Characterization of UV sex chromosomes and synteny-guided phylogenomic resolution of subgenomes in Bryopsida mosses

Bei Gao
Chinese Academy of Sciences https://orcid.org/0000-0002-5261-8107
Xiaoshuang Li
Xinjiang Institute of Ecology and Geography
Yuqing Liang
Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences
Jianhua Zhang
Hong Kong Baptist University https://orcid.org/0000-0002-3819-2437
Melvin Oliver (olivermj@missouri.edu)
University of Missouri https://orcid.org/0000-0002-5991-5788
Daoyuan Zhang
Chinese Academy of Sciences

Article

Keywords: genome evolution, sex, phylogenomic, Bryum argenteum, Physcomitrium patens

Posted Date: May 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2731574/v1

Abstract

The UV chromosomal sex-determination system primarily found in bryophytes, together with the XY and ZW chromosomal sex-determination systems, constitute the three principal types of chromosomal sex determination systems in nature. In this report, the genome of the moss *Bryum argenteum* was sequenced and assembled into 11 chromosomes containing 17,721 protein coding genes. A giant female U sex chromosome demonstrated conspicuously lower gene density, higher repeat coverage, and higher GC content compared to the ten autosomes. By further characterizing the sex-chromosomes and sex-linked genes in seven chromosomal-scale Bryopsida genomes, lower gene densities and distinct GC contents were revealed to be common in all moss sex chromosomes, likely resulting from degeneration. Weaker purifying selection, as evidenced by lower codon usage preference in sex-linked genes, was shared in most but not all Bryopsida mosses. Retained genomic syntenies between U/V sex chromosomes and ancestral Bryopsida chromosome 4 provided new evidence to support an autosomal origin for sex chromosomes. The seven ancestral Bryopsida chromosomes were reconstructed to decipher and depict chromosomal evolution; Dicranidae and Bryidae mosses demonstrated one and three chromosomal fusions and evolved 13 and 11 ancestral chromosomes, respectively. Ancient whole genome duplications characterized many plant lineages including the psi polyploidy event that occurred in the early diversification of Bryopsida. By aligning the modern chromosomes to the seven reconstructed ancestral chromosomes, genomic synteny-guided phylogenomic analyses provided strong evidence for the ancestral psi event shared by Dicranidae, Bryidae and *P. patens*. Furthermore, ancestral chromosomal projections and a phylogenomic resolution of Bryopsida subgenomes revealed distinct and lineage-specific chromosomal loss patterns in Dicranidae and Bryidae following the psi event. Our reconstructions reinforced the evolutionary significance and phylogenetic timing of the psi event and provided systemic insights into the sex and chromosomal evolution in mosses.

Introduction

Bryophytes are a monophyletic clade of early-diverging embryophyte remnants [1] composed of three principal lineages: mosses, liverworts, and hornworts. Of the three lineages, plants in the class Bryopsida (mosses) are the most diversified [2]. The initial sequencing of the genome of the model moss *Physcometrium patens* [3] provided extensive evolutionary insights into fundamental processes of plant growth and development and the transition of plants from aquatic to terrestrial habitats. Since the first draft, the genome sequence has been improved to a chromosome-level assembly and has revealed an unusual genome structure and an evolutionary history that included two recurrent rounds of whole genome duplications [4]. To date several moss genomes have been published including two Dicranidae genomes: *Ceratodon purpureus* [5] and *Syntrichia caninervis* [6], and six Bryidae genomes: *Pleurozium schreberi* [7], *Fontinalis antipyretica* [8], *Calohypnum plumiforme* [9], *Ceratodon purpureus* [5], *Pohlia nutans* [10], *Entodon seductrix* and *Hypnum curvifolium* [11], all of which belong to the class Bryopsida. Recently, two Sphagnum genomes, *S. divinum* and *S. angustifolium*, were released [12]. The sphagnum lineage diverged early in the radiation of mosses and has a common ancestor that is evolutionarily distant from the

Bryopsida [2]. Mosses, in addition to the occupation of an ideal phylogenetic position for reconstructing ancestral evolutionary changes that occurred in early-diverging embryophytes [13], also offer critical genetic components that enlighten our understanding of terrestrial adaption to environmental challenges such as extreme temperature variation [14], increased exposure to high light [15], and wide variations in water availability [16].

Previously, an integrated large-scale phylogenomic analyses revealed that a whole genome duplication event (the ψ event) occurred during the early diversification of Bryopsida, the first of the two whole genome duplication events reported for *P. patens*, which shaped most Bryopsida genomes. Functional analyses also indicated that recurrent large-scale gene duplications have contributed to the expansion of stress-responsive genes [17]. Besides the ancestral polyploidy events, the evolution of sex chromosomes have long been the focus of researchers and has led to the discovery of the XY, ZW and UV sex chromosomes [18]. It is generally acknowledged that the sex of the haploid gametophyte dominant bryophytes is determined by UV (U: female and V: male) sex chromosomes. The UV chromosomal sex-determination system has recently been identified and characterized in the *Sphagnum* genomes [12] and for *Ceratodon purpureus* [5], and the U- and V-linked genes reported in *C. purpureus* provided important markers to phylogenetically distinguish other Dicranidae and Bryidae U/V sex chromosomes.

Although there has been a focus on both whole genome duplications and chromosomal sex determination, questions remain as to how whole genome duplications have shaped the evolution of moss genomes and how sex chromosomes have evolved within Bryopsida lineages. The release of more genomes in Bryidae and Dicranidae, combined with the construction of the seven ancestral chromosomes [4] and the documentation of the psi WGD event, makes it possible to reconstruct and contrast the ancestral chromosomal evolutionary trajectories of these two lineages. Utilizing the reconstruction approach allows for an assessment of chromosomal level changes that may have occurred in each of the subgenomes (following the psi event) and to determine, if UV sex chromosomes are present in all sequenced Bryidae and Dicranidae genomes, whether there exist a general and common evolutionary feature that characterize and govern Bryopsida UV sex chromosomes and sex-linked genes. Integrated and systematic analyses of the increased number of moss genomes, in particular chromosome-level genomes, will generate a much clearer systematic picture of the evolutionary history of moss genomes and their components.

