines. *Evol. Ecol.* 34, 501–524. doi:10.1007/s10682-020-10050-4

Sottas, C., Reif, J., Kuczyński, L., and Reifová, R. (2018). Interspecific Competition Promotes Habitat and Morphological Divergence in a Secondary Contact Zone between Two Hybridizing Songbirds. *J. Evol. Biol.* 31, 914–923. doi:10.1111/jeb.13275

Storchová, R., Reif, J., and Nachman, M. W. (2010). Female Heterogamety and Speciation: Reduced Introgression of the Z Chromosome between Two Species of Nightingales. *Evolution* 64, 456–471. doi:10.1111/j.1558-5646.2009.00841.x

Sumner, A. T. (1972). A Simple Technique for Demonstrating Centromeric Heterochromatin. *Exp. Cel Res.* 75, 304–306. doi:10.1016/0014-4827(72)90558-7

Symonová, R., Sember, A., Majtánová, Z., and Ráb, P. (2015). "Characterization of Fish Genomes by GISH and CGH," in *Fish Cytogenetic Techniques*. Editors C. Ozouf-Costaz, E. Pisano, F. Foresti, and L. de Almeida (Boca Paton: CRC Press), 118–131. doi:10.1201/b18534-17

Teixeira, G. A., Aguiar, H. J. a. C., Petitclerc, F., Orivel, J., Lopes, D. M., and Barros, L. a. C. (2021). Evolutionary Insights into the Genomic Organization of Major Ribosomal DNA in Ant Chromosomes. *Insect Mol. Biol.* 30, 340–354. doi:10.1111/imb.12699

Torgasheva, A. A., Malinovskaya, L. P., Zadesenets, K. S., Karamysheva, T. V., Kizilova, E. A., Akberdina, E. A., et al. (2019). Germline-restricted Chromosome (GRC) Is Widespread Among Songbirds. *Proc. Natl. Acad. Sci. USA* 116, 201817373. doi:10.1073/pnas.1817373116

Udagawa, T. (1955). Karyogram Studies in Birds VI. The Chromosomes of Five Species of the Turdidae. *日本動物学彙報* 28 (4), 256–261.

Völker, M., Backström, N., Skinner, B. M., Langley, E. J., Bunzey, S. K., Ellegren, H., et al. (2010). Copy Number Variation, Chromosome Rearrangement, and Their Association with Recombination during Avian Evolution. *Genome Res.* 20, 503–511. doi:10.1101/gr.103663.109

Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., et al. (2020). Discovery and Population Genomics of Structural Variation in a Songbird Genus. *Nat. Commun.* 11, 3403. doi:10.1038/s41467-020-17195-4

Wellenreuther, M., Mérot, C., Berdan, E., and Bernatchez, L. (2019). Going beyond SNPs: The Role of Structural Genomic Variants in Adaptive Evolution and Species Diversification. *Mol. Ecol.* 28, 1203–1209. doi:10.1111/mec.15066

White, M. J. D. (1978). Chain Processes in Chromosomal Speciation. *Syst. Zoolog.* 27, 285. doi:10.2307/2412880

Zhang, G., Li, B., Li, B., Li, C., Gilbert, M. T. P., Jarvis, E. D., et al. (2014). Comparative Genomic Data of the Avian Phylogenomics Project. *GigaSci* 3, 26. doi:10.1186/2047-217X-3-26

Zhang, L., Reifová, R., Halenková, Z., and Gompert, Z. (2021). How Important Are Structural Variants for Speciation? *Genes* 12, 1084. doi:10.3390/genes12071084

Zhang, L., Sun, T., Woldesellassie, F., Xiao, H., and Tao, Y. (2015). Sex Ratio Meiotic Drive as a Plausible Evolutionary Mechanism for Hybrid Male Sterility. *Plos Genet.* 11, e1005073. doi:10.1371/journal.pgen.1005073

Zhou, J., Lemos, B., Dopman, E. B., and Hartl, D. L. (2011). Copy-Number Variation: The Balance between Gene Dosage and Expression in *Drosophila melanogaster*. *Genome Biol. Evol.* 3, 1014–1024. doi:10.1093/gbe/evr023

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 Genetics

REVIEW
published: 09 December 2021
doi: 10.3389/fgene.2021.727570

Evolution of B Chromosomes: From Dispensable Parasitic Chromosomes to Essential Genomic Players

Martina Johnson Pokorná[1,2,3]* *and Radka Reifová*[1]*

[1]*Department of Zoology, Charles University, Prague, Czech Republic,* [2]*Department of Ecology, Charles University, Prague, Czech Republic,* [3]*Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Liběchov, Czech Republic*

B chromosomes represent additional chromosomes found in many eukaryotic organisms. Their origin is not completely understood but recent genomic studies suggest that they mostly arise through rearrangements and duplications from standard chromosomes. They can occur in single or multiple copies in a cell and are usually present only in a subset of individuals in the population. Because B chromosomes frequently show unstable inheritance, their maintenance in a population is often associated with meiotic drive or other mechanisms that increase the probability of their transmission to the next generation. For all these reasons, B chromosomes have been commonly considered to be nonessential, selfish, parasitic elements. Although it was originally believed that B chromosomes had little or no effect on an organism's biology and fitness, a growing number of studies have shown that B chromosomes can play a significant role in processes such as sex determination, pathogenicity and resistance to pathogens. In some cases, B chromosomes became an essential part of the genome, turning into new sex chromosomes or germline-restricted chromosomes with important roles in the organism's fertility. Here, we review such cases of "cellular domestication" of B chromosomes and show that B chromosomes can be important genomic players with significant evolutionary impact.

Keywords: evolution, cytogenetics, supernumerary chromosomes, meiotic drive, cellular domestication

OPEN ACCESS

Edited by:
Ricardo Utsunomia, Federal Rural University of Rio de Janeiro, Brazil

Reviewed by:
Andreas Houben, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany
Stacey Hanlon, University of Connecticut, United States

***Correspondence:**
Martina Johnson Pokorná
martina.pokorna@natur.cuni.cz
Radka Reifová
radka.reifova@natur.cuni.cz

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *18 June 2021*
Accepted: *25 October 2021*
Published: *09 December 2021*

Citation:
Johnson Pokorná M and Reifová R (2021) Evolution of B Chromosomes: From Dispensable Parasitic Chromosomes to Essential Genomic Players. Front. Genet. 12:727570. doi: 10.3389/fgene.2021.727570

INTRODUCTION

B chromosomes are supernumerary dispensable chromosomes that occur only in some individuals or populations within a species, or only in a subset of cells or tissues within an individual (Beukeboom, 1994; Camacho, 2005; Houben et al., 2014; Ruban et al., 2020). Their presence in a species can thus be viewed as a type of genetic polymorphism. Unlike standard chromosomes, they often show irregular non-Mendelian inheritance (Jones, 1995). B chromosomes were observed for the first time by Wilson (1907) in true bugs (Hemiptera) from the genus *Metapodius*. Soon after, similar observations were published for cucumber beetles (Coleoptera) from the genus *Diabrotica* (Stevens, 1908). In plants, such structures were first observed in rye (*Secale cereale*) and named "k-chromosomes" by Gotoh (1924), and later described in maize (*Zea mays*) by Kuwada (1925) and Longley (1927) who labelled them as "supernumerary chromosomes." The term "B chromosomes" was introduced by Randolph (1928) and has been used by the scientific community ever since. In synchrony, all the other chromosomes in the genome are referred to as "A chromosomes".

Based on the number of species studied so far it has been estimated that approximately 15% of eukaryotic species have B chromosomes (Beukeboom, 1994; D'Ambrosio et al., 2017) with new findings of B chromosomes being regularly described (reviewed in Jones, 2017). A database collecting information about B chromosomes has been available since 2017 (D'Ambrosio et al., 2017). Out of the 2,828 eukaryotic species with B chromosomes reported there, 73.56% are plants, 25.95% animals and 0.49% fungi (D'Ambrosio et al., 2017). However, it is important to note that some groups of organisms have been cytogenetically studied less extensively than others and thus the representation of B chromosomes among specific taxonomic groups is currently difficult to compare (D'Ambrosio et al., 2017).

B chromosomes can occur in single or multiple copies per cell but usually their copy number is rather low. There is normally either one or a few copies of B chromosomes functioning usually as univalents (Camacho, 2005). However, in some species, extreme numbers of B chromosomes can be observed in a single cell, such as in some plant species of the genus *Pachyphylum* (Crassulaceae), which have up to 50 B chromosome copies (Uhl and Moran, 1973). Other organisms with high numbers of B chromosomes are maize (*Z. mays*) with up to 34 copies (Jones and Rees, 1982), the wood mouse (*Apodemus peninsulae*) with up to 24 copies (Volobuev and Timina, 1980) and the *Xylota nemorum* fly with up to 24 B chromosomes copies (Boyes and van Brink, 1967). In some cases, extensive variability in the number of B chromosomes can be observed among individuals or populations within a species (Camacho et al., 2000). It has been reported that these high numbers of B chromosomes could have negative effects on their hosts, particularly on their fertility and viability (Houben, 2017), especially if the B chromosomes occur in odd numbers (Camacho et al., 2004).

B chromosomes show high variability in their size across species. They can be similar in size to A chromosomes (Jones, 2018) but also, in some species, B chromosomes are considerably smaller than the smallest A chromosomes e.g., in the harvest mouse *Reithrodonomys megalotis* (Peppers et al., 1997) or the fly *Megaselia scalaris* (Wolf et al., 1991). On the other hand, B chromosomes bigger than the biggest A chromosomes have been reported in cyprinid fish *Alburnus alburnus* (Ziegler et al., 2003), the giant white-tailed rats *Uromys caudimaculatus* (Baverstock et al., 1982) and the characid fish *Aslyanax scabripinnis* (Mestriner et al., 2000). Variation in size can also be observed within a species e.g. in the grasshopper *Eyprepocnemis plorans* (López-León et al., 1993).

B chromosomes were for a long time considered to have no important function for the carrier individual and spread mostly as genome parasites (e.g., Östergren, 1945; Van Valen, 1977). Currently, however, the view of B chromosomes is changing with many active B chromosome genes with important functions for their hosts being discovered (reviewed in Houben et al., 2014; Ruban et al., 2017, Dala Benetta et al., 2020). Based on these findings, it has been proposed that the effects of B chromosomes on their host may shift back and forth from parasitic or neutral to beneficial (Camacho et al., 2000). In some species, it has been even hypothesised that B chromosomes may became an essential, stable part of the genome turning, for example, into new sex chromosomes or chromosomes restricted to germline that became essential for viability and fertility of their carriers (Carvalho, 2002; Nokkala et al., 2003; Dalíková et al., 2017; Torgasheva et al., 2019; Imarazene et al., 2021; Lewis et al., 2021). In this review we describe mechanisms of B chromosome origin, strategies of their inheritance and give examples of the "cellular domestication" of B chromosomes, where these chromosomes provide some important functions for their hosts. We reviewed possible pathways of B chromosome evolutionary dynamics with outcomes ranging from the classical view of B chromosomes as nonessential genetic elements spreading in the population as genomic parasites to important genomic players providing benefits to their hosts.

Origin of B Chromosomes

The question of where B chromosomes come from has been puzzling researchers since their discovery. Currently, the most likely explanation is that B chromosomes originate from A chromosomes as by-products of chromosomal rearrangements or unbalanced segregation, when a chromosome fragment or an extra copy of an A chromosome may develop into a proto-B chromosome (Peters, 1981; Jones and Rees, 1982; Talavera et al., 1990; Camacho et al., 2000; Mestriner et al., 2000; Dhar et al., 2002; Bauerly et al., 2014). Once a proto-B chromosome exists it may acquire additional genetic material through duplications from other A chromosomes (Martis et al., 2012; Marques et al., 2018; Ahmad et al., 2020; Blavet et al., 2021). Throughout the evolution of B chromosomes, various mobile elements and unique coding and noncoding sequences can be incorporated, sometimes amplifying and sometimes degenerating due to the very small selection pressure on the B chromosome (reviewed in Houben et al., 2014; Marques et al., 2018). Next-generation sequencing of B chromosomes in various species confirmed that these chromosomes are largely composed of A chromosome paralogous sequences, although in some plant species organellar DNA has also been shown to contribute to the B chromosomes (Valente et al., 2014; Ruban et al., 2017; Marques et al., 2018; Ruiz Ruano et al., 2018). For example, in rye (*S. cereale*), one of the best-studied plant models for B chromosome research, B chromosomes seem to harbor A chromosome derived sequences, mostly coming from 3R to 7R autosomes, with a significant contribution of organellar DNA (Martis et al., 2012). Similarly, in the goat grass *Aegilops speltoides,* B chromosomes share sequences not only with A chromosomes but also with the DNA of plastids and mitochondria, suggesting that organelle-to-nucleus DNA transfer affects B chromosome evolution (Ruban et al., 2014; Ruban et al., 2020). The level of homology between B chromosomal and A chromosomal paralogous sequences can be used to estimate the age of B chromosomes. In maize (*Z. mays*), such comparison revealed the very ancient origin of the B chromosomes (Blavet et al., 2021).

In several species, B chromosomes appear to have originated from sex chromosomes. This seems to have happened for example in the flies from the genus *Glossina*, the New Zealand endemic frog *Leiopelma hochstetteri* (reviewed in Camacho, 2005) and the grasshopper *E. plorans,* where the B chromosome is derived from a paracentromeric region of the

X chromosome (López León et al., 1994). The involvement of sex chromosomes in the origin of B chromosomes has been also demonstrated in the rodent group Oryzomyini (Ventura et al., 2015).

Although it seems to be rare, B chromosomes may also originate through interspecific hybridization. This has been described for jewel wasps *Nasonia vitripennis*. This species-specific B chromosome known as Paternal Sex Ratio (PSR) contains transposon-like sequences that appear to be absent from the *N. vitripennis* genome, but match sequences present in another two species of wasp from the genus *Trichomalopis* (McAllister, 1995; McAllister and Werren, 1997). This observation suggests that this B chromosome has been derived from a chromosome of another species that moved into the *N. vitripennis* genome by hybridization (McAllister, 1995; McVean, 1995; McAllister and Werren, 1997; Perfectti and Werren, 2001).

B chromosomes may also have their origin in incompletely expelled A chromosome from the sperm in pseudogamous parthenogens as documented in flatworms *Polycelis nigra*. In this species, sexual individuals are always diploid while pseudogamous parthenogens are usually triploid. In parthenogenetic individuals a sperm is required only to initiate the egg development and the paternal chromosomes never enter the oocyte nucleus. In purely parthenogenetic populations of this species, B chromosomes of three distinct subtypes were found in almost all individuals. These B chromosomes come from paternal A chromosomes which escaped the exclusion of the sperm genome and have been incorporated into the nucleus (Beukeboom et al., 1996; Sharbel at al., 1997).

B Chromosomes as Genomic Parasites

Traditionally, B chromosomes have been viewed as genomic parasites that do not provide any advantage to their host and sometimes can even be detrimental if they are present in high copy numbers (Camacho et al., 2004). Because their meiotic as well mitotic inheritance may be unstable especially if they occur in an odd copy number, B chromosomes evolved diverse ways to promote their own transmission, preventing their loss from the population. These include meiotic drive, mitotic drive associated with gonotaxis and preferential fertilization of the ovum by B chromosome carrying spermatozoa.

Meiotic drive (**Figure 1A,B**) promoting the transmission of B chromosomes can occur during female as well as male meiosis. Female meiotic drive (**Figure 1A**) is, however, more common (Jones, 2018). It stems from the asymmetry of female meiosis, where one ovum and three polar bodies (which do not participate in inheritance) are produced from a single diploid oogonia. Many B chromosomes have been shown to have a mechanism helping them to end up in the ovum rather than in a polar body (reviewed in Jones, 2018). Female meiotic drive is often associated with specific centromeric sequences, as centromeres bind to the meiotic spindle during chromosome segregation (Padro-Manuel De Villena and Sapienza, 2001). However, sometimes the total number of centromeres in each side of the meiotic spindle, rather than specific centromeric sequences, determine which chromosomes will end up in the egg and which in polar bodies (Padro-Manuel De Villena and Sapienza, 2001). If a higher number of centromeres preferentially segregate to the egg, the presence of an additional B chromosome bringing one extra centromere, may cause B chromosomes to preferentially segregate to the gamete. Generally, such a type of meiotic drive based on the total number of centromeres may also lead to a chromosome fission and the origin of acrocentric chromosomes from metacentric chromosomes (Padro-Manuel De Villena and Sapienza, 2001. Interestingly, it has been observed in mammals and insects that B chromosomes occur more often in species with acrocentric rather than metacentric chromosomes (Bidau and Martí, 2004; Palestis et al., 2004; Palestis et al., 2010), suggesting that female meiotic drive based on total number of centromeres may help to spread B chromosomes in a population. However, we cannot rule out the possibility that acrocentric chromosomes simply generate B chromosomes more frequently than metacentric chromosomes.