In this report a high-quality chromosomal scale genome of the desert moss *B. argenteum* is presented. The genome was derived from female gametophytes and as such contained a giant U chromosome with distinctive genomic characteristics that separated it from autosomes. Comparisons with published genomes indicated that U or V sex chromosomes were present in all the Bryidae and Dicranidae genomes and all had conspicuously lower gene densities and higher accumulation levels of repetitive elements, in particular the Gypsy LTR retrotransposons. The basic ancestral chromosome numbers in Bryidae and Dicranidae were revealed to be 11 and 13 which evolved from seven ancestral chromosomes following the psi polyploidy event and additional chromosomal fusions and losses. The *B. argenteum* autosomes

contained both intragenomic and whole-chromosomal level intergenomic collinearities with the other Bryopsida genomes. Based on inter- and intra-genomic synteny analyses, ancestral Bryopsida chromosomes were reconstructed to depict the way the psi polyploidy event shaped the evolution of moss genomes. Chromosomal-level subgenomes were phylogenetically resolved using a synteny-guided phylogenomic pipeline. Our integrated genomic and phylogenetic analyses provided a robust framework for Bryopsida genome evolution.

Results and Discussion

Genome assembly and annotation of Bryum argenteum

High quality genomic DNA was isolated from cultured *Bryum argenteum* gametophores and initially sequenced to generate Illumina reads for a genome assembly and genome survey analyses. Based on a k-mer (k = 17) analysis [19], the genome size was estimated at approximately 292.04 megabases (Mb), the small genome size and low heterozygosity rate of the haploid genome simplified the genome assembly. Corrected Nanopore long reads were assembled into a scaffold level genome with a total length of 258.85 Mb and N50 length of 15.94 Mb. Construction and sequencing of a Hi-C library and subsequent use of the genomic scaffolds contact information rendered 99.83% of the assembled sequences anchored on 11 chromosomes (Fig. 1 and Fig. S1) with only 92 gaps (represented by inserting 100 Ns in the genome assembly). The final genome assembly included fourteen unplaced small scaffolds containing a very small number of genes (Table S1). The quality of the genome assembly was assessed using LTR Assembly Indexes (LAI) as described previously [20]. The final *B. argenteum* genome assembly possessed an adjusted LAI of 15.29 (Fig. S2). The quality of the *B. argenteum* genome was comparable to the current reference genome for *P. patens* (LAI = 15.4) but not to the quality of the *C. purpureus* R40 genome (LAI = 21.87) which represents a gold standard level assembly (LAI > 20) [20]. With the *B. argenteum* genome assembly assessed at the high-quality reference level [20], *B. argenteum* (Bryidae), *C. purpureus* (Dicranidae) and *P. patens* (Funariidae) represent the three important lineages in Bryopsida, and thus provide robust genomic materials to study moss genomes with diverse phylogenetic positionings.

Prior to protein-coding gene annotation, repetitive sequences were annotated and masked to reveal that 51.95% of the *B. argenteum* genome was occupied by repetitive elements. Of the repeat elements that could be classified [21], the majority were LTRs (25.67%), DNA transposons (5.72%) and simple tandem repeats (4.09%) (Table 1). Non-coding RNAs were annotated within the genome, including 1,094 miRNAs, 424 tRNAs, 104 rRNAs and 431snRNAs (Table 1 and Table S2). By combining the *ab initio* predictions, RNA-Seq transcript alignments and homolog sequence-based annotations using EVidenceModeler [22], a total of 17,721 protein coding genes were annotated, among which 15,207 were considered of high confidence with supporting RNA-Seq reads evidence (Fig. S3). The completeness of the annotated protein dataset was evaluated using the BUSCO pipeline [23] employing two conserved gene datasets across viridiplantae (viridiplantae_odb10) and land plants (embryophyte_odb10). The annotated *B. argenteum*

genome captured 85.7% and 96.2% of the 1,614 viridiplantae and 425 embryophyta genes, respectively, which was comparable to the *Physcomitrium patens* genome (89.8% and 98.1%). The lower coverage of conserved viridiplantae genes in moss genomes likely resulted from the fact that few mosses are represented in the generation of the BUSCO dataset.

Table 1
Genome assembly and annotation of *Bryum argenteum*

Features	Value
Genome assembly size, Mb	258.86
No. of chromosomes	11
Genome GC content	41.85%
Repetitive elements, Mb (%)	134.49 (51.95)
Protein coding genes	17,721
Average CDS length, bp	1,317.73
No. of exons per gene	5.84
Average exon length, bp	225.69
Average intron length, bp	277.62
tRNA	424
rRNA	104
No. of transcription factor genes	604
No. of genes with SwissProt annotations (%)	12,698 (71.66)
No. of genes with Go annotations (%)	12,721 (71.78)
No. of genes with InterProScan annotations (%)	14,680 (82.84)

Diversity of Bryopsida genomes

In addition to the annotation data, alternate density distributions for genes and repeats (track III and IV in Fig. 1) and intra-genomic collinearities (track VII in Fig. 1) enabled a structural assessment of the 11 constructed chromosomes. Chromosome 1, the largest of the 11 chromosomes, was unique from the other 10 chromosomes in that it had a low density of protein coding genes and higher GC content (Fig. 1). Chromosome 1 was 44 Mb nucleotides in length and possessed only ~ 12.2 protein coding genes per Mb for a total of 539 genes, in contrast to the other ten chromosomes that encoded an average of 80 protein-coding genes per 1-Mb (Fig. 2B and Table S3). The GC content of the *B. argenteum* chromosome 1 was

approximately 43.75%, while GC contents of the other ten chromosomes did not exceed 42% (Fig. 2C and Table S3). Among the 11 sequenced Bryopsida genomes, eight were assembled to the chromosomal level (Fig. 2A). Similar to the *B. argenteum* genome, both of the hypnalean moss genomes, *H. curvifolium* and *E. seductrix*, also had 11 chromosomes. The *P. nutans* genome was characterized by a recent WGD to encompass 23 chromosomes, where two-fold syntenic chromosomal counterparts could be observed [10]. Based on the inter-genome synteny analyses of *P. nutans* and *B. argenteum*, *P. nutans* chromosomes 20 and 21 along with chromosome 19 are syntenic to chromosome 8 of *B. argenteum* (Fig. 4C). The *P. nutans* chromosomes 20 and 21 were likely derived from a chromosomal breakage after a lineage specific WGD event. Thus, the analyses suggest that the basic ancestral chromosome number for Bryidae genomes is 11.

A high-confidence species tree (all branches with 100% bootstrap support), consistent with previous phylogenic studies [17, 24], was constructed from a phylogenetic analysis of 132 single-copy homologous groups in the 11 Bryopsida mosses (Fig. 2A). *B. argenteum* and *P. nutans* were placed as sister to Hypnales within the Bryidae (Fig. 2A). Generally, genome sizes do not vary greatly among the sequenced genomes, and with the exception of *P. nutans*, none exceed 500 Mb. *B. argenteum* represented the smallest genome at 258.85 Mb assembled (Fig. 2A). Of the moss genomes that have been sequence most are composed of 40–60% repetitive elements except for the *P. schreberi* genome which contained less than 30% repetitive elements and the *P. nutans* genome with 65.53% repetitive sequences.

The sequenced moss genomes possessed a variable number of genes, ranging from less than 18,000 protein coding genes to more than 40 thousand genes. The two genomes in *P. nutans* [10] and *P. patens* [4], with the largest number of genes, were characterized by “recent” whole genome duplication events that would explain the increased gene numbers. Moss genomes also had relatively small numbers of transcription factor (TF) genes, less than 850, except for the *P. patens* and *P. nutans* genomes that had 1,136 and 1,304 TFs, respectively, in line with their larger gene repertoires. The *C. plumiforme* genome had only 685 TF genes, which was incongruent with the report of a recent WGD event characterized by a K_S peak of around 0.2 [9]. Analysis of the K_S distribution of 198 pairs of syntelogs in the *C. plumiforme* draft genome revealed a conspicuous K_S peak around 1.2–1.4 (Fig. S4), which also suggested that the “recent WGD” hypothesis for the *C. plumiforme* genome was in doubt. Specific TF family expansions were observed in the *P. patens* and *C. plumiforme* genomes, for example both *F. antipyretia* and *B. argenteum* have 9 AP2 genes, while *P. patens* and *C. plumiforme* have 17 AP2 genes in their genomes, but such instances are rare among the 53 TF families in mosses (Fig. S5). It was also possible that the recent K_S peak and large gene number might have been the result of accumulated small-scale gene duplications, which would require a chromosome-level genome assembly and comparative genomic synteny investigations for this genome.

Annotation of TFs in 2 chlorophyte, 5 charophyte, 3 hornwort and 7 representative tracheophyte genomes together with the 11 Bryopsida genomes and plotting their presence or absence (Fig. S5) provided a clear footprint of the origin and lineage-specific losses of TF genes. A number of TFs were conserved across

the viridiplantae, such as AP2, ERF, CPP, C2H2, C3H, MYB, SBP, bZIP and bHLH, coincident with the observation that a number of TF families originated in charophytes [25], the paraphyletic clade that is sister to the land plants. For example, the ARF, NAC, LFY, BBR-BPC, CAMTA, DBB, EIL, GRAS, HD-ZIP, RAV, TALE, TCP, WOX and ZF-HD TF families are all considered to have originated in the charophytes prior to the emergence of land plants. Some of the transcription factor families demonstrated relatively shallow origins, for example, the GeBP and SAP genes were absent in algae and bryophytes but emerged in vascular plants. The NZZ-SPL genes with low copy number were only found in angiosperms, lycophytes and in *P. patens*, which suggested an early origin and evolution via relatively independent processes (gains and losses) in distinct lineages. The VOZ gene was shown to be present in 9 of the moss genomes but absent in the *S. caninervis* and *P. schreberi* genomes, the *S. caninervis* genome also appeared to lack RAV, TCP and BBR-BPC genes. These TFs were well conserved across embryophytes, suggesting that their absence maybe the result of incomplete assembly or annotation of these two genomes or possible species-specific gene loss events. The Whirly transcription factors were present in all plant genomes except for the 11 Bryopsida genomes, which indicated a lineage-specific gene loss of the whirly family. One whirly TF gene was discovered in the peatmoss *Sphagnum fallax* (Sphagnopsida) genome, which suggested the whirly gene family loss occurred in the ancestor of Bryopsida, rather than in the ancestor of mosses. Whirly genes were reported to regulate metabolic processes for defense of external stresses [26] and were induced by both drought and salt stresses [27]. The absence of whirly TFs in the Bryopsida may suggest the presence of defense and regulatory mechanisms that are distinctive from other embryophytes.

Characterizations of sex chromosomes in Bryopsida

Previous reports suggested the presence of sex chromosomes in both Bryidae (*H. curvifolium* and *E. seductrix*) [11] and Dicranidae (*S. caninervis* and *C. purpureus*) [5, 6]. By analyzing the gene densities (Fig. 2B) and GC contents (Fig. 2C) for each of the chromosomes in the eight genomes (Table S3), sex chromosomes in each of the genomes conspicuously emerged (Fig. 2B and 2C). All identified sex chromosomes had lower gene densities compared to autosomes, and higher GC contents with the lone exception of chr. 11/V in *E. seductrix* (Fig. 2C). Notably, two putative sex chromosomes (chrs. 22 and 23) emerged in the *P. nutans* genomes (Fig. 2B and 2C). To determine whether the emerged sex chromosomes are U or V chromosomes, we extracted the genes on each of the sex chromosomes and searched for homologs of the 4 pairs of sex-linked genes reported in *C. purpureus* [5]. Only one pair of the marker genes, that has no known molecular function, delivered best-aligned homologs on each of the six putative sex chromosomes (Fig. 2D). We used synonymous and non-synonymous distances estimated between the *C. purpureus* sex-linked maker and corresponding homologs found on each of the putative sex chromosomes to determine their identity (Fig. 2D). Relatively lower substitutional rates for the *C. purpureus* V-linked marker genes were indicative of the male V chromosome, while lower substitutional levels for the U-linked marker were indicative of a female U chromosome (Fig. 2D). We constructed a maximum-likelihood tree for the genes in this homologous sex-linked cluster which indicated that two