B chromosomes can also increase their transmission by male meiotic drive (**Figure 1B**), where haploid cells without B chromosomes do not survive, although it seems to be much rarer than female meiotic drive (Jones and Rees, 1982). A specific case of male meiotic drive was described in the mealybug *Pseudococcus affinis*. In this species, paternally inherited chromosomes become heterochromatic during early embryonic development. During male meiosis, paternal and maternal chromosomes segregate to different poles and only meiotic products with euchromatic maternal chromosomes form functional sperm. The B chromosome, although paternally inherited, segregates with the maternal chromosomes and ends up in the functional sperm (Nur, 1962).

Mitotic drive leading to the preferential segregation of B chromosomes to the germline (gonotaxis) (**Figure 1C,D**) may also increase the chance of B chromosome transmission to the next generation. At the same time, it can lead to the multiplication of B chromosome copies in the cell. Mitotic drive occurs through the nondisjunction of B chromosomes during mitosis (Jones, 1991). It can happen before meiosis during the germline cell division and in plants also after meiosis during the gametophytic phase. Premeiotic mitotic drive (**Figure 1C**) was described for the first time in the grasshoppers *Calliptamus palaestinensis* and *Cammula pelucida* (Nur, 1963; Nur, 1969) and leads to an amplification of B chromosomes in spermatocytes. Similar phenomenon was later observed in other animals and plants (Rutishauser and Röthlisberger, 1966; Kayano, 1971; Viseras et al., 1990; Pardo et al., 1995; Jones, 2018). The postmeiotic mitotic drive (**Figure 1D**) is known from plants which have, compared to animals, a haploid gametophyte phase (reviewed by Houben, 2017). This type of drive was first observed by Hasegawa (1934) in rye, where the B chromosome moved with its two nondisjunct chromatids towards the generative nucleus in the first pollen grain mitosis. The nondisjunction in rye is controlled by the region on the long arm of the B chromosome where various tandem repeats have been identified (reviewed in Houben et al., 2014; Marques et al., 2018).

In some plants, the overtransmission of B chromosomes can be caused by preferential fertilization of the ovum by the B-carrying spermatozoid (Jones, 2018). This has been observed

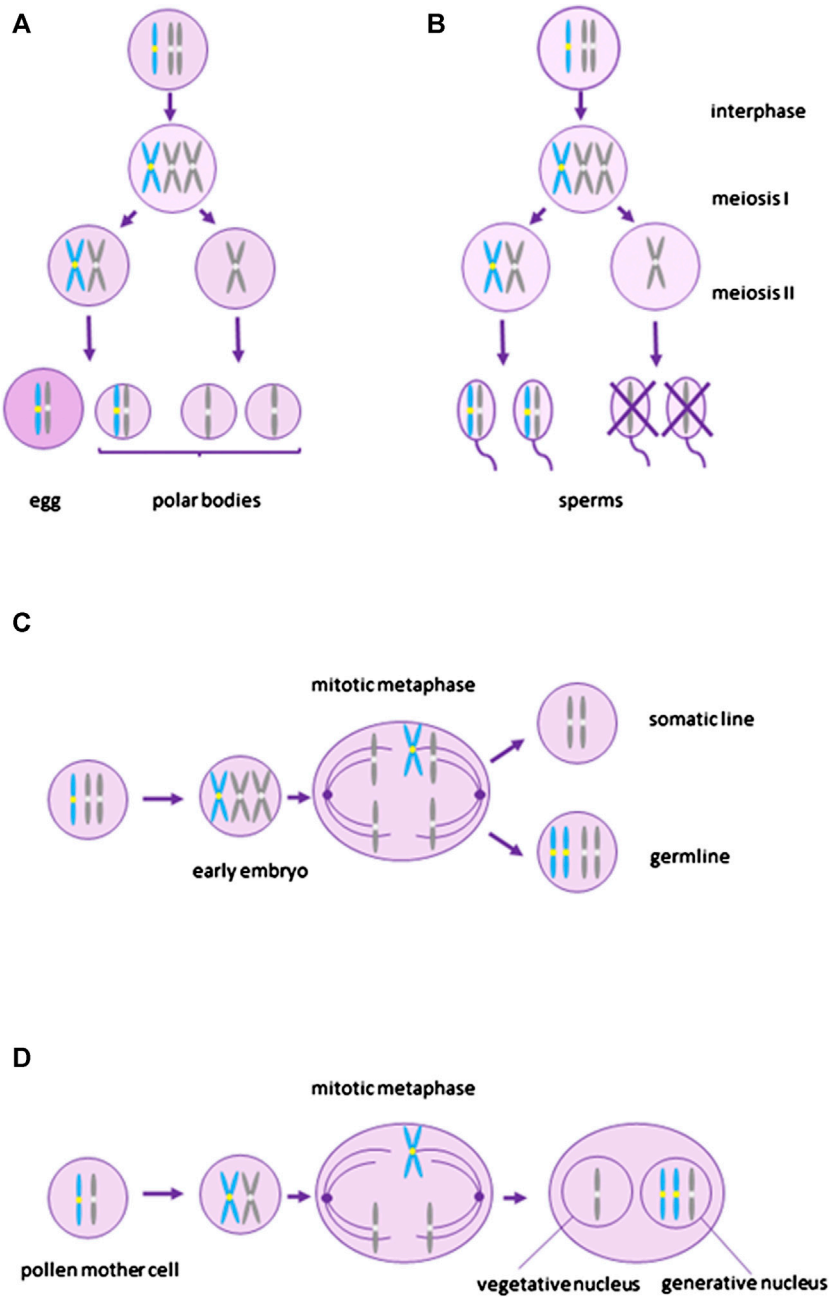

FIGURE 1 | Depiction of transmission mechanisms of B chromosomes. **(A)** and **(B)** represent meiotic drive. **(A)**–female meiotic drive where the B chromosome segregates preferentially into the egg, **(B)**–male meiotic drive where sperms without B do not survive. **(C)** and **(D)** represent mitotic drive associated with gonotaxis where B chromosomes preferentially segregate into the germline **(C)**–premeiotic mitotic drive during early embryo development when the germline is being determined, **(D)**–postmeiotic mitotic drive in plants during gametophytic phase. Blue represents B chromosomes and grey represents A chromosomes.

for example in maize, where mitotic drive at the second pollen mitosis caused by nondisjunction of B chromosome chromatids results in two unequal sperms. The egg is then preferentially fertilized by the sperm bearing the B chromosome during double fertilization (Blavet et al., 2021).

Although different mechanisms of unequal transmission already play very powerful roles in B chromosome inheritance, some B chromosomes found a way to enhance the drive effect. For example, some cichlid fish from Lake Victoria and Malawi carry a female-specific B chromosome (Yoshida et al., 2011; Clark and Kocher, 2019), which exhibits meiotic drive ending up in more than half of oocytes. Interestingly, offspring of B-transmitting females show a strong female biased sex ratio and genotyping of these offspring revealed that the B chromosome carries a female sex determining gene that is epistatically dominant to an XY system. Therefore, the

outcome is that an XY fish with a B chromosome becomes a female and an XY fish without a B chromosome becomes a male. It has been suggested that the new sex determining function of the B chromosome evolved to enhance the female meiotic drive of the B chromosome without providing any beneficial effect to the host (Clark and Kocher, 2019).

Another example of the B chromosome, which can manipulate the sex of its carrier to enhance its own transmission, is PSR known from the jewel wasp *N. vitripennis* and other haplodiploid arthropods (Werren et al., 1987; Werren, 1991; Werren and Beukeboom, 1993; Werren and Stouthamer, 2003). PSR is transmitted strictly paternally and causes the complete elimination of the paternal chromosomes (except for the PSR itself) after fertilization (Reed and Werren, 1995; Dalla Benetta et al., 2020). As a result, all diploid fertilized eggs which would normally develop into females are turned into males causing an extremely male biased sex ratio (Werren and Beukeboom, 1993; Werren and Stouthamer, 2003).

B chromosomes might also be involved in the genetic control of apomixis (e.g., asexual reproduction via seeds). This has been described in *Boechera stricta* and *B. holboellii* (Brassicaceae). In these species, there is an additional *Het* chromosome which in some cases went through fission resulting in a largely heterochromatic *Het'* chromosome in all apomictic individuals and a smaller *Del* chromosome in aneuploid apomictic individuals. Because these chromosomes are present exclusively in apomictic plants, it has been suggested that they could play a role in the genetic control of apomixis (Sharbel et al., 2004; Kantama et al., 2007; Mandáková et al., 2015). As asexual reproduction can ensure maintaining the stable combination of chromosomes, B chromosomes involved in the transition to asexuality could theoretically gain advantage in their own transmission. However, Mandáková et al. (2021) argued that these chromosomes may be more a consequence of apomixis rather than its cause.

New genomic approaches enabling the sequencing of B chromosomes allow the identification of specific genes and regions on the B chromosome causing the drive (Dalla Benetta et al., 2020). A nice example of a gene involved in its own drive is the *haploidizer* located on PSR in *N. vitripennis* which is responsible for the elimination of paternal chromosomes except for PSR (Dalla Benetta et al., 2020). Banaei-Moghaddam et al. (2012) and Banaei-Moghaddam et al. (2013) identified the B-specific centromeric sequence responsible for the extended cohesion of the B chromatids during the first pollen mitosis and their preferential segregation to the generative nucleus in rye. Using whole B chromosome assembly, Blavet et al. (2021) determined regions on the maize B chromosome including B chromosome-specific repeat concentrated around the centromere and *trans*-acting factors on the long arm involved in B chromosome nondisjunction at the second pollen mitosis. Interestingly, centromeric region also played a role in a preferential fertilization of the egg by sperm carrying B-specific centromere (Blavet et al., 2021). In addition, in several organisms, genes involved in cell cycle, cell division, chromosome segregation or cell differentiation have been identified on B chromosomes (Graphodatsky et al., 2005; Makunin et al., 2016; Navarro Domínguez et al., 2017; Makunin et al., 2018; Marques et al., 2018; Kichigin et al., 2019; Martins and Jehangir, 2021). These might represent candidate genes for B chromosome drive. The origin of such sequences promoting the B chromosome transmission is assumed to play an important role during early B chromosome evolution (Houben et al., 2014). Without such sequences the B chromosome is likely destined to be lost from the population if it does not provide any advantage to the host (Camacho et al., 2000; Ahmad et al., 2020).

Cellular Domestication of B Chromosomes

As described in the previous section, for a long time B chromosomes were mostly viewed as nonessential selfish elements which had either no effect on the host or a negative one, and spread through the population solely as genome parasites. However, there is accumulating evidence that B chromosomes in many organisms carry important functions for their hosts, which can help them to spread in the population without selfish mechanisms such as drive (**Table 1**).

A well-known case where a B chromosome is beneficial for its carrier can be seen in the chive plant *Allium schoenoprasum.* It has been observed that individuals with B chromosomes have better survival rates in their natural habitats than those without them. These B chromosomes have a positive effect on the development from seeds as they boost the germination rate in drought conditions (Holmes and Bougourd, 1989; Plowman and Bougourd, 1994) It is interesting that in *A. schoenoprasum* no mechanisms of drive have been found (Bougourd and Parker, 1979; Holmes and Bougourd, 1989), suggesting that the presence of B chromosomes in a population may be maintained by their positive effect on the carrier. A higher survival rate of plants with B chromosome compared to those without it was observed also in ryegrass, *Lolium peremne* (Rees and Hutchinson, 1974). In rye a comparison between individuals with and without B chromosomes suggests that rye B chromosomes may play a role in heat tolerance and may protect plants against damage caused by heat stress (Pereira et al., 2017).

In the fungus *Nectria haematococca,* resistance to antibiotics, which are naturally produced by pea plants, is associated with the presence of B chromosomes which show stable inheritance under certain conditions (Miao et al., 1991; Enkerli et al., 1997). These B chromosomes thus increase the pathogenicity of their host. Observations of an increase in pathogenicity under the influence of a B chromosome has also been demonstrated in other fungi species e.g., in *Magnaporthe oryzae*, *Fusarium oxysporum*, *F.* sp. *radicis-cucumerinum, Alternaria alternata, Cochliobolus heterostrophus* and *Leptoshaeria maculans* (Ma et al., 2010; Balesdent et al., 2013; van Dam et al., 2017). Interestingly, sequencing of the fungal B chromosomes revealed that these chromosomes display different genomic properties compared to the A chromosomes, including faster evolutionary rates, higher density of transposable elements and more gene duplications (Vanheule et al., 2016; Yang et al., 2020). It has been suggested that such a "two-speed" genome may bring an advantage to the pathogens by allowing more rapid adaptations to the host and new environments through more frequent mutations on the faster evolving B

TABLE 1 | List of examples where B chromosomes have a beneficial or necessary function for their hosts.

Species	Group	Function	Reference
Nectria haematococca	fungus (Ascomicota)	resistance to antibiotics increased pathogenicity on pea roots utilization of unique carbon/nitrogen sources	Miao et al. (1991) Enkerli et al. (1997) Rodriguez-Carres et al. (2008) Coleman et al. (2009)
Magnaporthe oryzae	fungus (Ascomicota)	increased pathogenicity	Ma et al. (2010) Balesdent et al. (2013) van Dam et al. (2017)
Fusarium oxysporum	fungus (Ascomicota)	increased pathogenicity	Armitage et al. (2018) Balesdent et al. (2013) Ma et al. (2010) Thatcher et al. (2016) Williams et al. (2016) van Dam et al. (2017)
Fusarium sp. *radicis-cucumerinum*	fungus (Ascomicota)	increased pathogenicity	Ma et al. (2010) Balesdent et al. (2013) van Dam et al. (2017)
Alternaria alternata	fungus (Ascomicota)	increased pathogenicity caused by production of host-specific toxins	Hatta et al. (2002) Akagi et al. (2009) Ma et al. (2010) Balesdent et al. (2013) van Dam et al. (2017)
Cochliobolus heterostrophus	fungus (Ascomicota)	increased pathogenicity	Ma et al. (2010) Balesdent et al. (2013) van Dam et al. (2017)
Leptoshaeria maculans	fungus (Ascomicota)	increased pathogenicity	Ma et al. (2010) Balesdent et al. (2013) van Dam et al. (2017)
Avena sativa	plant (Poaceae)	resistance to rust	Dherawattana and Sadanaga (1973)
Lolium peremne	plant (Poaceae)	higher survival rate	Williams (1970) Rees and Hutchinson (1974)
Allium schoenoprasum	plant (Amaryllidaceae)	boost of the germination rate	Holmes and Bougourd (1989) Plowman and Bougourd (1994)
Secale cereale	plant (Poaceae)	heat tolerance	Pereira et al. (2017)
Tischeria ekebladella	moth (Lepidoptera)	sex chromosome	Dalíková et al. (2017)
Plutella xylostella	moth (Lepidoptera)	sex chromosome	Dalíková et al. (2017)
Cameraria ohridella	moth (Lepidoptera)	sex chromosome	Dalíková et al. (2017)
Dryas iulia	butterfly (Lepidoptera)	sex chromosome	Lewis et al. (2021)
Cacopsylla peregrina	plant lice (Psylloidea, Homoptera)	sex chromosome	Nokkala et al. (2003)
Astyanax mexicanus	cavefish (Actinopterygii)	sex chromosome	Imarazene et al. (2021)
numerous species	passerine birds (Passeriformes)	germline-restricted chromosome	Torgasheva et al. (2019) Kinsella et al. (2019)

chromosomes (Croll and McDonald 2012). B chromosomes can also increase the resistance of their host to pathogens, such as, for example, in the oat plant, *Avena sativa,* where they increase the resistance of the plants to rust caused by the fungus *Puccinia coronata* f. sp. *avenae* (Dherawattana and Sadanaga, 1973).