clades conspicuously separated female- and male-linked genes with a tree topology similar to an ancestral gene duplication prior to U-V separation (Fig. 2E), which is consistent to the substitutional rate analyses results (Fig. 2D). Chr. 1 and chr. 13 in *B. argenteum* and *S. caninervis* were determined to be U chromosomes (Fig. 2E), which is consistent with previous analyses of the *S. caninervis* sex chromosome using a different sex-linked marker (Sc_g00229) gene [6]. *H. curvifolium* chr. 11 and *E. seductrix* chr. 11 were both determined to be V chromosomes (Fig. 2E). Intriguingly, the Antarctic moss *P. nutans* contained two sex chromosomes with chr. 22 phylogenetically determined to be V and chr. 23 as a U chromosome (Fig. 2E). In this scenario, two hypotheses might arise: it is possible that the *P. nutans* genome was contaminated or incorporated sporophyte materials that included both U and V chromosomes; or the subgenome donors for the recent WGD were from female and a male gametophytes, and an economic hermaphroditic reproductive system could have been established[28]. This later scenario could perhaps have been beneficial for survival in the Antarctic environment, and would also suggest that the sex-linked maker gene had not yet eroded in the fractionation process given that the WGD event is pretty “young” with a K_S peak around 0.11 [10]. We speculate that the latter might be the case because our analyses indicated the breakage of an ancestral chromosome that generated chromosomes 20 and 21, homologous to separate segments of the chr. 19 (Fig. 4C), which suggested that the *P. nutans* genome was progressing towards de-polyploidy. If so, our reconstructions would provide a notable genomic example of hermaphroditism that might be caused by allopolyploidy or hybridization [29], even though the majority of mosses are found to be dioecious [28]. Species information described in the Flora of North America (provided by Jonathan Shaw, (https://floranorthamerica.org/Pohlia_nutans)), describes *P. nutans* as exhibiting a “sexual condition paroicous, rarely dioicous”, which supports the observations described above.

The 44.09 Mb U sex chromosome, with no large regions of synteny to any of the autosomes, was the largest chromosome in the *B. argenteum* genome occupying about 17% of the genome. To elucidate the expansion of the *B. argenteum* chr. 1/U, we plotted more detailed coverage for different classes of repetitive elements along each of the chromosomes (Fig. 3A). We observed conspicuous peaks of Copia retrotransposons on each of the *B. argenteum* chromosomes, which may serve as centromeres similar to that observed for *P. patens* [4] and *C. purpureus* [5]. As seen for the sex chromosomes of the *C. purpureus* genome, the *B. argenteum* U chromosome possessed conspicuously high contents of Gypsy LTR retrotransposon elements and hAT DNA transposons. Mutator transposons were also abundantly accumulated in the *B. argenteum* U chromosome compared to the autosomes, while PIF-Harbinger, Tc1-Mariner and Helitron DNA transposons were conspicuously less abundant. Notably, the coverage peaks of LINE retrotransposons observed in most autosomes were not present in the U chromosome.

Like the U and V sex chromosomes characterized in *C. purpureus*, the sex-linked genes in *S. caninervis*, *B. argeteum* and *E. seductrix* genomes all demonstrated a higher number of effective codons (ENC), lower frequency of optimal codons (fop) and a lower bias of GC contents on the third positions of codons (GC3s), consistent with a decreased level of purifying selection compared to autosomal genes (Fig. 3B). The largest difference in codon usage properties between sex-linked and autosomal genes were observed

in *S. caninervis*, a desiccation-tolerant moss found in the biological soil crusts in the Gurbantunggut desert of central Asia [6, 30, 31]. *S. caninervis* U-linked genes used two more codons than autosomal genes and the ENC of *S. caninervis* autosomal genes was the lowest among the moss genomes. The fop and GC3s for *S. caninervis* genes were also the highest among the investigated moss genomes (Fig. 3B). These observations are consistent with higher levels of selection for *S. caninervis* genes compared to other analyzed mosses which suggests a connection between strong purifying selection and the extreme desert environment; this, however, would require further investigation. It is noteworthy that *B. argenteum* was also originally collected from the Gurbantunggut desert, but its genome reflected a relatively average selection pressure on genes. Weaker purifying of sex-linked genes was not common among the analyzed Bryopsida genomes, for example, the fops for *P. nutans* sex-linked genes are much higher than autosomal genes (Fig. 3B). Autosomal genes of *H. curvifolium*, used more than three codons on average than sex-linked genes, while sex-linked genes exhibited higher fop values and GC3s contents, which were all consistent with the observation that sex-linked genes in *H. curvifolium* exhibited high levels of codon bias than autosomal genes.

Genomic syntenies between the *B. argenteum* chr. 4, *C. purpureus* chrs. 5 and sex chromosomes were identified (Fig. 4A) to be syntenically clustered within the ancestral Bryopsida chromosome 4 (or element D, detailed in methods below). It has been reported that the distribution of K_S values for syntenic homologs could reflect the divergency of the corresponding chromosomal fragments, tracing the phylogenetic timing of speciation or duplication [32]. The K_S values for the detected syntelog pairs for the *B. argenteum* chromosome 4 revealed a total of 82 and 94 syntelog pairs between *C. purpureus* U and V chromosomes, respectively (Fig. 4B) that had median K_S values ~ 1.1 with no statistical difference (Mann-Whitney test, $p = 0.9935$). Syntelogs between *C. purpureus* U and V chromosomes exhibited conspicuously lower overall divergency. These results suggested a common origin of genes for the three chromosomes and that they were acquired from the ancestral Bryopsida chromosome 4. In addition, the 67 pairs of syntelogs between *B. argenteum* chromosomes 4 and 7, likely relics of the ancestral psi polyploidy event in the early diversification of Bryopsida, exhibited a much higher median K_S value of ~ 1.37, which was significantly higher than the divergency level between *B. argenteum* chr. 4 and the *C. purpureus* sex chromosomes (Mann-Whitney test, $p < 0.0001$). These observations suggested that the psi polyploidy event predated the acquisition of sex-linked genes from the ancestral chromosome 4, the homologous chromosome to *C. purpureus* chromosome 5, which was assumed to have fused to establish the current the sex chromosomes [5].

Tracking the evolution of modern genomes from reconstructed ancestral Bryopsida chromosomes

From the inter-genomic synteny alignments between *B. argenteum* and *P. nutans* that represent a relatively short evolutionary distance, the dominant collinear pattern we observed were the two-fold homologous autosomal counterparts in *P. nutans* (Fig. 4C). With longer evolutionary distance syntenic

tracing back to the early diversification of Bryopsida, the genomic synteny alignments between *B. argenteum* and *P. patens* revealed a dominant 2:4 syntenic ratio and confidently retrieved the seven ancestral Bryopsida chromosomes described previously [4] (Fig. 4D). For example, the *B. argenteum* chromosomes 8 and 10 together with *P. patens* chromosomes 8, 20, 23 and 24 are modern chromosomal descendants of the ancestral chromosome 7 (i.e. ancestral element F as described in *C. purpureus*) [4, 5]. In both whole genome synteny alignments (Fig. 4C and 4D), the sex chromosomes demonstrated very little synteny with other chromosomes and so we only included the autosomes for each genome for subsequent tracking of the trajectory of their evolution from the newly reconstructed ancestral chromosomes (see Methods) and the karyotype estimation analyses.