It has been proposed that B chromosomes can become involved in sex determination and start to function as sex chromosomes (e.g., Camacho, 2005). Although, this crucial function can at the beginning represent a way how to enhance the B chromosome transmission as has been described in some cichlid fish (Clark and Kocher, 2019), it can later become essential for the host. In addition, the B chromosome can theoretically turn into a sex chromosome when it starts to pair with the X or Z in the system where Y or W is not present (i.e. X0 or Z0). Possible examples of such B to sex chromosome transition have been described in moth and butterfly species (Lepidoptera). Lukhtanov (2000) mentioned examples, which may represent different phases of such evolutionary transition, from the stage where a newly formed W chromosome (originally B chromosome) is still dispensable and individuals with or without it can be found in a population, to the stage where the W chromosome has the sex determining function and is fully fixed and essential. Although these examples support the possibility of the sex chromosome origin within B chromosomes, Lukhtanov (2000) provides alternative explanations of these observations. Also, in the family Tischeriidae and in the clade Ditrysia (Lepidoptera), the W possibly arose from a B chromosome which started to pair with the Z chromosome (Dalíková et al., 2017). However, a possibility that at least in Tischeriidae W chromosome evolved from the fusion of a Z chromosome and an autosome has not been ruled out (Dalíková et al., 2017; Hejníčková et al., 2019). Recent data from genome sequencing in another butterfly

species *Dryas iulia* (tribe Heliconiini) support the origin of a W sex chromosome from a B chromosome and suggest that this event may have happened multiple times during the evolution of butterflies (Lewis et al., 2021). It has also been proposed that the ancestral Y chromosome in *Drosophila* may have originated from a B chromosome (Hackstein et al., 1996; Carvalho, 2002). None of the Y-linked genes in *Drosophila* have homologs on the X chromosome and all identified paralogs are autosomal. This implies that the origin and evolutionary history of the Y chromosome is different than simply being a degenerated counterpart of the X chromosome and Carvalho (2002) proposed that it could have its origin in a B chromosome that became a segregational partner of the X chromosome in an X0 system. However, it is also possible that the present Y in *Drosophila* could be an outcome of a fusion of an ancestral Y chromosome with an autosome (Bachtrog, 2013). In the plant lice *Cacopsylla peregrina*, the Y chromosome has most likely evolved from a B chromosome that was integrated into a segregation system with the X chromosome and later became fixed in the karyotype (Nokkala et al., 2003). In some cases, such as in the cichlid fish from the tribe Oreochromini, the B chromosome fused with the original sex chromosomes and become a stable part of the genome (Conte et al., 2021). In the cavefish *Astyanax mexicanus* the B chromosome contains a gene which determines male sex and therefore the B chromosome functions as a sex chromosome with a dominant male determining role (Imarazene et al., 2021).

In passerine birds (Passeriformes), an extra chromosome has been observed in the germline. This germline-restricted chromosome (GRC) is excluded by programmed DNA elimination from somatic cells during early embryogenesis (Wang and Davis, 2014). The GRC was first described in the zebra finch (*Taenopygia guttata*) (Pigozzi and Solari, 1998), and it has been revealed recently that it is likely present in all passerine birds (Passeriformes) (Kinsella et al., 2019; Torgasheva et al., 2019). The GRC is present in two copies in oocytes forming a bivalent that undergoes recombination. In contrast, in spermatocytes there is only a single copy of this chromosome which forms a heterochromatic element that is eliminated from the nucleus during the first meiotic division (Schoenmakers et al., 2010; del Priore and Pigozzi, 2014). Camacho et al. (2000) suggested that the GRC may be originally a supernumerary B chromosome which acquired an essential function for birds, possibly a germline determining role, preventing its loss. Genomic analysis of the GRC in zebra finch revealed that similarly as in B chromosomes, most GRC-linked genes are paralogs to genes on A chromosomes, which have been subsequently added to the GRC during passerine evolution (Kinsella et al., 2019). Like B chromosomes, GRC shows high variability in size ranging from the largest chromosome in the karyotype in some species to microchromosome in others (Torgasheva et al., 2019). Dedukh and Krasikova (2021) pointed out yet another similarity which can be found in the programed GRC elimination from somatic cells which strongly resembles mechanisms of tissue-specific B chromosome elimination as described in goatgrass *Aegilops speltoides* where the B chromosome is eliminated from roots (Ruban et al., 2020).

Evolutionary Dynamics of B Chromosomes

Originally, two theoretical models were proposed to explain the occurrence of B chromosomes in populations and their maintenance in relatively stable frequencies. The first model assumed the spread of B chromosomes by some selfish drive mechanisms, opposed by negative effects of the B chromosomes on the carrier's fitness if they are in high copy numbers (Jones 1995; Camacho et al., 2000). The second model (White, 1973) proposed that B chromosomes may be maintained in the population without drive mechanisms if they have a beneficial effect on their carriers in small numbers but start to be detrimental in high copy numbers. Empirical data reviewed in this paper supports both models providing evidence for the selfish spread of B chromosomes in populations through drive in many species (Hasegawa, 1934; Nur, 1962; Nur, 1963; Rutishauser and RÖthlisberger, 1966; Nur, 1969; Kayano, 1971; Jones and Rees, 1982; Gregg at al., 1984; Murray, 1984; Viseras et al., 1990; Jones, 1991; Pardo et al., 1995; Houben, 2017; Jones, 2018; Clark and Kocher, 2019; Blavet et al., 2021) as well as identifying beneficial effects of B chromosomes for their hosts (Williams, 1970; Dherawattana and Sadanaga, 1973; Rees and Hutchinson, 1974; Holmes and Bougourd, 1989; Miao et al., 1991; Plowman and Bougourd, 1994; Enkerli et al., 1997; Hatta et al., 2002; Nokkala et al., 2003; Rodriguez-Carres et al., 2008; Akagi et al., 2009; Coleman et al., 2009; Ma et al., 2010; Balesdent et al., 2013; Thatcher et al., 2016; Williams et al., 2016; Dalíková et al., 2017; Pereira et al., 2017; van Dam et al., 2017; Armitage et al., 2018; Kinsella et al., 2019; Torgasheva et al., 2019; Imarazene et al., 2021; Lewis et al., 2021). From the example in rye where B chromosomes have beneficial function but are still driving (Pereira et al., 2017) we see that there could be even co-occurrence of drive and beneficial function which indicates the rather extensive complexity of B chromosome evolution.

As Camacho et al. (1997) suggested, parasitic B chromosomes initially spreading in a population by drive may be neutralized by the evolution of drive suppressors located on the A chromosomes if they harm the hosts. Since the elimination of already common B chromosomes from the population may be slow, the new drive genes may arise on the B before this elimination leading to a new cycle of B chromosome accumulation. The dynamics between B chromosomes and the rest of the genome may thus resemble a classical arms race between a parasite and its host. Certain studied species seem to show this type of dynamic where A chromosomes try to suppress the accumulation of B chromosomes and B chromosomes try to escape this pressure (Bosemark, 1954; Carlson, 1969; Nur and Brett, 1985, Shaw and Hewitt, 1985; Shaw et al., 1985; Nur and Brett, 1987, Nur and Brett, 1988; Romera et al., 1991; Cebriá et al., 1994; Jiménez et al., 1995; Herrera et al., 1996; Rosato et al., 1996; reviewed in Shaw and Hewitt, 1990; Camacho et al., 2000).

The examples collected in this review also suggest that sometimes B chromosomes can become stable and essential parts of the genome by gaining some vital function such as a role in sex determination or germline specific function. In fact, B chromosomes may be predisposed to become sex chromosomes

or germline-restricted chromosomes by their selfish mechanisms of transmission including sex-specific meiotic drive and gonotaxis. In some cases, B chromosomes can also become accommodated into the genome by being translocated to autosomes or sex chromosomes as observed in several organisms (e.g., grasshoppers and maize) or could acquire regular behavior during meiosis, when two B chromosomes start to pair, both of which ensures their stable inheritance and maintenance in the population (e.g., Houben et al., 2000; Cabrero et al., 2003; Hsu et al., 2003).

CONCLUSION

This review aims to portray B chromosomes as highly dynamic elements, with variable effects on their hosts, and rich evolutionary pathways. We are demonstrating that, although it was originally believed that B chromosomes behave mostly as genomic parasites with neutral or negative effects on the host being spread in the population by selfish drive mechanisms, a growing number of studies have shown that they can also have a positive effect on their hosts. Here we collected such examples including cases where B chromosomes increase the pathogenicity of their hosts or increase the survival rate in particular habitats. Moreover, we provide examples where B chromosomes likely became a stable and essential part of the genome by turning into new sex chromosomes or germline-restricted chromosomes. In this light, B chromosomes can be viewed as a reservoir of genetic material for the evolution of important genomic novelties with potentially significant evolutionary impacts.

AUTHOR CONTRIBUTIONS

MJP: study conception and design, RR: study conception and design. Both authors drafted the manuscript and approved the final version of the article.

FUNDING

This work was supported by the Czech Science Foundation (grant no. 20-23794S) and Charles University (grant PRIMUS/19/SCI/008).

ACKNOWLEDGMENTS

We are very grateful to Manuelita Sotelo, Manon Poignet, Stephen Schlebusch, Dmitrij Dedukh, Christopher Johnson and two reviewers for all their valuable comments and suggestions.

REFERENCES

Ahmad, S. F., Jehangir, M., Cardoso, A. L., Wolf, I. R., Margarido, V. P., Cabral-de-Mello, D. C., et al. (2020). B Chromosomes of Multiple Species Have Intense Evolutionary Dynamics and Accumulated Genes Related to Important Biological Processes. *BMC genomics* 21, 1–25. doi:10.1186/s12864-020-07072-1

Akagi, Y., Akamatsu, H., Otani, H., and Kodama, M. (2009). Horizontal Chromosome Transfer, a Mechanism for the Evolution and Differentiation of a Plant-Pathogenic Fungus. *Eukaryot. Cel.* 8, 1732–1738. doi:10.1128/EC.00135-09

Armitage, A. D., Taylor, A., Sobczyk, M. K., Baxter, L., Greenfield, B. P. J., Bates, H. J., et al. (2018). Characterisation of Pathogen-specific Regions and Novel Effector Candidates in *Fusarium Oxysporum F. Sp.* Cepae. *Sci. Rep.* 8, 13530. doi:10.1038/s41598-018-30335-7

Bachtrog, D. (2013). Y-chromosome Evolution: Emerging Insights into Processes of Y-Chromosome Degeneration. *Nat. Rev. Genet.* 14, 113–124. doi:10.1038/nrg3366

Balesdent, M. H., Fudal, I., Ollivier, B., Bally, P., Grandaubert, J., Eber, F., et al. (2013). The Dispensable Chromosome of *Leptosphaeria maculans* Shelters an Effector Gene Conferring Avirulence towards *Brassica rapa*. *New Phytol.* 198, 887–898. doi:10.1111/nph.12178

Banaei-Moghaddam, A. M., Meier, K., Karimi-Ashtiyani, R., and Houben, A. (2013). Formation and Expression of Pseudogenes on the B Chromosome of rye. *The Plant Cell* 25, 2536–2544. doi:10.1105/tpc.113.111856

Banaei-Moghaddam, A. M., Schubert, V., Kumke, K., Weiβ, O., Klemme, S., Nagaki, K., et al. (2012). Nondisjunction in Favor of a Chromosome: the Mechanism of rye B Chromosome Drive during Pollen Mitosis. *Plant Cell* 24, 4124–4134. doi:10.1105/tpc.112.105270

Bauerly, E., Hughes, S. E., Vietti, D. R., Miller, D. E., McDowell, W., and Hawley, R. S. (2014). Discovery of Supernumerary B Chromosomes in *Drosophila melanogaster*. *Genetics* 196, 1007–1016. doi:10.1534/genetics.113.160556

Baverstock, P. R., Gelder, M., and Jahnke, A. (1982). Cytogenetic Studies of the Australian Rodent, *Uromys caudimaculatus*, a Species Showing Extensive Heterochromatin Variation. *Chromosoma* 84, 517–533. doi:10.1007/BF00292853

Beukeboom, L. W. (1994). Bewildering Bs: an Impression of the 1st B-Chromosome Conference. *Heredity* 73, 328–336. doi:10.1038/hdy.1994.140

Beukeboom, L. W., Seif, M., Mettenmeyer, T., Plowman, A. B., and Michiels, N. K. (1996). Paternal Inheritance of B Chromosomes in a Parthenogenetic Hermaphrodite. *Heredity* 77, 646–654. doi:10.1038/hdy.1996.192

Bidau, C. J., and Martí, D. A. (2004). B Chromosomes and Robertsonian Fusions of *Dichroplus pratensis* (Acrididae): Intraspecific Support for the Centromeric Drive Theory. *Cytogenet. Genome Res.* 106, 347–350. doi:10.1159/000079311

Blavet, N., Yang, H., Su, H., Solanský, P., Douglas, R. N., Karafiátová, M., et al. (2021). Sequence of the Supernumerary B Chromosome of maize Provides Insight into its Drive Mechanism and Evolution. *Proc. Natl. Acad. Sci. USA* 118, e2104254118. doi:10.1073/pnas.2104254118

Bosemark, N. O. (1954). On Accessory Chromosomes in *Festuca pratensis*. II. Inheritance of the Standard Type of Accessory Chromosome. *Hereditas* 40, 425–437.

Bougourd, S. M., and Parker, J. S. (1979). The B-Chromosome System of *Allium schoenoprasum*. *Chromosoma* 75, 369–383. doi:10.1007/bf00293478

Boyes, J. W., and van Brink, J. M. (1967). Chromosomes of Syrphidae. *Chromosoma* 22, 417–455. doi:10.1007/BF00286546

Cabrero, J., Perfectti, F., Gómez, R., Camacho, J. P., and López-León, M. D. (2003). Population Variation in the A Chromosome Distribution of Satellite DNA and Ribosomal DNA in the Grasshopper *Eyprepocnemis plorans*. *Chromosome Res.* 11, 375–381. doi:10.1023/a:1024127525756

Camacho, J. P. M. (2005). "B Chromosomes," in *The Evolution of the Genome* (Academic Press), 223–286. doi:10.1016/b978-012301463-4/50006-1

Camacho, J. P. M., Perfectti, F., Teruel, M., López-León, M. D., and Cabrero, J. (2004). The Odd-Even Effect in Mitotically Unstable B Chromosomes in Grasshoppers. *Cytogenet. Genome Res.* 106, 325–331. doi:10.1159/000079307

Camacho, J. P. M., Sharbel, T. F., and Beukeboom, L. W. (2000). B-chromosome Evolution. *Phil. Trans. R. Soc. Lond. B* 355, 163–178. doi:10.1098/rstb.2000.0556

Carlson, W. R. (1969). Factors Affecting Preferential Fertilization in maize. *Genetics* 62, 543–554. doi:10.1093/genetics/62.3.543

Carvalho, A. (2002). Origin and Evolution of the *Drosophila* Y Chromosome. *Curr. Opin. Genet. Dev.* 12, 664–668. doi:10.1016/s0959-437x(02)00356-8

Cebriá, A., Navarro, M. L., and Puertas, M. J. (1994). Genetic Control of B Chromosome Transmission in *Aegilops speltoides* (Poaceae). *Am. J. Bot.* 81, 1502–1507. doi:10.1002/j.1537-2197.1994.tb15636.x

Clark, F. E., and Kocher, T. D. (2019). Changing Sex for Selfish Gain: B Chromosomes of Lake Malawi Cichlid Fish. *Sci. Rep.* 9, 1–10. doi:10.1038/s41598-019-55774-8

Coleman, J. J., Rounsley, S. D., Rodriguez-Carres, M., Kuo, A., Wasmann, C. C., Grimwood, J., et al. (2009). The Genome of *Nectria haematococca*: Contribution of Supernumerary Chromosomes to Gene Expansion. *Plos Genet.* 5, e1000618. doi:10.1371/journal.pgen.1000618

Conte, M. A., Clark, F. E., Roberts, R. B., Xu, L., Tao, W., Zhou, Q., et al. (2021). Origin of a Giant Sex Chromosome. *Mol. Biol. Evol.* 38, 1554–1569. doi:10.1093/molbev/msaa319