Unlike the *P. nutans* and *P. patens* genomes that include an extra round of WGD, as revealed by whole-genome synteny comparisons with *B. argenteum* (Fig. 4C and 4D), the other Bryopsida genomes including *C. purpureus*, *S. caninervis*, *E. seductrix*, *H. cuvifolium* and *B. argenteum* had generally similar syntenic depth ratios (Fig. 5A). This is featured more prominently in the inter-genomic comparisons for the two hypnalean genomes of *E. seductrix* and *H. cuvifolium* which demonstrated more intense one-to-one chromosome-level syntenies with each other (Fig. 5A), presumably because of their closer phylogenetic relationship and short divergence time [11].

To track the evolution of extant sequenced moss chromosomes from the ancestral Bryopsida chromosomes more clearly it was necessary to update and refine the construction of the ancestral chromosomes described previously[4]. To accomplish this, we selected seven *P. patens* chromosomes to represent the ancestral chromosomes as the backbone and aligned them to the other 20 *P. patens* chromosomes along with the autosomes from other six Bryopsida species (*C. purpureus*, *S. caninervis*, *E. seductrix*, *H. cuvifolium*, *P. nutans* and *B. argenteum*) and included the insertion of additional syntenic genes into the backbone chromosomes (detailed in methods). Once the ancestral chromosomes were constructed, schematic karyotyping for two Bryidae genomes in *B. argenteum* (Fig. 5B and 5D) and *H. cuvifolium* (Fig. 5C and 5E) and two Dicranidae genomes in *S. caninervis* (Fig. 5F and 5H) and *C. purpureus* (Fig. 5G and 5I) were constructed by aligning the genomes to the reconstructed ancestral chromosomes. The two Bryidae genomes of *B. argenteum* and *H. curvifolium* had similar karyotypic patterns and indicated the presence of three ancestral chromosomal fusions, including fusions between ancestral chromosomes 3–5, 2–6 and 2–4 (Fig. 5D and 5E). The two Dicranidae genomes of *S. caninervis* and *C. purpureus* also exhibited similar karyotypic patterns but with only one shared ancestral chromosomal fusion event that occurred between the ancestral chromosomes 2–7 (Fig. 5H and 5I), resulting in the 13 modern chromosomes found in both genomes. The *B. argenteum* chromosome 11 was syntenically aligned to only a fraction of the *H. cuvifolium* chromosome 4 (Fig. 5A) and ancestral chromosome 1 (Fig. 5B), suggesting a lineage-specific segmental loss in *B. argenteum*.

Synteny-guided phylogenomic resolution of Bryopsida subgenomes

For subgenome analyses, the Bryopsida genomes were aligned to the seven reconstructed ancestral chromosomes (Fig. 6A) and a table of syntelogs were generated to facilitate syntelog clustering (Table S4). Maximum-likelihood gene trees were constructed for each of the syntelog clusters and used to estimate a coalescence-based consensus gene tree that represented the two subgenomes derived from each of the seven ancestral chromosomes (Fig. 6B). Two hypotheses exist in the literature to explain the wide-spread WGDs observed in Bryopsida genomes [4]; one assumed that each multiple WGD occurred in each of the lineages (as depicted in Fig. 6C), while the another found phylogenomic evidence for an ancestral WGD event, termed B3 [33] or psi [17] that occurred in the ancestor of BDTF (Bryidae, Dicranidae, Timmiidae and Funariidae) clade (Fig. 6D). Only two species-poor early-diverging subclasses, Diphysciidae and Buxbaumiidae, were unaffected by the psi event which supports the early diversification of the Bryiosida class.

The phylogenic topologies for all the ancestral chromosomes presented here strongly support the psi event hypotheses as illustrated in Fig. 6D. Each of the subgenome pairs generated from the psi event traced to the ancestor of the Bryidae, Dicranidae and *P. patens* with high posterior possibility values. There was only one exception in the phylogenetic tree topology where the four modern chromosomes evolved from ancestral chromosome 7, the *P. patens* chromosomes 8, 20, 23 and 24, and were placed outside of the Bryidae and Dicranidae chromosomes, however, the posterior probability is very low (0.01), suggesting the instability of this tree topology. In short, our integrated phylogenomic reconstructions rejected the multiple lineage specific WGD hypotheses as depicted in Fig. C and reinforced the phylogenetic timing of the psi event (Fig. 6F).

Lineage-specific chromosomal-level losses could also be observed from Bryopsida subgenome analyses. The Dicranidae were absent from the subgenome 4G clade and were assumed to be homologous to the modern components of subgenome 4H (*C. purpureus* chr. 5 and *S. caninervis* chr. 8). The “lost Dicranidae subgenome 4G” genes were likely acquired by the sex chromosomes after the psi event which is consistent with our observed syntenic relationships between *B. argenteum* chr. 4 (subgenome 4G) and *C. purpureus* sex chromosomes as described earlier. The data indicates that genes from the subgenome 4G have fused to and can be partially traced to the modern *C. purpureus* and *S. caninervis* sex chromosomes, which is consistent with previous phylogenomic evidence [5].

While Dicranidae genomes retained chromosomes in both the 5I and 5J subgenome clades, Bryidae was absent in the 5J subgenome clade, suggesting an ancestral lineage-specific chromosome loss, as the chromosomal loss from the *P. patens* genome was a counterpart of the 5I subgenome. Thus *P. patens* and the Bryidae have independently lost different subgenomes of the ancestral chromosome 5 following the psi event. The distinct retention and fractionation patterns of ancestral chromosome 5 among the Bryidae, Dicranidae and *P. patens* would thus represent another important signature of genomic evolution that differentiates the three Bryopsida lineages.