Croll, D., and McDonald, B. A. (2012). The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. *Plos Pathog.* 8, e1002608. doi:10.1371/journal.ppat.1002608

D'Ambrosio, U., Alonso-Lifante, M. P., Barros, K., Kovařík, A., Mas de Xaxars, G., and Garcia, S. (2017). B-chrom: a Database on B-chromosomes of Plants, Animals and Fungi. *New Phytol.* 216, 635–642. doi:10.1111/nph.14723

Dalíková, M., Zrzavá, M., Hladová, I., Nguyen, P., Šonský, I., Flegrová, M., et al. (2017). New Insights into the Evolution of the W Chromosome in Lepidoptera. *J. Hered.* 108, 709–719. doi:10.1093/jhered/esx063

Dalla Benetta, E., Antoshechkin, I., Yang, T., Nguyen, H. Q. M., Ferree, P. M., and Akbari, O. S. (2020). Genome Elimination Mediated by Gene Expression from a Selfish Chromosome. *Sci. Adv.* 6, eaaz9808. doi:10.1126/sciadv.aaz9808

Dedukh, D., and Krasikova, A. (2021). Delete and Survive: Strategies of Programmed Genetic Material Elimination in Eukaryotes. *Biol. Rev.* doi:10.1111/brv.12796

del Priore, L., and Pigozzi, M. I. (2014). Histone Modifications Related to Chromosome Silencing and Elimination during Male Meiosis in Bengalese Finch. *Chromosoma* 123, 293–302. doi:10.1007/s00412-014-0451-3

Dhar, M. K., Friebe, B., Koul, A. K., and Gill, B. S. (2002). Origin of an Apparent B Chromosome by Mutation, Chromosome Fragmentation and Specific DNA Sequence Amplification. *Chromosoma* 111, 332–340. doi:10.1007/s00412-002-0214-4

Dherawattana, A., and Sadanaga, K. (1973). Cytogenetics of a Crown Rust-Resistant Hexaploid Oat with 42 + 2 Fragment Chromosomes 1. *Crop Sci.* 13, 591–594. doi:10.2135/cropsci1973.0011183X001300060002x

Enkerli, J., Bhatt, G., and Covert, S. F. (1997). Nht1, a Transposable Element Cloned from a Dispensable Chromosome in *Nectria haematococca*. *Mpmi* 10, 742–749. doi:10.1094/MPMI.1997.10.6.742

Gotoh, K. (1924). Uuml;ber die Chromosomenzahl von *Secale cereale*, L. *Shokubutsugaku Zasshi* 38, 135–152. doi:10.15281/jplantres1887.38.135

Graphodatsky, A. S., Kukekova, A. V., Yudkin, D. V., Trifonov, V. A., Vorobieva, N. V., Beklemisheva, V. R., et al. (2005). The Proto-Oncogene C-KIT Maps to Canid B-Chromosomes. *Chromosome Res.* 13, 113–122. doi:10.1007/s10577-005-7474-9

Gregg, P. C., Webb, G. C., and Adena, M. A. (1984). The Dynamics of B Chromosomes in Populations of the Australian Plague Locust, *Chortoicetes terminifera* (Walker). *Can. J. Genet. Cytol.* 26, 194–208. doi:10.1139/g84-033

Hackstein, J. H. P., Hochstenbach, R., Hauschteck-Jungen, E., and Beukeboom, L. W. (1996). Is the Y Chromosome of *Drosophila* an Evolved Supernumerary Chromosome? *Bioessays* 18, 317–323. doi:10.1002/bies.950180410

Hasegawa, N. (1934). A Cytological Study on 8-chromosome rye. *Cytologia* 6, 68–77. doi:10.1508/cytologia.6.68

Hatta, R., Ito, K., Hosaki, Y., Tanaka, T., Tanaka, A., Yamamoto, M., et al. (2002). A Conditionally Dispensable Chromosome Controls Host-specific Pathogenicity in the Fungal Plant Pathogen *Alternaria alternata*. *Genetics* 161, 59–70. doi:10.1093/genetics/161.1.59

Hejníčková, M., Koutecký, P., Potocký, P., Provazníková, I., Voleníková, A., Dalíková, M., et al. (2019). Absence of W Chromosome in Psychidae Moths and Implications for the Theory of Sex Chromosome Evolution in Lepidoptera. *Genes* 10, 1016. doi:10.3390/genes10121016

Herrera, J. A., López-León, M. D., Cabrero, J., Shaw, M. W., and Camacho, J. P. M. (1996). Evidence for B Chromosome Drive Suppression in the Grasshopper *Eyprepocnemis plorans*. *Heredity* 76, 633–639. doi:10.1038/hdy.1996.90

Holmes, D. S., and Bougourd, S. M. (1989). B-chromosome Selection in *Allium schoenoprasum*. I. Natural Populations. *Heredity* 63, 83–87. doi:10.1038/hdy.1989.78

Houben, A. (2017). B Chromosomes - A Matter of Chromosome Drive. *Front. Plant Sci.* 08, 210. doi:10.3389/fpls.2017.00210

Houben, A., Banaei-Moghaddam, A. M., Klemme, S., and Timmis, J. N. (2014). Evolution and Biology of Supernumerary B Chromosomes. *Cell. Mol. Life Sci.* 71, 467–478. doi:10.1007/s00018-013-1437-7

Houben, A., Wanner, G., Hanson, L., Verlin, D., Leach, C. R., and Timmis, J. N. (2000). Cloning and Characterisation of Polymorphic Heterochromatic Segments of *Brachycome dichromosomatica*. *Chromosoma* 109, 206–213. doi:10.1007/s004120050430

Hsu, F. C., Wang, C. J., Chen, C. M., Hu, H. Y., and Chen, C. C. (2003). Molecular Characterization of a Family of Tandemly Repeated DNA Sequences, TR-1, in Heterochromatic Knobs of maize and its Relatives. *Genetics* 164, 1087–1097. doi:10.1093/genetics/164.3.1087

Imarazene, B., Du, K., Beille, S., Jouanno, E., Feron, R., Pan, Q., et al. (2021). A Supernumerary "B-sex" Chromosome Drives Male Sex Determination in the Pachón Cavefish, *Astyanax mexicanus*. *Curr. Biol.* doi:10.1016/j.cub.2021.08.030

Jiménez, M. M., Romera, F., Gallego, A., and Puertas, M. J. (1995). Genetic Control of the Rate of Transmission of rye B Chromosomes. II. 0B × 2B Crosses. *Heredity* 74, 518–523. doi:10.1038/hdy.1995.73

Jones, N. (2017). New Species with B Chromosomes Discovered since 1980. *Nucleus* 60, 263–281. doi:10.1007/s13237-017-0215-6

Jones, R. N. (1995). B Chromosomes in Plants. *New Phytol.* 131, 411–434. doi:10.1111/j.1469-8137.1995.tb03079.x

Jones, R. N. (1991). B-chromosome Drive. *The Am. Naturalist* 137, 430–442. doi:10.1086/285175

Jones, R. N., and Rees, H. (1982). *B Chromosomes*. Academic Press.

Jones, R. N. (2018). Transmission and Drive Involving Parasitic B Chromosomes. *Genes* 9, 388. doi:10.3390/genes9080388

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.* 104, 14026–14031. doi:10.1073/pnas.0706647104

Kayano, H. (1971). Accumulation of B Chromosomes in the Germ Line of *Locusta migratoria*. *Heredity* 27, 119–123. doi:10.1038/hdy.1971.76

Kichigin, I. G., Lisachov, A. P., Giovannotti, M., Makunin, A. I., Kabilov, M. R., O'Brien, P. C. M., et al. (2019). First Report on B Chromosome Content in a Reptilian Species: The Case of *Anolis carolinensis*. *Mol. Genet. Genomics.* 294, 13–21. doi:10.1007/s00438-018-1483-9

Kinsella, C. M., Ruiz-Ruano, F. J., Dion-Côté, A.-M., Charles, A. J., Gossmann, T. I., Cabrero, J., et al. (2019). Programmed DNA Elimination of Germline Development Genes in Songbirds. *Nat. Commun.* 10, 5468. doi:10.1038/s41467-019-13427-4

Kuwada, Y. (1925). On the Number of Chromosomes in maize. *Shokubutsugaku Zasshi* 39, 227–234. doi:10.15281/jplantres1887.39.465_227

Lewis, J. J., Cicconardi, F., Martin, S. H., Reed, R. D., Danko, C. G., and Montgomery, S. H. (2021). The *Dryas iulia* Genome Supports Multiple Gains of a W Chromosome from a B Chromosome in Butterflies. *Genome Biol. Evol.* 13, evab128. doi:10.1093/gbe/evab128

Longley, A. E. (1927). Supernumerary Chromosomes in *Zea mays*. *J. Agric. Res.* 35, 769–784.

López-León, M. D., Neves, N., Schwarzacher, T., Heslop-Harrison, J. S., Hewitt, G. M., and Camacho, J. P. (1994). Possible Origin of a B Chromosome Deduced from its DNA Composition Using Double FISH Technique. *Chromosome Res.* 2, 87–92. doi:10.1007/BF01553487

López-León, M. D., Cabrero, J., Pardo, M. C., Viseras, E., Camacho, J. P. M., and Santos, J. L. (1993). Generating High Variability of B Chromosomes in *Eyprepocnemis plorans* (Grasshopper). *Heredity* 71, 352–362. doi:10.1038/hdy.1993.149

Lukhtanov, V. A. (2000). Sex Chromatin and Sex Chromosome Systems in Nonditrysian Lepidoptera (Insecta). *J. Zoolog. Syst.* 38, 73–79. doi:10.1046/j.1439-0469.2000.382130.x

Ma, L.-J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., et al. (2010). Comparative Genomics Reveals mobile Pathogenicity Chromosomes in *Fusarium*. *Nature* 464, 367–373. doi:10.1038/nature08850

Makunin, A. I., Kichigin, I. G., Larkin, D. M., O'Brien, P. C., Ferguson-Smith, M. A., Yang, F., et al. (2016). Contrasting Origin of B Chromosomes in Two Cervids (Siberian Roe Deer and Grey Brocket Deer) Unravelled by Chromosome-specific DNA Sequencing. *BMC Genomics* 17, 618–714. doi:10.1186/s12864-016-2933-6

Makunin, A., Romanenko, S., Beklemisheva, V., Perelman, P., Druzhkova, A., Petrova, K., et al. (2018). Sequencing of Supernumerary Chromosomes of Red Fox and raccoon Dog Confirms a Non-random Gene Acquisition by B Chromosomes. *Genes* 9, 405. doi:10.3390/genes9080405

Mandáková, T., Ashby, K., Price, B. J., Windham, M. D., Carman, J. G., and Lysak, M. A. (2021). Genome Structure and Apomixis in *Phoenicaulis* (Brassicaceae; Boechereae). *J. Syst. Evol.* 59, 83–92. doi:10.1111/jse.12555

Mandáková, T., Schranz, M. E., Sharbel, T. F., de Jong, H., and Lysak, M. A. (2015). Karyotype Evolution in apomictic *Boechera* and the Origin of the Aberrant Chromosomes. *Plant J.* 82, 785–793. doi:10.1111/tpj.12849

Marques, A., Klemme, S., and Houben, A. (2018). Evolution of Plant B Chromosome Enriched Sequences. *Genes* 9, 515. doi:10.3390/genes9100515

Martins, C., and Jehangir, M. (2021). A Genomic Glimpse of B Chromosomes in Cichlids. *Genes Genom.* 43, 199–208. doi:10.1007/s13258-021-01049-4

Martis, M. M., Klemme, S., Banaei-Moghaddam, A. M., Blattner, F. R., Macas, J., Schmutzer, T., et al. (2012). Selfish Supernumerary Chromosome Reveals its Origin as a Mosaic of Host Genome and Organellar Sequences. *Proc. Natl. Acad. Sci.* 109, 13343–13346. doi:10.1073/pnas.1204237109

McAllister, B. F. (1995). Isolation and Characterization of a Retroelement from B Chromosome (PSR) in the Parasitic Wasp *Nasonia vitripennis*. *Insect Mol. Biol.* 4, 253–262. doi:10.1111/j.1365-2583.1995.tb00031.x

McAllister, B. F., and Werren, J. H. (1997). Hybrid Origin of a B Chromosome (PSR) in the Parasitic Wasp *Nasonia vitripennis*. *Chromosoma* 106, 243–253. doi:10.1007/s004120050245

McVean, G. T. (1995). Fractious Chromosomes: Hybrid Disruption and the Origin of Selfish Genetic Elements. *Bioessays* 17, 579–582. doi:10.1002/bies.950170702

Mestriner, C. A., Galetti, P. M., Valentini, S. R., Ruiz, I. R. G., Abel, L. D. S., Moreira-Filho, O., et al. (2000). Structural and Functional Evidence that a B Chromosome in the Characid Fish *Astyanax scabripinnis* is an Isochromosome. *Heredity* 85, 1–9. doi:10.1046/j.1365-2540.2000.00702.x

Miao, V. P., Covert, S. F., and VanEtten, H. D. (1991). A Fungal Gene for Antibiotic Resistance on a Dispensable ("B") Chromosome. *Science* 254, 1773–1776. doi:10.1126/science.1763326

Murray, B. G. (1984). The Structure, Meiotic Behaviour and Effects of B Chromosomes in *Briza humilis* Bieb. (Gramineae). *Genetica* 63 (3), 213–219.

Navarro-Domínguez, B., Ruiz-Ruano, F. J., Cabrero, J., Corral, J. M., López-León, M. D., Sharbel, T. F., et al. (2017). Protein-coding Genes in B Chromosomes of the Grasshopper *Eyprepocnemis plorans*. *Sci. Rep.* 7, 1–12. doi:10.1038/srep45200

Nokkala, S., Grozeva, S., Kuznetsova, V., and Maryanska-Nadachowska, A. (2003). The Origin of the Achiasmatic XY Sex Chromosome System in *Cacopsylla peregrina* (Frst.) (Psylloidea, Homoptera). *Genetica* 119, 327–332. doi:10.1023/b:gene.0000003757.27521.4d

Nur, U. (1963). A Mitotically Unstable Supernumerary Chromosome with an Accumulation Mechanism in a Grasshopper. *Chromosoma* 14, 407–422. doi:10.1007/bf00326786

Nur, U. (1962). A Supernumerary Chromosome with an Accumulation Mechanism in the Lecanoid Genetic System. *Chromosoma* 13, 249–271. doi:10.1007/bf00577042

Nur, U., and Brett, B. L. H. (1987). Control of Meiotic Drive of B Chromosomes in the Mealybug, *Pseudococcus affinis (obscurus)*. *Genetics* 115, 499–510. doi:10.1093/genetics/115.3.499

Nur, U., and Brett, B. L. H. (1988). Genotypes Affecting the Condensation and Transmission of Heterochromtic B Chromosomes in the Mealybug *Pseudococcus affinis*. *Chromosoma* 96, 205–212. doi:10.1007/bf00302359

Nur, U., and Brett, B. L. H. (1985). Genotypes Suppressing Meiotic Drive of a B Chromosome in the Mealybug, *Pseudococcus obscurus*. *Genetics* 110, 73–92. doi:10.1093/genetics/110.1.73

Nur, U. (1969). Mitotic Instability Leading to an Accumulation of B-Chromosomes in Grasshoppers. *Chromosoma* 27, 1–19. doi:10.1007/BF00326108

Östergren, G. (1945). Parasitic Nature of Extra Fragment Chromosomes. *Bot. Not.* 2, 157–163.