The modern chromosomes derived from ancestral chromosome 2 found in *S. caninervis* (chrs. 1 and 11) and *C. purpureus* (chrs. 3 and 9) provides evidence for a conspicuous chromosome break as the modern

chromosomes aligned to different but consecutive parts of the ancestral chromosome 2 and could together cover the entire ancestral chromosome 2 syntenically (Fig. 5F and 5G). There are different scenarios which could explain this observation. In one scenario, when the ancestral chromosome 2 duplicated during the psi event, one chromosome was lost and the other broke into two. In this scenario chromosomes 1 and 11 of *S. caninervis* would belong to the same subgenome clade as would chromosomes 3 and 9 of *C. purpureus*. In another scenario, the ancestral chromosome 2 duplicated into 2C and 2D, subsequently 2C and 2D recombined but lost different segments of the chromosomes. In this scenario the modern chromosomes 1 and 11 of *S. caninervis* would have evolved from two different subgenomes, as would the *C. purpureus* chromosomes 3 and 9. The synteny-guided phylogenomic analyses of the Bryopsida subgenomes placed *S. caninervis* chr. 11 together with *C. purpureus* chr. 3 in the 2C subgenome clade (Fig. 6E), and the *S. caninervis* chr. 1 and *C. purpureus* chr. 9 were placed into the 2D subgenome clade (Fig. 6E). These results suggest that the latter scenario better explained the evolutionary trajectory of these chromosomes (Fig. 6G). As for the reciprocal chromosomal segmental losses found in the Dicranidae genomes, similar segmental losses in 2D subgenome were also be observed in the Bryidae genomes (Fig. 5B and 5C), but in this case the Bryidae 2C subgenomes remained relative intact (Fig. 5B and 5C) and thus represent another novel karyotypic difference between the Dicranidae and Bryidae genomes (Fig. 6G and 6H).

Conclusions

Though non-recombining sex-linked genes exhibited weaker purifying selection on codon usage bias for most Bryopsida moss genomes, our integrated analyses suggested the existence of diversified codon usage selection trends for sex-linked and autosomal genes. The detection of synteny between the ancestral Bryopsida chromosome 4 and the sex chromosomes, in conjunction with corresponding ancestral losses of subgenomes 2 and 4G, provided strong evidence that sex chromosomes acquired genes from ancestral chromosomes 2 and 4G. Our reconstructed evolution of modern chromosomes from ancestral chromosomes and the tailored synteny-guided phylogenomic analyses of Bryopsida subgenomes generated robust evidence for the critical phylogenetic timing of the psi event during the early diversification of Bryopsida. These analyses also revealed diversified chromosomal evolutionary trajectories within the Bryidae, Dicranidae and *P. patens* genomes revolving around ancestral chromosomes 2, 4 and 5. Our reconstructions generated a systematic depiction of the evolution of Bryopsida genomes (Fig. 7) and a robust evolutionary framework for characterizing future moss genomes.

Materials and Methods

Plant material preparation and genome sequencing. Moss gametophytes were originally collected from the biological soil crusts of the Gurbantungut desert. The *Bryum argenteum* gametophytes were sub-cultured to obtain abundant plant materials and harvested as described previously for DNA extraction and genome sequencing [16, 34]. The *B. argenteum* genome was sequenced using a hybrid sequencing

strategy incorporating both Illumina and Nanopore reads. A total of 55.5 Gb of Illumina reads were initially generated to conduct K-mer analyses to estimate genome size. Based on the estimated genome size, 100x Illumina 150 bp pair-end and Nanopore long reads were generated, assembled, and polished using the NextDenovo v2.0 pipeline (https://github.com/Nextomics/NextDenovo). In addition, Hi-C (high-throughput chromatin conformation capture) libraries were constructed [35], including chromatin cross-linking and digestion, followed by DNA ligation, purification, fragmentation and sequenced on an Illumina HiSeq X platform. A total of 11.3 million 150 bp pair-end reads were generated and were then cleaned and aligned to the assembled contigs to order and anchor the assembled contigs by using the HiCUP and ALLHiC softwares [36, 37] to obtain the chromosome-scale genome assembly for subsequent analyses.

Genome annotation. A *de novo* repeat library was generated using the RepeatModeler pipeline (http://www.repeatmasker.org/RepeatModeler.html) with default parameters [38]. All repeat sequences with lengths long than 100 bp and gaps less than 5% were retained and constituted the *de novo* transposable element (TE) library. The *de novo* library and Repbase [39] were combined and processed to remove redundant sequences to generate a custom library, which was supplied to RepeatMasker (http://repeatmasker.org/) for whole genome repeat identification.

Consensus protein-coding gene models were annotated by combining ab initio, homology-based and RNA-Seq assisted predictions. The *ab initio* gene prediction pipeline consisted of Augustus v3.2.3 [40], geneid v1.4.4 [41], GENASCAN v1.0 [42], GlimmerHMM v3.04 [43] and SNAP v2013-11-29 [44]. For homolog-based gene predictions, the proteomes of *Physcomitrium patens*, *Marchantia polymorpha*, *Anthoceros angustus*, *Spirogloea muscicola*, *Arabidopsis thaliana*, *Oryza sativa* were aligned to the *B. argenteum* genome using GeneWise v2.4.1 [45] to predict the gene structures contained in each protein. A non-redundant *de novo* transcriptome using Trinity v2.1.1 [46] was generated using different moss tissues and gametophytes at different hydration stages as described by our laboratory[47]. Transcripts were aligned to the genome using the PASA pipeline (Program to Assemble Spliced Alignment) to identify exons and splicing positions [48]. Finally, the consensus gene structures were predicted by merging gene structures predicted by the three methods with EVidenceModeler v1.1.1 [22].

Gene functional annotations were assigned by aligning the *B. argenteum* protein sequences to the Swiss-Prot database using diamond [49]. The motifs and domains were annotated using InterProScan v5.31-70 [50]. The Gene Ontology (GO) annotations were conducted using PANNZER2 [51]. The final gene set was aligned to KEGG pathways and the best match for each gene was identified. The tRNAs were predicted using tRNAscan-SE [52] and highly conserved rRNAs were annotated using Barrnap v0.9 (https://github.com/tseemann/barrnap). Other ncRNAs, including miRNAs, snRNAs were identified by searching against the Rfam database with default parameters using infernal software (http://infernal.janelia.org/). Transcription factors were identified and classified using PlantTFDB v5.0 classification rules [53].

Phylogenetic analyses. Protein sequences from the 11 moss genomes were clustered using orthofinder v2.5.4 [54] and a total of 132 single copy gene families were identified. Amino acid sequences for each gene were aligned using muscle v5.1 (https://drive5.com/muscle5/) with default parameters of the "-align" algorithm [55]. Aligned sequences were back-translated and trimmed using trimal v1.4.rev15 [56] with the "-ignorestopcodon -backtrans -automated1" options. Maximum likelihood trees were constructed for each of the resulting sequence alignments using iqtree v2.2.0.3 [57] with the best-fit model [58] and the 1000 ultrafast bootstrap tests (-B 1000) [59]. Identifiers for each gene were then replaced by species names in the resultant tree files and imported into ASTRAL v5.7.1 for coalescence-based species tree inference with posterior probabilities [60].

Sex chromosome analyses. Sex chromosomes were identified by comparing gene densities and overall GC contents between sex chromosomes and autosomes, properties that differentiate the two types of chromosomes. Two sex chromosomes in the *Pohlia nutans* genome assembly and one sex chromosome in each of the Bryidae and Dicranidae genomes were probed. Among the four pairs of sex-linked marker genes, previously identified in *C. purpureus* [5], we were able to find sex chromosome-linked homologs for only one pair (CepurGG1.UG068700.1 and CepurR40.VG026200.1) on all the moss sex chromosomes. Synonymous (K_S) and nonsynonymous (K_A) distances were estimated using Nei-Gojobori approach [61] implemented in KaKs_calculator v2.0 [62] between identified sex-linked genes and U- and V-linked markers from *C. purpureus*; this homologous cluster was also used to construct a maximum likelihood gene tree using iqtree v2.2.0.3 [57] to further distinguish U and V chromosomes. The gene tree was visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/Figtree/).

Codon usage analyses. The CodonW v1.4.2 was employed to analyze the codon usage of sex-linked and autosomal genes; only coding sequences of genes with 200 or more codons were retained for further analyses. Correspondence analyses were performed for each of the gene sets to estimate the optimal codon sets in each species. Subsequently, genes annotated on each chromosome were separated into two different sets (sex chromosomes and autosomes), and the frequency of optimal codon usage (fop) [63], effective number of codons (ENC) [64], and GC contents of the third codon positions were assessed for U/V-linked and autosomal genes. The analyses were performed by the recommended procedures described by https://codonw.sourceforge.net/Tutorial.html. We conducted nonparametric tests used in the 'wilcox.test' function for pairwise comparisons and plots were generated using 'vioplot' in R v4.2.2 statistic environment.

ABK construction and karyotype projection. Based on previous genomic synteny analyses in *P. patens* and *C. purpureus* genomes, seven ancestral moss chromosomes were identified. *P. patens* chromosomes 14, 11, 3, 16, 9, 18 and 20 were chosen as the reference backbone for ancestral chromosomes 1 through 7 (equivalent to ancestral elements A through G in *C. purpureus*) [4, 5]. Synteny alignments were performed using WGDI [65] to identify collinear genes in the 7 selected reference chromosomes and the autosomes from the selected and published moss genomes (*Bryum*, *Syntrichia*, *Ceratodon*, *Hypnum*, *Entodon* and *Pohlia*) in addition to the remaining 20 *P. patens* chromosomes. Each backbone chromosome was

augmented by inserting collinear genes into the reference chromosomes using WGDI "-akr" as described previously [65, 66]. These reconstructed seven ancestral moss chromosomes were deposited in the FigShare repository. We aligned each of the Dicranidae and Bryidae genomes to the seven ancestral chromosomes to scrutinize the karyotypes for each focal genome using mcscan [67] implemented in jcvi v1.2.7 (https://github.com/tanghaibao/jcvi).

Synteny-guided phylogenomics of Bryopsida subgenomes. All autosomes of the Bryopsida genomes were used in the synteny guided phylogenomic analyses, to determine the divergence of subgenomes [68]. Inter-genomic synteny alignments were conducted between the seven reconstructed ancestral chromosomes and moss genomes using jcvi v1.2.7 to identify syntelogs with syntenic ratios of 2 in *Bryum*, *Syntrichia*, *Ceratodon*, *Hypnum*, *Entodon* and of 4 in *Physcomitrium and Pohlia* genomes. Syntelog clusters were joined with 'ancestral' moss genes in a table that constituted the seven ancestral chromosomes in the first column with 2 or 4 columns reserved for each of the moss genomes (Table S3). The first column of the table were genes constituting the reconstructed ancestral chromosomes followed by columns of syntelogs in each of the Bryopsida genomes identified using mcscan[67] with different quotas[69]. Then genes that fell into each row constituted syntelog clusters, and the clustered syntelogs were homologous to each other at the sequence level, but also carry evolutionary signatures with retained and conserved collinear gene orders. Syntenlog clusters containing more than ten genes were retained for further phylogenomic analyses. An additional column was added to the table to include homologous genes identified using blastp (-evalue 1e-5) from either *Diphyscium foliosum* or *Buxbaumia aphylla* generated by OneKP [33], as outgroups to properly root the trees. Maximum likelihood trees were generated for each of the syntelog clusters as described above. The resultant gene trees for syntelog clusters were then divided into seven pools based on the corresponding 'ancestral' genes that belonged to the seven ancestral Bryopsida chromosomes. Gene identifiers were then replaced by their corresponding "species_chr" to generate compatible tree files which were imported into ASTRAL v5.7.1 for the coalescence-based phylogenetic analyses [60] of sub-genomes.

Declarations

Data availability. The *Bryum argenteum* genome has been deposited in the National Genomics Data Center (NGDC; https://ngdc.cncb.ac.cn/bioproject/) under the BioProject accession number PRJCA015400. The genome assembly, annotation and reconstructed ancestral Bryopsida chromosomes supporting the key findings of this study are publicly available in figshare repository (https://doi.org/10.6084/m9.figshare.21226328.v6).

Author Contributions: B.G. and D.Z. designed and managed the project. B.G. performed the bioinformatic analyses and wrote the manuscript. Y.L., X.L. and R.R.Y. collected and prepared samples. J.Z. and M.J.O. contributed to the experimental design of the project. J.Z., M.J.O. and D.Z. revised the manuscript. All authors read and approved the manuscript.

Competing Interest Statement: The authors declare no competing interests.

Acknowledgments: This research was supported by the National Natural Science Foundation of China (NSFC Grant No.: 32100256 to BG), the Third Xinjiang Scientific Expedition Program (2021xjkk0501), the financial support provided by the Ministry of Human Resources and Social Security of the People's Republic of China (2022000005 and 2022000243 to BG) and the U.S. National Science Foundation Dimensions of Biodiversity Program Award (1638972 to MJO).