Padro-Manuel De Villena, F. P.-M., and Sapienza, C. (2001). Female Meiosis Drives Karyotypic Evolution in Mammals. *Genetics* 159, 1179–1189. doi:10.1093/genetics/159.3.1179

Palestis, B. G., Burt, A., Jones, R. N., and Trivers, R. (2004). B Chromosomes Are More Frequent in Mammals with Acrocentric Karyotypes: Support for the Theory of Centromeric Drive. *Proc. Biol. Sci.* 271 Suppl 3, S22–S24. doi:10.1098/rsbl.2003.0084

Palestis, B. G., Cabrero, J., Trivers, R., and Camacho, J. P. M. (2010). Prevalence of B Chromosomes in Orthoptera is Associated with Shape and Number of A Chromosomes. *Genetica* 138, 1181–1189. doi:10.1007/s10709-010-9509-1

Pardo, M. C., López-León, M. D., Viseras, E., Cabrero, J., and Camacho, J. P. M. (1995). Mitotic Instability of B Chromosomes during Embryo Development in *Locusta migratoria*. *Heredity* 74, 164–169. doi:10.1038/hdy.1995.24

Peppers, J. A., Wiggins, L. E., and Baker, R. J. (1997). Nature of B Chromosomes in the Harvest Mouse *Reithrodontomys megalotis* by Fluorescence *in situ* Hybridization (FISH). *Chromosome Res.* 5, 475–479. doi:10.1023/a:1018421114607

Pereira, H. S., Delgado, M., Viegas, W., Rato, J. M., Barão, A., and Caperta, A. D. (2017). Rye (*Secale cereale*) Supernumerary (B) Chromosomes Associated with Heat Tolerance during Early Stages of Male Sporogenesis. *Ann. Bot.* 119, 325–337. doi:10.1093/aob/mcw206

Perfectti, F., and Werren, J. H. (2001). The Interspecific Origin of B Chromosomes: Experimental Evidence. *Evol.* 55, 1069–1073. doi:10.1111/j.0014-3820.2001.tb00624.x10.1554/0014-3820(2001)055[1069:tioobc]2.0.co;2

Peters, G. B. (1981). Germ Line Polysomy in the Grasshopper *Atractomorpha similis*. *Chromosoma* 81, 593–617. doi:10.1007/bf00285852

Pigozzi, M. I., and Solari, A. J. (1998). Germ Cell Restriction and Regular Transmission of an accessory Chromosome that Mimics a Sex Body in the Zebra Finch, *Taeniopygia guttata*. *Chromosome Res.* 6, 105–113. doi:10.1023/a:1009234912307

Plowman, A. B., and Bougourd, S. M. (1994). Selectively Advantageous Effects of B Chromosomes on Germination Behaviour in *Allium schoenoprasum* L. *Heredity* 72, 587–593. doi:10.1038/hdy.1994.81

Randolph, L. F. (1928). Types of Supernumerary Chromosomes in maize. *Anat. Rec.* 41, 102.

Reed, K. M., and Werren, J. H. (1995). Induction of Paternal Genome Loss by the Paternal-Sex-Ratio Chromosome and Cytoplasmic Incompatibility Bacteria (*Wolbachia*): A Comparative Study of Early Embryonic Events. *Mol. Reprod. Dev.* 40, 408–418. doi:10.1002/mrd.1080400404

Rees, H., and Hutchinson, J. (1974). *Nuclear DNA Variation Due to B chromosomes." in Cold Spring Harbor Symposia on Quantitative Biology*, 38. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. doi:10.1101/sqb.1974.038.01.021

Rodriguez-Carres, M., White, G., Tsuchiya, D., Taga, M., and VanEtten, H. D. (2008). The Supernumerary Chromosome of *Nectria haematococca* that Carries Pea-Pathogenicity-Related Genes Also Carries a Trait for Pea Rhizosphere Competitiveness. *Appl. Environ. Microbiol.* 74, 3849–3856. doi:10.1128/aem.00351-08

Romera, F., Jiménez, M. M., and Puertas, M. J. (1991). Genetic Control of the Rate of Transmission of rye B Chromosomes. I. Effects in 2B × 0B Crosses. *Heredity* 66, 61–65. doi:10.1038/hdy.1991.8

Rosato, M., Chiavarino, A. M., Naranjo, C. A., Puertas, M. J., and Poggio, L. (1996). Genetic Control of B Chromosome Transmission Rate in *Zea mays* ssp. *mays* (Poaceae). *Am. J. Bot.* 83, 1107–1112. doi:10.1002/j.1537-2197.1996.tb13890.x

Ruban, A., Fuchs, J., Marques, A., Schubert, V., Soloviev, A., Raskina, O., et al. (2014). B Chromosomes of *Aegilops speltoides* Are Enriched in Organelle Genome-Derived Sequences. *PLoS ONE* 9, e90214. doi:10.1371/journal.pone.0090214

Ruban, A., Schmutzer, T., Scholz, U., and Houben, A. (2017). How Next-Generation Sequencing Has Aided Our Understanding of the Sequence Composition and Origin of B Chromosomes. *Genes* 8, 294. doi:10.3390/genes8110294

Ruban, A., Schmutzer, T., Wu, D. D., Fuchs, J., Boudichevskaia, A., Rubtsova, M., et al. (2020). Supernumerary B Chromosomes of *Aegilops speltoides* Undergo Precise Elimination in Roots Early in Embryo Development. *Nat. Commun.* 11, 1–12. doi:10.1038/s41467-020-16594-x

Ruiz-Ruano, F. J., Cabrero, J., López-León, M. D., Sánchez, A., and Camacho, J. P. M. (2018). Quantitative Sequence Characterization for Repetitive DNA Content in the Supernumerary Chromosome of the Migratory Locust. *Chromosoma* 127, 45–57. doi:10.1007/s00412-017-0644-7

Rutishauser, A., and Rothlisberger, E. (1966). Boosting Mechanism of B-Chromosomes in *Crepis capilaris*. *Chrom. Today* 1, 28–30.

Schoenmakers, S., Wassenaar, E., Laven, J. S. E., Grootegoed, J. A., and Baarends, W. M. (2010). Meiotic Silencing and Fragmentation of the Male Germline Restricted Chromosome in Zebra Finch. *Chromosoma* 119, 311–324. doi:10.1007/s00412-010-0258-9

Sharbel, T. F., Beukeboom, L. W., and Pijnacker, L. P. (1997). Multiple Supernumerary Chromosomes in the Pseudogamous Parthenogenetic flatworm *Polycelis nigra*: Lineage Markers or Remnants of Genetic Leakage? *Genome* 40, 850–856. doi:10.1139/g97-810

Sharbel, T. F., Voigt, M.-L., Mitchell-Olds, T., Kantama, L., and De Jong, H. (2004). Is the Aneuploid Chromosome in an Apomictic *Boechera holboellii* a Genuine B Chromosome? *Cytogenet. Genome Res.* 106, 173–183. doi:10.1159/000079284

Shaw, M. W., Hewitt, G. M., and Anderson, D. A. (1985). Polymorphism in the Rates of Meiotic Drive Acting on the B-Chromosome of *Myrmeleotettix maculatus*. *Heredity* 55, 61–68. doi:10.1038/hdy.1985.72

Shaw, M. W., and Hewitt, G. M. (1990). B Chromosomes, Selfish DNA and Theoretical Models: where Next? *Oxf. Surv. Evol. Biol.* 7, 197–223.

Shaw, M. W., and Hewitt, G. M. (1985). The Genetic Control of Meiotic Drive Acting on the B Chromosome of *Myrmeleotettix maculatus* (Orthoptera: Acrididae). *Heredity* 54, 259–268. doi:10.1038/hdy.1985.25

Stevens, N. M. (1908). The Chromosomes in *Diabrotica vittata*, *Diabrotica soror*, *Diabrotica 12-punctata*. A Contribution to the Literature on Heterochromosomes and Sex Determination. *J. Exp. Zool.* 5, 453–470. doi:10.1002/jez.1400050402

Talavera, M., López-Leon, M. D., Cabrero, J., and Camacho, J. P. M. (1990). Male Germ Line Polysomy in the Grasshopper *Chorthippus binotatus*: Extra Chromosomes Are Not Transmitted. *Genome* 33, 384–388. doi:10.1139/g90-058

Thatcher, L. F., Williams, A. H., Garg, G., Buck, S.-A. G., and Singh, K. B. (2016). Transcriptome Analysis of the Fungal Pathogen *Fusarium oxysporum* f. sp. *medicaginis* during colonisation of resistant and susceptible *Medicago truncatula* hosts Identifies Differential Pathogenicity Profiles and Novel Candidate Effectors. *BMC Genomics* 17, 860. doi:10.1186/s12864-016-3192-2

Torgasheva, A. A., Malinovskaya, L. P., Zadesenets, K. S., Karamysheva, T. V., Kizilova, E. A., Akberdina, E. A., et al. (2019). Germline-restricted Chromosome (GRC) is Widespread Among Songbirds. *Proc. Natl. Acad. Sci. U S A.* 116, 11845–11850. doi:10.1073/pnas.1817373116

Uhl, C. H., and Moran, R. (1973). The Chromosomes of *Pachyphytum* (Crassulaceae). *Am. J. Bot.* 60, 648–656. doi:10.1002/j.1537-2197.1973.tb05969.x

Valente, G. T., Conte, M. A., Fantinatti, B. E. A., Cabral-de-Mello, D. C., Carvalho, R. F., Vicari, M. R., et al. (2014). Origin and Evolution of B Chromosomes in the Cichlid Fish *Astatotilapia latifasciata* Based on Integrated Genomic Analyses. *Mol. Biol. Evol.* 31, 2061–2072. doi:10.1093/molbev/msu148

van Dam, P., Fokkens, L., Ayukawa, Y., van der Gragt, M., ter Horst, A., Brankovics, B., et al. (2017). A mobile Pathogenicity Chromosome in *Fusarium oxysporum* for Infection of Multiple Cucurbit Species. *Sci. Rep.* 7, 9042. doi:10.1038/s41598-017-07995-y

Van Valen, L. (1977). The Red Queen. *Am. Naturalist* 111, 809–810. doi:10.1086/283213

Vanheule, A., Audenaert, K., Warris, S., van de Geest, H., Schijlen, E., Höfte, M., et al. (2016). Living Apart Together: Crosstalk between the Core and Supernumerary Genomes in a Fungal Plant Pathogen. *BMC Genomics* 17, 670–718. doi:10.1186/s12864-016-2941-6

Ventura, K., O'Brien, P. C. M., do Nascimento Moreira, C., Yonenaga-Yassuda, Y., and Ferguson-Smith, M. A. (2015). On the Origin and Evolution of the Extant System of B Chromosomes in Oryzomyini Radiation (Rodentia, Sigmodontinae). *PLoS ONE* 10, e0136663. doi:10.1371/journal.pone.0136663

Viseras, E., Camacho, J. P. M., Cano, M. I., and Santos, J. L. (1990). Relationship between Mitotic Instability and Accumulation of B Chromosomes in Males and Females of *Locusta migratoria*. *Genome* 33 23–29. doi:10.1139/g90-005

Volobuev, V. T., and Timina, N. Iu. (1980). Unusually High Number of B-Chromosomes and Mosaicism Among Them in the Asiatic forest Mouse *Apodemus peninsulae* (Rodentia, Muridae). *Tsitol Genet.* 14, 43–45.

Wang, J., and Davis, R. E. (2014). Programmed DNA Elimination in Multicellular Organisms. *Curr. Opin. Genet. Develop.* 27, 26–34. doi:10.1016/j.gde.2014.03.012

Werren, J. H., and Beukeboom, L. W. (1993). Population Genetics of a Parasitic Chromosome: Theoretical Analysis of PSR in Subdivided Populations. *Am. Naturalist* 142, 224–241. doi:10.1086/285536

Werren, J. H., Nur, U., and Eickbush, D. (1987). An Extrachromosomal Factor Causing Loss of Paternal Chromosomes. *Nature* 327, 75–76. doi:10.1038/327075a0

Werren, J. H., and Stouthamer, R. (2003). PSR (Paternal Sex Ratio) Chromosomes: The Ultimate Selfish Genetic Elements. *Genetica* 117, 85–101. doi:10.1023/a:1022368700752

Werren, J. H. (1991). The Paternal-Sex-Ratio Chromosome of *Nasonia*. *Am. Naturalist* 137, 392–402. doi:10.1086/285172

Williams, A. H., Sharma, M., Thatcher, L. F., Azam, S., Hane, J. K., Sperschneider, J., et al. (2016). Comparative Genomics and Prediction of Conditionally Dispensable Sequences in Legume-Infecting *Fusarium oxysporum formae speciales* Facilitates Identification of Candidate Effectors. *BMC Genomics* 17, 191. doi:10.1186/s12864-016-2486-8

Williams, P. (1970). *Genetical Effects of B-Chromosomes in* Lolium. [dissertation]. Cardiff: University of Wales.

Wilson, E. B. (1907). The Supernumerary Chromosomes of Hemiptera. *Sci. New Ser.* 26, 870–871. doi:10.1126/science.26.660.258-a

Wolf, K. W., Mertl, H. G. n., and Traut, W. (1991). Structure, Mitotic and Meiotic Behaviour, and Stability of Centromere-like Elements Devoid of Chromosome Arms in the Fly *Megaselia scalaris* (Phoridae). *Chromosoma* 101, 99–108. doi:10.1007/BF00357059

Yang, H., Yu, H., and Ma, L.-J. (2020). Accessory Chromosomes in *Fusarium oxysporum*. *Phytopathology* 110, 1488–1496. doi:10.1094/phyto-03-20-0069-ia

Yoshida, K., Terai, Y., Mizoiri, S., Aibara, M., Nishihara, H., Watanabe, M., et al. (2011). B Chromosomes Have a Functional Effect on Female Sex Determination in Lake Victoria Cichlid Fishes. *Plos Genet.* 7, e1002203. doi:10.1371/journal.pgen.1002203

Ziegler, C. G., Lamatsch, D. K., Steinlein, C., Engel, W., Schartl, M., and Schmid, M. (2003). The Giant B Chromosome of the Cyprinid Fish *Alburnus alburnus* Harbours a Retrotransposon-Derived Repetitive DNA Sequence. *Chromosome Res.* 11, 23–35. doi:10.1023/a:1022053931308

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 14 March 2022
doi: 10.3389/fgene.2022.793734

Repeat Age Decomposition Informs an Ancient Set of Repeats Associated With Coleoid Cephalopod Divergence

Alba Marino[1,2,3]*, *Alena Kizenko*[1], *Wai Yee Wong*[1], *Fabrizio Ghiselli*[2] and *Oleg Simakov*[1]

[1]Department for Neurosciences and Developmental Biology, University of Vienna, Vienna, Austria, [2]Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy, [3]Institute of Evolutionary Science of Montpellier, University of Montpellier, Montpellier, France

In comparison with other molluscs and bilaterians, the genomes of coleoid cephalopods (squid, cuttlefish, and octopus) sequenced so far show remarkably different genomic organization that presumably marked the early evolution of this taxon. The main driver behind this genomic rearrangement remains unclear. About half of the genome content in coleoids is known to consist of repeat elements; since selfish DNA is one of the powerful drivers of genome evolution, its pervasiveness could be intertwined with the emergence of cephalopod-specific genomic signatures and could have played an important role in the reorganization of the cephalopod genome architecture. However, due to abundant species-specific repeat expansions, it has not been possible so far to identify the ancient shared set of repeats associated with coleoid divergence. By means of an extensive repeat element re-evaluation and annotation combined with network sequence divergence approaches, we are able to identify and characterize the ancient repeat complement shared by at least four coleoid cephalopod species. Surprisingly, instead of the most abundant elements present in extant genomes, lower-copy-number DNA and retroelements were most associated with ancient coleoid radiation. Furthermore, evolutionary analysis of some of the most abundant families shared in *Octopus bimaculoides* and *Euprymna scolopes* disclosed within-family patterns of large species-specific expansions while also identifying a smaller shared expansion in the coleoid ancestor. Our study thus reveals the apomorphic nature of retroelement expansion in octopus and a conserved complement composed of several DNA element types and fewer LINE families.

Keywords: cephalopods, genome architecture, evolution, repeat elements, LINEs, SINEs, ancient repeat complement

OPEN ACCESS

Edited by:
Ricardo Utsunomia, Federal Rural University of Rio de Janeiro, Brazil

Reviewed by:
Frederic Guy Brunet, UMR5242 Institut de Génomique Fonctionnelle de Lyon (IGFL), France
René Massimiliano Marsano, University of Bari Aldo Moro, Italy

***Correspondence:**
Alba Marino
alba.marino@etu.umontpellier.fr

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *12 October 2021*
Accepted: *14 February 2022*
Published: *14 March 2022*

Citation:
Marino A, Kizenko A, Wong WY, Ghiselli F and Simakov O (2022) Repeat Age Decomposition Informs an Ancient Set of Repeats Associated With Coleoid Cephalopod Divergence. Front. Genet. 13:793734. doi: 10.3389/fgene.2022.793734

Abbreviations: BLAST, Basic Local Alignment Search Tool; GO, Gene Ontology; HGT, Horizontal Gene Transfer; LINE, Long Interspersed Nuclear Element; LTR, Long Terminal Repeat; PCA, Principal Component Analysis; SINE, Short Interspersed Nuclear Element; TE, Transposable Element.

INTRODUCTION

Coleoid cephalopods (squid, cuttlefish, and octopus) are characterized by a highly derived body plan compared to the other molluscs, with the main novelties being a partial or complete loss of the shell, a crown of flexible arms provided with suckers (Boletzky, 2003), camera-type eyes, and a nervous system considered to be the most complex among invertebrates (Young, 1963). Such phenotypic features are further closely related to the active predatory lifestyle and the wide variety of behaviors in extant cephalopods (Hanlon and Messenger, 2018). In recent years, cephalopods emerged as intriguing organisms in the genome evolution field as they showcase several types of genomic features, including rearrangements of bilaterian-conserved local gene linkages, gene duplications, orphan gene emergence, and repeat element expansions (Albertin et al., 2015; Belcaid et al., 2019). These signatures at different levels of genome organization were associated with the evolution of distinct organs within a single organism (Belcaid et al., 2019) and are likely to have co-evolved, comprising a complex evolutionary genome signature that ultimately contributed to the phenotypic novelties of cephalopods (Ritschard et al., 2019). Even though transposable elements (TEs) were initially classified as "junk" (Ohno, 1972) or "selfish" DNA (Doolittle and Sapienza, 1980; Orgel and Crick, 1980), their role as important mutation sources and therefore as determinants in the evolution of their hosts is now established. Indeed, depending on the target and mode of their transposition and recombination, mobile elements can be exapted to new *cis*-regulatory elements (Britten, 1996; Marino-Ramirez et al., 2005), disrupt or rewire regulatory networks (Feschotte, 2008; Moschetti et al., 2020; Sundaram and Wysocka, 2020), and cause chromosomal-level rearrangements (Gray, 2000). Besides, TEs are important tools for the development of new genomic integration (Sandoval-Villegas et al., 2021) and expression vector technologies (Palazzo and Marsano, 2021). TEs are present in every eukaryotic genome in very different proportions and classes (Wells and Feschotte, 2020), with both random drift and natural selection contributing to their differential amplification in divergent lineages (Lynch and Conery, 2003; Kent et al., 2017). About half of every sequenced coleoid cephalopod genome comprises repetitive DNA, whose composition significantly differs across lineages: SINEs are the main components of *Octopus bimaculoides* and *O. vulgaris* transposomes; LINEs prevail in *O. minor* and *Euprymna scolopes*, whereas mostly DNA elements are present in the *Architeuthis dux* genome (Albertin et al., 2015; Kim et al., 2018; Belcaid et al., 2019; Zarrella et al., 2019; Fonseca et al., 2020). Unlike coleoids, the *Nautilus pompilius* genome is smaller, is less repetitive (31%), and lacks the many genomic features of coleoid cephalopods (Zhang et al., 2021). Despite no functional survey being available, TEs are found to be extensively expressed in *O. bimaculoides* and *O. vulgaris* tissues (Albertin et al., 2015; Petrosino et al., 2021); furthermore, regions nearby loci that underwent rearrangements in coleoid cephalopods are rich in repeats in *O. bimaculoides*, just as orphan genes associated with novel structures are in *E. scolopes* (Albertin et al., 2015; Belcaid et al., 2019; Petrosino et al., 2021). Such observations highlight the central role that TEs might have played in cephalopod diversification. Although many of the repeat families have expanded recently in individual lineages, their role in shaping the ancestral coleoid cephalopod genome remains elusive. Furthermore, information about repeats in mollusks is fragmented as it is not usually presented with a wide comparative purpose (Zhang et al., 2012; Simakov et al., 2013; Wang et al., 2017; Powell et al., 2018; Kenny et al., 2020; Zeng et al., 2020); additionally, the number of sequenced cephalopod species is scarce. This hinders the systematic comparison of TE content within a clade, making it hard to have an overview of the present and past cephalopod repeat landscape. Our study aims to make a first step in this direction by providing a common repeat annotation of the main cephalopod lineages and extrapolating with a comparative approach the ancient TE landscape that possibly existed in the stem coleoid lineage. To this end, we considered the genome assemblies of the coleoids *O. vulgaris*, *O. bimaculoides*, *A. dux*, *E. scolopes*, and *N. pompilius*. Octopuses' common ancestor dates back to ~25 Mya (Uribe and Zardoya, 2017) and that of coleoids dates back to ~270 Mya (Tanner et al., 2017), while *Nautilus* lineage diverged ~415 Mya from coleoids (Bergmann et al., 2006; Kröger et al., 2011). We characterized both the total and divergence-based repeat contents in every species. Based on sequence divergence, we identified shared ancient TE families present across coleoid genomes. Finally, using sequence similarity network approaches, we could reveal complements of closely related squid and octopus sequences among the most abundant TE families, possibly hinting at their common origin back in the coleoid lineage.

METHODS

We used the scaffold-level genome assemblies of *O. vulgaris*, *O. bimaculoides*, *A. dux*, and *N. pompilius*, publicly available under GenBank accession numbers GCA_003957725.1, GCA_001194135.1, GCA_006491835.1, and GCA_018389105.1, respectively. A chromosomal-scale assembly generated with LACHESIS (Burton et al., 2013) was used for *E. scolopes* (Schmidbaur et al., in review, http://metazoa.csb.univie.ac.at/data/v2/). Completeness of genomes was assessed with BUSCO 5.2.2 (Manni et al., 2021) by considering the 954 conserved orthologs of the metazoa_odb10 database and with technical statistics supplied by Quast 5.0.2 (Gurevich et al., 2013) (**Supplementary Table S1**). For each assembly, the same repeat annotation workflow was employed: a family library was generated with RepeatModeler 2.0 (Flynn et al., 2020) and used to annotate and mask each starting assembly with RepeatMasker 4.0.9 (Smit et al., 2020); in order to uncover further sequences that were not detected in the first masking round, these steps were performed a second time on the previously hard-masked genome (double-masking, as employed in Meyer et al., 2021); a defragmentation step of all the obtained sequences was then carried out with RepeatCraft in the "strict" merge mode (Wong and Simakov, 2019).

Custom Bash, Python, and R scripts were used to filter and parse the data for the assessment of repeat content. Because the "Unknown" and "Simple_repeat" categories constituted a

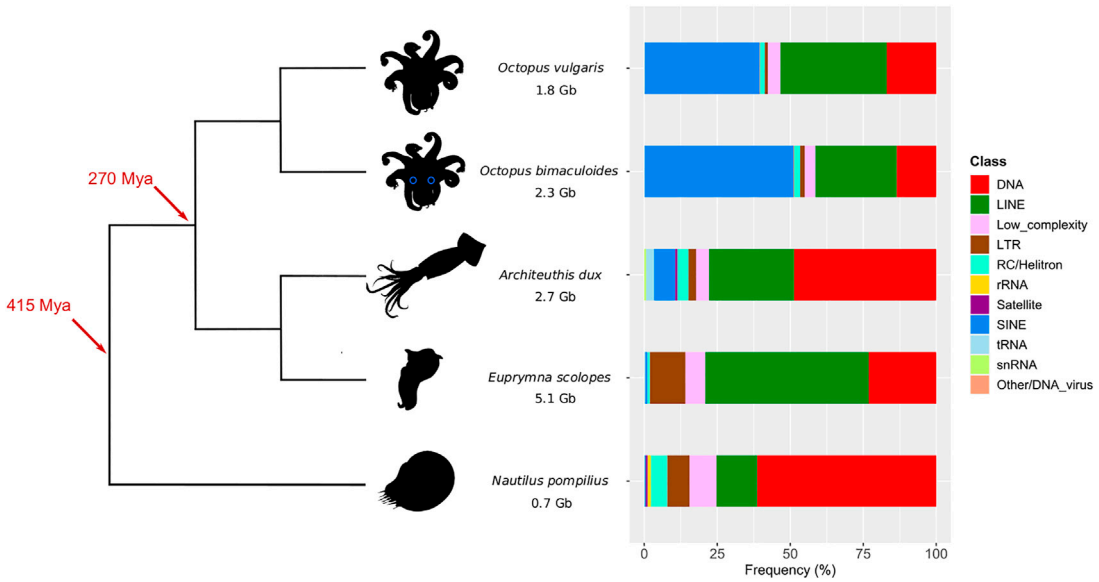

FIGURE 1 | Phylogeny of *O. vulgaris* (common octopus), *O. bimaculoides* (California two-spot octopus), *A. dux* (giant squid), *E. scolopes* (Hawaiian bobtail squid), and *N. pompilius* (chambered nautilus) with their corresponding assembly size and total content of repeat subclasses. Time marks on the tree refer to the coleoid last common ancestor (270 Mya) and the split of the nautiloid outgroup (415 Mya) based on Tanner et al. (2017). Percentages in the barplot are calculated as subclass counts out of the total number of elements in every assembly (unknown and simple repeats are excluded).

significant portion of the total repeats (see **Supplementary Table S1**) but were not of interest for our purpose, they were discarded to obtain a clearer landscape of the known TEs. Any repeat content that is henceforth referred to is therefore intended as deprived of unknown and simple repeats. Total repeat composition was assessed for every assembly in terms of subclass and family raw counts. Such content was then split into three contiguous intervals of divergence from consensus, namely, 0–10, 10–30, and >30%, as defined by RepeatMasker estimation with the Kimura distance-based method. We then looked for expression evidence by comparing RNA-seq data from different tissues with the repeat annotations to have an overview of the repeat complement activity of every species, except *A. dux*, for which transcriptomic data are not available (for data accessions, see **Supplementary Table S3**). After adapter and quality trimming (TrimGalore 0.6.5, Krueger, 2015), the reads were mapped to their genome with Hisat2 2.1.0 (Kim et al., 2019) and their coordinates were intersected with the repeat annotations in bedtools 2.29.2 with an overlap of 100% for the repeat sequences (Quinlan & Hall, 2010). Regardless of the expression pattern, weighted TE family composition in every bin, both with all families and with only shared families, was used to estimate Euclidean distances between species and carry out a principal component analysis (PCA). An "ancient" repeat subset was extracted by retaining only TE families represented in the >30% bin of every species. A 30% cutoff was chosen to identify old repeat copies as this distance is close to the RepeatMasker distance detection limit (around 50%): indeed, 5% maximum of the total elements was detected beyond this distance, and even fewer elements were found above 40% divergence (**Supplementary Table S1**). Such a complement was further characterized in *O. bimaculoides* and *E. scolopes*. For each family, the relationship between raw repeat counts per chromosome and chromosome sizes was estimated in *E. scolopes*. Finally, octopus and Hawaiian bobtail squid sequences of all divergence values from some of the most abundant families—CR1, RTE-BovB, Dong-R4, Penelope, and TcMar-Tc1—were compared with blastn from ncbiblast+ 2.10.0 with search options -task blastn and -word_size 18 (Altschul et al., 1990). A distance calculated as the number of mismatches/alignment length was assigned to each pairwise hit and used to resolve intra- and inter-species relations within each TE family. The R packages igraph, ggplot2, RcolorBrewer, and plyr were used for graphically representing the distances. Since the overall repeats were too many to be handled by R, the entire set of sequences in a bin was retained when possible, but in most cases, a downsampling of 1% or 10% was applied to obtain a readable graph. In addition to this distance-based network approach, we looked for homologies between cephalopod repeats and sequences of distantly related taxa that could hint at potential horizontal gene-transfer events (HGT) underlying cephalopod repeat bursts. To do this, we conducted BLAST searches of the TE family consensi in Dfam 3.5 (Storer et al., 2021) by considering all the hits with an e-value < 1e-50 and a bit-score > 50 significant.

RESULTS

Improved Annotation of the Cephalopod Repeat Complement

Roughly 40–50% of the total coleoid assembly lengths were masked in the first round, whereas only 30% of the *Nautilus pompilius*

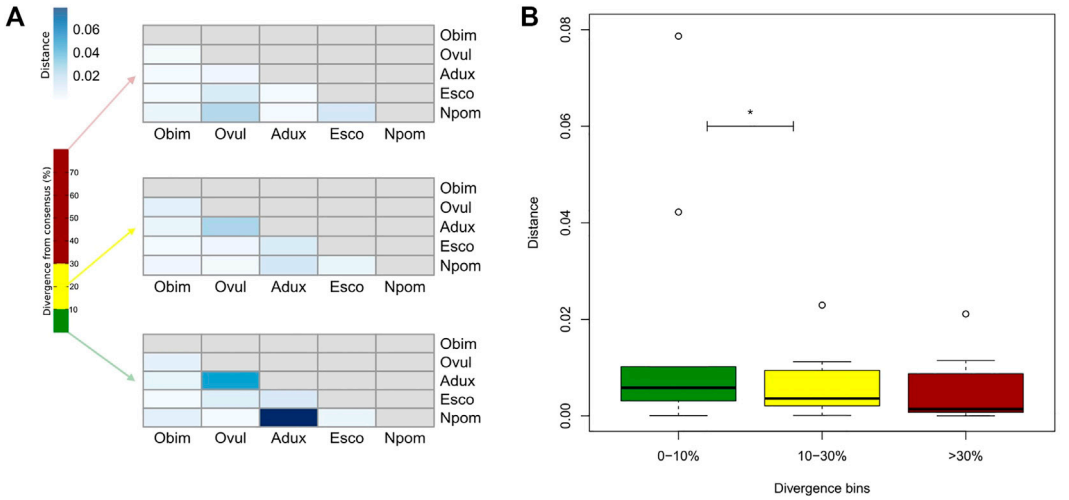

FIGURE 2 | (A) Euclidean distances between species according to divergence from consensus of their repeats. Distance (quantitative variable in blue) is calculated on the normalized raw frequencies of just TE families shared by all species in each divergence bin (qualitative variable in percentage). Divergence bins are defined as 0–10% (green), 10–30% (yellow), and >30% (red) ranges. **(B)** The same quantitative distance values for the same qualitative divergence bins are displayed in a boxplot. Black horizontal lines correspond to medians, boxes' lower and upper ends respectively to the first and third quartiles, whiskers' lower and upper ends respectively to the minimum and maximum values, and empty circles to the outlier distance values. Asterisks indicate $p < 0.05$ for the Wilcoxon test calculated between the respective distance sets. Adux = *A. dux*; Esco = *E. scolopes*; Npom = *N. pompilius*; Obim = *O. bimaculoides*; Ovul = *O. vulgaris*.

genome was masked. An additional 2–6% was uncovered in the second round of the hard-masked genome, highlighting the importance of the second round of genome masking. As a result, the double masking revealed the repeat content to constitute about half of all the genomes considered, except for *Nautilus* (**Supplementary Table S1**). The double masking has been proven to be a useful approach for capturing huge amounts of repetitive DNA in noticeably big genomes, such as that of *Neoceratodus forsteri* (Meyer et al., 2021). In our case, cephalopod genomes are around 10-fold smaller and less repetitive than the Australian lungfish genome. Even so, TE annotation was enhanced in terms of both sequence quantity and number of detected families; for instance, the second masking round allowed to identify SINEs in *E. scolopes*, which were completely unannotated after just one round. The RepeatCraft step was then able to merge from a minimum of about 53,000 repeat copies in *O. vulgaris* to a maximum of 152,000 in *E. scolopes* (**Supplementary Table S1**), allowing for the reconstruction of degenerated and fragmented elements.

Total TE Composition and Activity of TEs in Cephalopod Genomes

As shown in **Figure 1**, octopus TE subclass compositions are similar between each other, with a major SINE (~40%) and LINE portion (~30%), followed by DNA elements (~17%). Decapodiformes show instead a different landscape: *E. scolopes* features mostly LINEs (56%) and secondly DNA (23%) and LTR subclasses (12%), while SINEs are very scarcely represented (<1%); *A. dux* repeat content mainly consists of DNA elements (49%) and LINEs (29%). Despite having a much more restrained genome (see **Supplementary Table S1**), the *Nautilus* repeatome is similar to the giant squid one in that the first major subclass is DNA (61%) and the second one is LINE (14%). At the TE family level, tRNA-Core and tRNA-Deu are the main contributors to the octopus-like SINE complement; in *E. scolopes*, LINEs and LTRs are mainly represented by CR1 (29%) and Gypsy elements (11%), respectively. Both *A. dux* and *Nautilus* DNA repeat contents are not defined by one prevailing family but by diverse ones, such as TcMar-Mariner, hAT-Charlie, TcMar-Tc1, hAT-Tip100, and TcMar-Tigger, which also contribute to the DNA element content of the other species (**Supplementary Figure S1**, **Supplementary Table S2**).

Overall, the portion of the repetitive genome and subclass composition of each species are consistent with the literature (Albertin et al., 2015; Belcaid et al., 2019; Zarrella et al., 2019; Fonseca et al., 2020; Zhang et al., 2021 (see Discussion for details). The mapping of transcriptomic data against genomes and calculating their overlap with repeat annotations revealed the proportion of elements expressed in at least one of the sampled tissues. A substantial proportion of repeat loci showed putative expression. While we found large differences in the proportion of loci with at least one transcriptomic read, with *O. vulgaris* having the lowest (39%) and *N. pompilius* having the highest (92%), this is likely a result of the underlying assembly quality. Moreover, the counts for each repeat category vary between tissues, which may be a result of tissue-specific TE activity within a single organism (**Supplementary Table S3**).

Divergence Decomposition Reveals an Ancient Repeat Subset

We find only a slight decrease in the transcriptional activity of older element loci (>30% divergence) in *E. scolopes* and *O.*

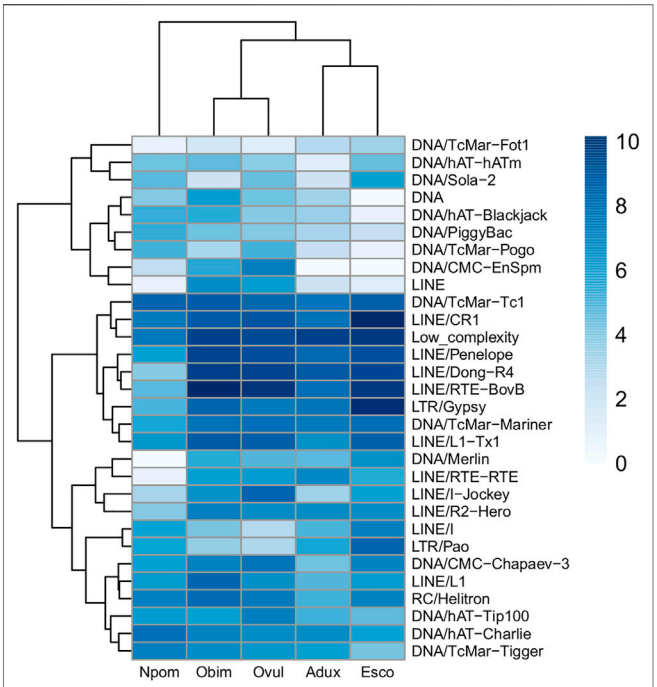

FIGURE 3 | Anciently shared repeat complement obtained from TE families retained by all species at >30% divergence values. TE families are clustered with the 'complete' method of the pheatmap package, and species are clustered according to the phylogeny in **Figure 1**. Log-scaled values of raw element counts are shown.

vulgaris compared to the younger age categories, both overall and at the tissue level (**Supplementary Table S3**). 0–10 and 10–30% divergence complements are in general more abundant in the genome than in the >30% subset for both the number of TE families in at least one genome and the maximum raw count for a family in a given assembly. Lineage-specific expansions such as those of tRNA-Core, tRNA-Deu, and CR1 recur throughout all the bins as well as some more abundant elements shared by all species, such as LINEs Penelope, Dong-R4 and RTE-BovB, and the DNA elements Mariner and Tc1 (**Supplementary Figure S1**). Interspecies distances calculated on both shared families and all families are higher in the 0–10% divergence complement and tend to lower as the divergence increases (**Figure 2**; **Supplementary Figure S2**). The highest distances are those of *A. dux* against *Nautilus* and *O. vulgaris* and are generally consistent with the differences in repeat family abundance and weights on principal components (PCs) in each bin (**Supplementary Figures S1, S3**). The extracted anciently shared repeat complement is formed by 15 DNA families, 11 LINEs, 2 LTRs, and Helitrons, (plus tRNA and low-complexity elements) (**Figure 3**). Almost all families show vastly different genomic abundances across species: in particular, CR1 and Gypsy elements stand out in *E. scolopes*, just as RTE-BovB does in octopuses. Moreover, a specific subset composed of LINEs RTE-BovB, Dong-R4, Penelope, L1-Tx1, CR1, LTR/Gypsy, and TcMar-Tc1 and Mariner DNA elements is expanded in three coleoids, while *Nautilus* and *Architeuthis* show significantly lower copy numbers (p-Wilcoxon < 0.05). Although SINEs are very abundant in octopuses, they are underrepresented in Decapodiformes and completely missing from this common ancient coleoid cephalopod repeat set. Raw abundance counts per chromosome of sequences at all divergence levels have linear relationships with chromosome sizes (**Supplementary Figure S4**). Consistent with the previous observations of possible lineage-specific expansions, the BLAST analysis revealed at least two LINE CR1 bursts in the *E. scolopes* genome and just as many RTE-BovB expansions in the *O. bimaculoides* genome. We also identify smaller expansions of LINE families Dong-R4 and Penelope and DNA/TcMar-Tc1 as octopus- and Hawaiian bobtail squid-specific. Furthermore, the sequence similarity search highlights considerable octopus-squid copy co-groupings for all the families considered (**Figure 4**). Despite the effort made to make inter- and intraspecies sequence hit proportions as balanced as possible, exactly even retention of both in the search output was not reached (**Supplementary Figure S5**). The possibility that the marked bias favoring same-species matches could affect to some extent the net plot arrangement should be taken into consideration. The research in Dfam gave significant hits for 12 DNA and 3 LINE families, with TcMar-Tc1, Mariner, and Tigger having the highest number of hits in the database and *A. dux* being the species with the highest number of overall matches (**Supplementary Table S4**).

DISCUSSION

The Repeat Landscape of Cephalopods

By considering five cephalopod species as a proxy of the present diversity, we were able to integrate a common repeat annotation of the available representatives of this clade and to identify the diverging expansion histories that characterize each lineage. Our results at the subclass level are strongly consistent with the literature, and our annotations at the TE family level add valuable knowledge in the context of cephalopod genome architectures. The discrepancy in the number of active elements (as inferred by RNA-seq mapping) across species could be correlated with genome assembly quality. It is worth noting that the *Nautilus* genome, which has the highest proportion of active repeats, is also the only gapless assembly and the one with the highest alignment score of RNA-seq reads. In addition to the technical limitations of fragmented genomes, another reason could be an actual stronger inhibition of transcription that might be at play in genomes more extensively colonized by selfish elements. Despite these factors, the results reported in **Supplementary Table S3** are consistent with the expression of a substantial portion of the annotated repeatomes.

Sequence Divergence Decomposition Accounts for Different Phylogenetic Signal Between Species

The general trend of lower genomic counts above the 30% divergence level measured from the repeat consensus is due to

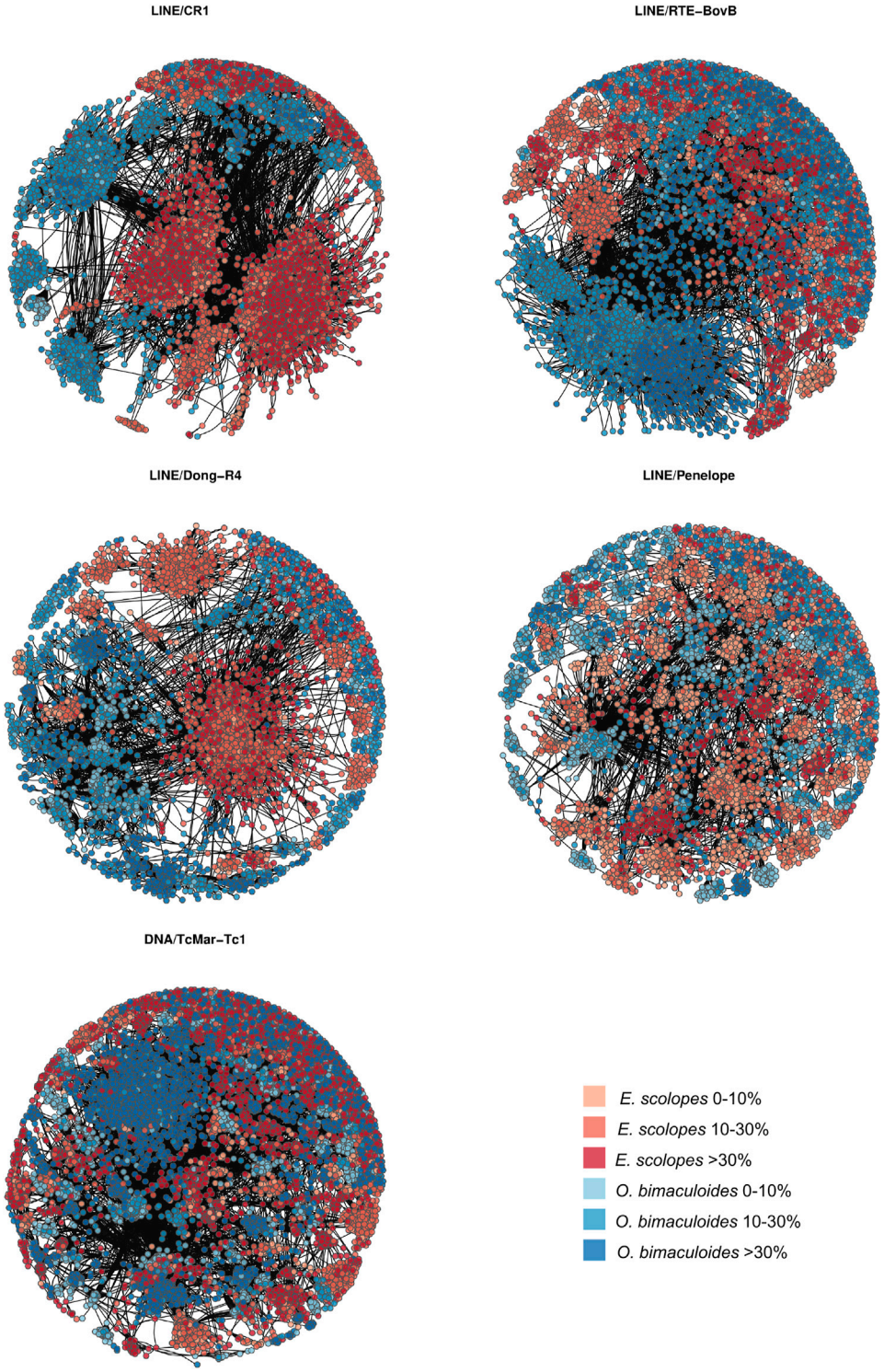

FIGURE 4 | Sequence similarity-based net plots for LINEs CR1, RTE-BovB, Dong-R4, Penelope, and DNA/TcMar-Tc1. Each point corresponds to one TE copy, whose color and shade correspond to a given species and divergence range as per legend and whose position depends on the ratio number of mismatches/alignment length assigned by BLAST.

the decreasing ability of RepeatMasker to find repeats as their divergence to the consensus increases as well as many ancient sequences being lost from the genome. However, we consider the recurrence of a given TE set in the highest divergence bin of all species as a strong signal of TE basal retention across coleoids and some in their outgroup. In support of this, interspecies distances are on the whole higher in the 0–10% interval and progressively lower in the 10–30% and >30% intervals (**Figure 2**, **Supplementary Figure S2**). Repeat composition at different divergence windows can thus be accounted for with a good approximation for more recent or ancestral scenarios: 0–10% complements tend to mirror specific novel TE bursts or new family emergence, causing more marked differences; conversely, >30% divergence contents should consist of conserved families which make species more akin to each other. TE activity patterns can significantly vary among lineages, even in the case of a recent evolutionary split (Boulesteix et al., 2006), meaning that the comparison of TE content does not necessarily reflect species phylogeny. In our observation, distance results calculated considering all families were similar to those based only on shared ancient families. 0–10% divergence-based PCA places species according to phylogeny along PC1, and among TE families mainly responsible for differences are LINE/RTE-BovB and CR1, the main ones subject to differential expansions (**Supplementary Figure S3**). Moreover, as divergence increases, octopuses generally cluster together, while *Nautilus* tends to move closer to Decapodiformes, especially *A. dux*, consistently with the different repeat expansion patterns highlighted in the ancient repeat complement (see next paragraph).

The Ancestral Coleoid Repeat Complement: TE Subclass Composition Insights From the Comparison Across Species

The anciently shared repeat complement obtained primarily consists of LINEs, DNA elements, and one LTR family. SINEs are not present, as reflected in their low counts in *E. scolopes* and *Nautilus*. The considerable length of the *E. scolopes* genome (5.1 Gb) combined with the difficulty in sequencing short interspersed elements could have misled SINE representation. Nevertheless, as suggested by Albertin et al. (2015) and considering the lack of SINE enrichments in other Decapodiformes, the SINEs that we were able to recover likely constitute expansions specific to octopuses. It is important to note that the ancient repeat set shared across *coleoid* species does include some SINE families (**Supplementary Figure S1**), suggesting that these retroelements could have been active in the genome of their common ancestor. The slow evolutionary rate and the repeat content found in the *Nautilus* genome by Zhang et al. (2021) might suggest the retention of signatures similar to those of the pre-radiating coleoid ancestor. Therefore, the fact that *Nautilus* generally lacks highly divergent SINEs points to their actual absence in the ancient repeat complement of cephalopods. Whether and to what extent SINEs also initially contributed to the ancestral cephalopod genome remain unclear due to SINE fast evolution and sequence decay that may have occurred during more than 270 million years. As shown by **Supplementary Figure S3**, *Nautilus* and *A. dux* cluster separately from the other species because of the weaker genomic expansion of their shared complement, especially LINEs and LTRs; DNA elements instead display more restrained expansion patterns in all species (**Figure 3**). Assuming that these TE subclasses were all present in the common ancestor, this suggests the cephalopod and molluscan plesiomorphic and conserved nature of the DNA transposon complement and the dynamic nature and more recent activity of some LINEs that expanded in the coleoid ancestor.

Chromosomal Distribution and Expansion Patterns of Anciently Shared TE Families

The most enriched families emerging in the ancient complement are LINEs Penelope, Dong-R4, CR1, L1-Tx1, L2, RTE-BovB, and DNA/TcMar-Tc1, as well as LTR/Gypsy. Among them, as already mentioned, CR1, RTE-BovB, and Gypsy elements show clear lineage-specific expansions. The linear relationships of element count against chromosome size revealed that TE families belonging to the ancestral complement are not arranged into any chromosomal hotspots in *E. scolopes*: the pattern is the same for both sequences close to and divergent from consensus, meaning that both recent and older TE outbreaks did not occur in specific chromosomes in this species. However, this remains to be verified in other species and does not rule out possible enrichments at finer scales and linked to different terms such as Gene Ontology (GO) or cephalopod-specific synteny (gene order) loci. The scattered distribution of TEs across the genome of *E. scolopes* agrees, however, with the scenario of the extensive and long-standing reshuffling that has arisen in coleoid genomes (Albertin and Simakov, 2020). Additionally, the directly proportional contribution of TEs to chromosome lengths is consistent with the hypothesis that genome size is directly influenced by repetitive DNA (Kidwell, 2002; Naville et al., 2019).

The fact that repeat sets that we deem as apomorphic are still included in the ancient complement stresses the limit of sequence divergence-based methods as we are not able to clearly isolate actual ancestral repeat subgroups. Notwithstanding, the network-based approach identifies clusterings that do not conform with the divergence bins we defined, as both independent outbursts and interspecies groupings appear to consist of all divergence values (**Figure 4**). This might be a valuable approach for discriminating between recently proliferated elements and the more interspecies connected ancestral and conserved copies that are putative remnants of the ancient expansions. The common octopus-squid clusters could thus be informative in revealing ancient repeats across such divergent lineages, potentially pointing to conserved TE subsets in coleoids.

Although the similarity networks and our Dfam similarity analysis suggest that repeat bursts occurred through vertical transmission, we cannot rule out occasional horizontal transfer events for more ancient elements. While we did not find evidence for homology across long-diverged taxa for CR1, Penelope, and

Dong-R4, hits were obtained for RTE-BovB and TcMar-Tc1 (in addition to other DNA elements), mostly corresponding to aquatic vertebrates. Nevertheless, most of these species were the most closely related to cephalopods in Dfam. The origin of these repeat elements in cephalopods is therefore equally likely via vertical transmission.

Conclusion and Next Steps

The family repeat content was outlined in five cephalopod species, and a preliminary assessment of an ancestral TE set was made by considering the most divergent repeat sequences. This allowed us to distinguish between lineage-specific, shared, and stem-coleoid expanded repeat elements. An additional sequence similarity-based analysis of some ancestrally shared families revealed more accurate patterns of independent and interspecies expansions, therefore highlighting a possible partially shared history of such repeat families. The comparative profiling here described is preliminary work, and the inclusion of new key species and chromosome-level data will be essential for making the coleoid and cephalopod TE landscape more robust. Indeed, the recent genome sequencing of *Nautilus* added an important comparative point to our study as the only coleoid outgroup, and future acquisition of new data regarding nautiloids and new coleoid species will be fundamental for investigating the cephalopod repeatome evolution. Similarly, further studies such as gene ontology enrichment, orthology construction, and synteny breakage enrichment could shed light on whether the TE subgroups obtained with our method were actually involved in cephalopod genome reshuffling and to test our approach to track down the repeat complement of the early (coleoid) cephalopods.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: https://www.ncbi.nlm.nih.gov/assembly/GCA_018389105.1/ GenBank assembly accession: GCA_018389105.1, https://www.ncbi.nlm.nih.gov/assembly/GCA_003957725.1/ GenBank assembly accession: GCA_003957725.1, https://www.ncbi.nlm.nih.gov/assembly/GCF_001194135.1/ GenBank assembly accession: GCA_001194135.1, https://www.ncbi.nlm.nih.gov/assembly/GCA_006491835.1/ GenBank assembly accession: GCA_006491835.1, http://metazoa.csb.univie.ac.at/data/v2/, https://www.ncbi.nlm.nih.gov/sra/?term=SRR2047118 SRA accession: SRR2047118, https://www.ncbi.nlm.nih.gov/sra/?term=SRR7645642 SRA accession: SRR7645642, https://www.ncbi.nlm.nih.gov/sra/?term=SRR2047116 SRA accession: SRR2047116, https://www.ncbi.nlm.nih.gov/sra/?term=SRR2047109 SRA accession: SRR2047109, https://www.ncbi.nlm.nih.gov/sra/?term=SRR2047111 SRA accession: SRR2047111, https://www.ncbi.nlm.nih.gov/sra/?term=SRR2857274 SRA accession: SRR2857274, https://www.ncbi.nlm.nih.gov/sra/?term=SRR7548187 SRA accession: SRR7548187, https://www.ncbi.nlm.nih.gov/sra/?term=SRR13005724 SRA accession: SRR13005724, https://www.ncbi.nlm.nih.gov/sra/?term=SRR8159234 SRA accession: SRR8159234, https://www.ncbi.nlm.nih.gov/sra/?term=SRR8172522 SRA accession: SRR8172522, https://www.ncbi.nlm.nih.gov/sra/?term=SRR3493852 SRA accession: SRR3493852, https://www.ncbi.nlm.nih.gov/sra/?term=SRR2857280 SRA accession: SRR2857280, https://www.ncbi.nlm.nih.gov/sra/?term=SRR13131286 SRA accession: SRR13131286.

AUTHOR CONTRIBUTIONS

AM and OS designed the project. AM performed the analyses, with help from AK and WW. OS supervised the project. AM and OS wrote the manuscript. AK, FG, and WW gave critical feedback and contributed to the final version of the manuscript.

FUNDING

AM, AK, WW, and OS were supported by the Austrian Science Fund (FWF) grant P30686-B29. AM was supported by the 2020/2021 Erasmus+ Mobility for Traineeship and by the Department of Biological, Geological, and Environmental Sciences of the University of Bologna with a scholarship for the preparation of the thesis abroad.

ACKNOWLEDGMENTS

We wish to thank Hannah Schmidbaur for providing the assembly data for *E. scolopes*. Computations were performed for the most part on the Life Science Computer Cluster at the University of Vienna.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.793734/full#supplementary-material

REFERENCES

Albertin, C. B., and Simakov, O. (2020). Cephalopod Biology: at the Intersection between Genomic and Organismal Novelties. *Annu. Rev. Anim. Biosci.* 8, 71–90. doi:10.1146/annurev-animal-021419-083609

Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., and Edsinger-Gonzales, E. (2015). The octopus Genome and the Evolution of Cephalopod Neural and Morphological Novelties. *Nature* 524 (7564), 220–224. doi:10.1038/nature14668

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215 (3), 403–410. doi:10.1016/S0022-2836(05)80360-2

Belcaid, M., Casaburi, G., McAnulty, S. J., Schmidbaur, H., Suria, A. M., Moriano-Gutierrez, S., et al. (2019). Symbiotic Organs Shaped by Distinct Modes of Genome Evolution in Cephalopods. *Proc. Natl. Acad. Sci.* 116 (8), 3030–3035. doi:10.1073/pnas.1817322116

Bergmann, S., Lieb, B., Ruth, P., and Markl, J. (2006). The Hemocyanin from a Living Fossil, the Cephalopod *Nautilus Pompilius*: Protein Structure, Gene Organization, and Evolution. *J. Mol. Evol.* 62 (3), 362–374. doi:10.1007/s00239-005-0160-x

Boletzky, S. V. (2003). Biology of Early Life Stages in Cephalopod Molluscs. *Adv. Mar. Biol.* 44, 144–204. doi:10.1016/s0065-2881(03)44003-0

Boulesteix, M., Weiss, M., and Biémont, C. (2006). Differences in Genome Size between Closely Related Species: the *Drosophila melanogaster* Species Subgroup. *Mol. Biol. Evol.* 23 (1), 162–167. doi:10.1093/molbev/msj012

Britten, R. J. (1996). Cases of Ancient mobile Element DNA Insertions that Now Affect Gene Regulation. *Mol. Phylogenet. Evol.* 5 (1), 13–17. doi:10.1006/mpev.1996.0003

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 (12), 1119–1125. doi:10.1038/nbt.2727

Doolittle, W. F., and Sapienza, C. (1980). Selfish Genes, the Phenotype Paradigm and Genome Evolution. *Nature* 284 (5757), 601–603. doi:10.1038/284601a0

Feschotte, C. (2008). Transposable Elements and the Evolution of Regulatory Networks. *Nat. Rev. Genet.* 9 (5), 397–405. doi:10.1038/nrg2337

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families. *Proc. Natl. Acad. Sci.* 117 (17), 9451–9457. doi:10.1073/pnas.1921046117

Fonseca, D., Couto, A., Machado, A. M., Brejova, B., Albertin, C. B., Silva, F., et al. (2020). A Draft Genome Sequence of the Elusive Giant Squid, Architeuthis Dux. *GigaScience* 9 (1), giz152. doi:10.1093/gigascience/giz152

Gray, Y. H. (2000). It Takes Two Transposons to Tango: Transposable-Element-Mediated Chromosomal Rearrangements. *Trends Genet.* 16 (10), 461–468. doi:10.1016/S0168-9525(00)02104-1

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* 29 (8), 1072–1075. doi:10.1093/bioinformatics/btt086

Hanlon, R. T., and Messenger, J. B. (2018). *Cephalopod Behaviour*. Cambridge University Press.

Kenny, N. J., McCarthy, S. A., Dudchenko, O., James, K., Betteridge, E., and CortonWilliams, C. S. T. (2020). The Gene-Rich Genome of the Scallop *Pecten maximus*. *GigaScience* 9 (5), giaa037. doi:10.1093/Fgigascience/Fgiaa037

Kent, T. V., Uzunović, J., and Wright, S. I. (2017). Coevolution between Transposable Elements and Recombination. *Phil. Trans. R. Soc. B: Biol. Sci.* 372 (1736), 20160458. doi:10.1098/rstb.2016.0458

Kidwell, M. G. (2002). Transposable Elements and the Evolution of Genome Size in Eukaryotes. *Genetica* 115 (1), 49–63. doi:10.1023/a:1016072014259

Kim, B. M., Kang, S., Ahn, D. H., Jung, S. H., Rhee, H., Yoo, J. S., et al. (2018). The Genome of Common Long-Arm octopus Octopus Minor. *Gigascience* 7 (11), giy119. doi:10.1093/gigascience/giy119

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* 37 (8), 907–915. doi:10.1038/s41587-019-0201-4

Kröger, B., Vinther, J., and Fuchs, D. (2011). Cephalopod Origin and Evolution: a Congruent Picture Emerging from Fossils, Development and Molecules: Extant Cephalopods Are Younger Than Previously Realised and Were under Major Selection to Become Agile, Shell-Less Predators. *Bioessays* 33 (8), 602–613. doi:10.1002/bies.201100001

Krueger, F. (2015). Trim Galore!: A Wrapper Tool Around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files. Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Lynch, M., and Conery, J. S. (2003). The Origins of Genome Complexity. *science* 302 (5649), 1401–1404. doi:10.1126/science.1089370

Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A., and Zdobnov, E. M. (2021). *BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes*. arXiv preprint arXiv:2106.11799. doi:10.1093/molbev/msab199

Marino-Ramirez, L., Lewis, K. C., Landsman, D., and Jordan, I. K. (2005). Transposable Elements Donate Lineage-specific Regulatory Sequences to Host Genomes. *Cytogenet. Genome Res.* 110 (1-4), 333–341. doi:10.1159/000084965

Meyer, A., Schloissnig, S., Franchini, P., Du, K., Woltering, J. M., and IrisarriSchartl, I. M. (2021). Giant Lungfish Genome Elucidates the Conquest of Land by Vertebrates. *Nature* 590 (7845), 284–289. doi:10.1038/s41586-021-03198-8

Moschetti, R., Palazzo, A., Lorusso, P., Viggiano, L., and Massimiliano Marsano, R. (2020). "What You Need, Baby, I Got it": Transposable Elements as Suppliers of Cis-Operating Sequences in drosophila. *Biology* 9 (2), 25. doi:10.3390/biology9020025

Naville, M., Henriet, S., Warren, I., Sumic, S., Reeve, M., Volff, J. N., et al. (2019). Massive Changes of Genome Size Driven by Expansions of Non-autonomous Transposable Elements. *Curr. Biol.* 29 (7), 1161–1168. doi:10.1016/j.cub.2019.01.080

Ohno, S. (1972). So Much'junk'DNA in Our Genome. In Evolution of Genetic Systems. *Brookhaven Symp. Biol.*, 366–370.

Orgel, L. E., and Crick, F. H. (1980). Selfish DNA: the Ultimate Parasite. *Nature* 284 (5757), 604–607. doi:10.1038/284604a0

Palazzo, A., and Marsano, R. M. (2021). Transposable Elements: a Jump toward the Future of Expression Vectors. *Crit. Rev. Biotechnol.*, 1–27. doi:10.1080/07388551.2021.1888067

Petrosino, G., Ponte, G., Volpe, M., Zarrella, I., Langella, C., Di Cristina, G., et al. (2021). *Identification of LINE Retrotransposons and Long Non-coding RNAs Expressed in the octopus Brain*. bioRxiv [preprint](Accessed May 8 2021). doi:10.1101/2021.01.24.427974

Powell, D., Subramanian, S., Suwansa-Ard, S., Zhao, M., O'Connor, W., Raftos, D., et al. (2018). The Genome of the Oyster Saccostrea Offers Insight into the Environmental Resilience of Bivalves. *DNA Res.* 25 (6), 655–665. doi:10.1093/dnares/dsy032

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26 (6), 841–842. doi:10.1093/bioinformatics/btq033

Ritschard, E. A., Whitelaw, B., Albertin, C. B., Cooke, I. R., Strugnell, J. M., and Simakov, O. (2019). Coupled Genomic Evolutionary Histories as Signatures of Organismal Innovations in Cephalopods: Co-evolutionary Signatures across Levels of Genome Organization May Shed Light on Functional Linkage and Origin of Cephalopod Novelties. *Bioessays* 41 (12), 1900073. doi:10.1002/bies.201900073

Sandoval-Villegas, N., Nurieva, W., Amberger, M., and Ivics, Z. (2021). Contemporary Transposon Tools: A Review and Guide through Mechanisms and Applications of Sleeping Beauty, piggyBac and Tol2 for Genome Engineering. *Int. J. Mol. Sci.* 22 (10), 5084. doi:10.3390/ijms22105084

Simakov, O., Marletaz, F., Cho, S. J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., et al. (2013). Insights into Bilaterian Evolution from Three Spiralian Genomes. *Nature* 493 (7433), 526–531. doi:10.1038/nature11696

Smit, A. F. A., Hubley, R., and Green, P. (2020). RepeatMasker. Available at: http://repeatmasker.org (Accessed October, 2020).

Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., and Smit, A. F. (2021). The Dfam Community Resource of Transposable Element Families, Sequence Models, and Genome Annotations. *Mobile DNA* 12 (1), 1–14. doi:10.1186/s13100-020-00230-y

Sundaram, V., and Wysocka, J. (2020). Transposable Elements as a Potent Source of Diverse Cis-Regulatory Sequences in Mammalian Genomes. *Philosophical Trans. R. Soc. B* 375 (1795), 20190347. doi:10.1098/rstb.2019.0347

Tanner, A. R., Fuchs, D., Winkelmann, I. E., Gilbert, M. T. P., Pankey, M. S., Ribeiro, M., et al. (2017). Molecular Clocks Indicate Turnover and Diversification of Modern Coleoid Cephalopods during the Mesozoic Marine Revolution. *Proc. R. Soc. B: Biol. Sci.* 284 (1850), 20162818. doi:10.1098/rspb.2016.2818

Uribe, J. E., and Zardoya, R. (2017). Revisiting the Phylogeny of Cephalopoda Using Complete Mitochondrial Genomes. *J. Molluscan Stud.* 83 (2), 133–144. doi:10.1093/mollus/eyw052

Wang, S., Zhang, J., Jiao, W., Li, J. I., Xun, X., Sun, Y., et al. (2017). Scallop Genome Provides Insights into Evolution of Bilaterian Karyotype and Development. *Nat. Ecol. Evol.* 1 (5), 1–12. doi:10.1038/s41559-017-0120

Wells, J. N., and Feschotte, C. (2020). A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* 54, 539–561. doi:10.1146/annurev-genet-040620-022145

Wong, W. Y., and Simakov, O. (2019). RepeatCraft: a Meta-Pipeline for Repetitive Element De-fragmentation and Annotation. *Bioinformatics* 35 (6), 1051–1052. doi:10.1093/bioinformatics/bty745

Young, J. Z. (1963). "The Number and Sizes of Nerve Cells in Octopus. Proceedings of the Zoological Society of London," in 140(2). Oxford, UK: Blackwell Publishing Ltd, 229–254. March. doi:10.1111/j.1469-7998.1963.tb01862.x

Zarrella, I., Herten, K., Maes, G. E., Tai, S., Yang, M., Seuntjens, E., et al. (2019). The Survey and Reference Assisted Assembly of the *Octopus vulgaris* Genome. *Scientific data* 6 (1), 1–8. doi:10.1038/s41597-019-0017-6

Zeng, X., Zhang, Y., Meng, L., Fan, G., Bai, J., Chen, J., et al. (2020). Genome Sequencing of Deep-Sea Hydrothermal Vent Snails Reveals Adaptions to Extreme Environments. *GigaScience* 9 (12). doi:10.1093/Fgigascience/Fgiaa139

Zhang, G., Fang, X., Guo, X., Li, L. I., Luo, R., Xu, F., et al. (2012). The Oyster Genome Reveals Stress Adaptation and Complexity of Shell Formation. *Nature* 490 (7418), 49–54. doi:10.5524/10003010.1038/nature11413

Zhang, Y., Mao, F., Mu, H., Huang, M., Bao, Y., Wang, L., et al. (2021). The Genome of *Nautilus Pompilius* Illuminates Eye Evolution and Biomineralization. *Nat. Ecol. Evol.*, 1–12. doi:10.1038/s41559-021-01448-6

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

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