References

1. Puttick MN*, et al.* The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte. *Current biology : CB* **28**, 733-745 e732 (2018).
2. Shaw AJ, Szovenyi P, Shaw B. Bryophyte Diversity and Evolution: Windows into the Early Evolution of Land Plants. *American journal of botany* **98**, 352-369 (2011).
3. Rensing SA*, et al.* The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64-69 (2008).
4. Lang D*, et al.* The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. *The Plant journal : for cell and molecular biology* **93**, 515-533 (2018).
5. Carey SB*, et al.* Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. *Science Advances* **7**, eabh2488 (2021).
6. Silva AT*, et al.* To dry perchance to live: Insights from the genome of the desiccation-tolerant biocrust moss Syntrichia caninervis. *The Plant journal : for cell and molecular biology* **105**, 1339-1356 (2021).
7. Pederson ERA, Warshan D, Rasmussen U. Genome Sequencing of Pleurozium schreberi: The Assembled and Annotated Draft Genome of a Pleurocarpous Feather Moss. *G3 Genes|Genomes|Genetics* **9**, 2791-2797 (2019).
8. Yu J*, et al.* Draft genome of the aquatic moss Fontinalis antipyretica (Fontinalaceae, Bryophyta). *Gigabyte* **2020**, 1-9 (2020).
9. Mao L*, et al.* Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 12472-12480 (2020).
10. Liu S*, et al.* The Antarctic Moss Pohlia nutans Genome Provides Insights Into the Evolution of Bryophytes and the Adaptation to Extreme Terrestrial Habitats. *Frontiers in plant science* **13**, (2022).
11. Yu J*, et al.* Chromosome-Level Genome Assemblies of Two Hypnales (Mosses) Reveal High Intergeneric Synteny. *Genome Biology and Evolution* **14**, evac020 (2022).
12. Healey AL*, et al.* Newly identified sex chromosomes in the Sphagnum (peat moss) genome alter carbon sequestration and ecosystem dynamics. *Nature Plants*, (2023).
13. Bowman JL*, et al.* Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell* **171**, 287-304 e215 (2017).
14. Zhuo L*, et al.* Thermal tolerance of dried shoots of the moss Bryum argenteum. *Journal of Thermal Biology* **89**, 102469 (2020).

15. Schroeter B, Green TGA, Kulle D, Pannewitz S, Schlensog M, Sancho LG. The moss Bryum argenteum var. muticum Brid. is well adapted to cope with high light in continental Antarctica. *Antarctic Science* **24**, 281-291 (2012).
16. Liang Y*, et al.* Dehydration rates impact physiological, biochemical and molecular responses in desert moss Bryum argenteum. *Environmental and Experimental Botany* **183**, 104346 (2021).
17. Gao B*, et al.* Ancestral gene duplications in mosses characterized by integrated phylogenomic analyses. *Journal of Systematics and Evolution* **60**, 144-159 (2022).
18. Bachtrog D*, et al.* Are all sex chromosomes created equal? *Trends in Genetics* **27**, 350-357 (2011).
19. Liu B*, et al.* Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv: Genomics*, (2013).
20. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Research* **46**, e126-e126 (2018).
21. Wicker T*, et al.* A unified classification system for eukaryotic transposable elements. *Nature reviews Genetics* **8**, 973-982 (2007).
22. Haas BJ*, et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology* **9**, R7 (2008).
23. Waterhouse RM*, et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* **35**, 543-548 (2017).
24. Liu Y*, et al.* Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nature Communications* **10**, (2019).
25. Wilhelmsson PKI, Muhlich C, Ullrich KK, Rensing SA. Comprehensive Genome-Wide Classification Reveals That Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae. *Genome Biology and Evolution* **9**, 3384-3397 (2017).
26. Desveaux D, Maréchal A, Brisson N. Whirly transcription factors: defense gene regulation and beyond. *Trends in plant science* **10**, 95-102 (2005).
27. Akbudak MA, Filiz E. Whirly (Why) transcription factors in tomato (Solanum lycopersicum L.): genome-wide identification and transcriptional profiling under drought and salt stresses. *Molecular biology reports* **46**, 4139-4150 (2019).
28. McDaniel SF, Atwood J, Burleigh JG. RECURRENT EVOLUTION OF DIOECY IN BRYOPHYTES. *Evolution* **67**, 567-572 (2013).
29. Robertson K, Goldberg EE, Igić B. Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae. *Evolution* **65**, 139-155 (2011).
30. Zhao J, Zheng Y, Zhang B, Chen Y, Zhang Y. Progress in the study of algae and mosses in biological soil crusts. *Frontiers of Biology in China* **4**, 143-150 (2009).
31. Belnap J. The world at your feet: desert biological soil crusts. *Frontiers in Ecology and the Environment* **1**, 181-189 (2003).

32. Bergero R, Charlesworth D. The evolution of restricted recombination in sex chromosomes. *Trends in Ecology & Evolution* **24**, 94-102 (2009).
33. Leebens-Mack JH*, et al.* One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679-685 (2019).
34. Gao B, Zhang D, Li X, Yang H, Zhang Y, Wood AJ. De novo transcriptome characterization and gene expression profiling of the desiccation tolerant moss Bryum argenteum following rehydration. *BMC Genomics* **16**, 416 (2015).
35. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268-276 (2012).
36. Wingett S*, et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* **4**, 1310 (2015).
37. Zhang X, Zhang S, Zhao Q, Ming R, Tang H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat Plants* **5**, 833-845 (2019).
38. Flynn JM*, et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451 (2020).
39. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**, 11 (2015).
40. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**, ii215-225 (2003).
41. Guigó R, Knudsen S, Drake N, Smith T. Prediction of gene structure. *Journal of molecular biology* **226**, 141-157 (1992).
42. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* **268**, 78-94 (1997).
43. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878-2879 (2004).
44. Korf I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
45. Birney E, Durbin R. Using GeneWise in the Drosophila annotation experiment. *Genome research* **10**, 547-548 (2000).
46. Haas BJ*, et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**, 1494-1512 (2013).
47. Gao B*, et al.* Desiccation tolerance in bryophytes: The dehydration and rehydration transcriptomes in the desiccation-tolerant bryophyte Bryum argenteum. *Scientific reports* **7**, 7571 (2017).
48. Haas BJ*, et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654-5666 (2003).
49. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **12**, 59-60 (2015).

50. Jones P*, et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
51. Törönen P, Medlar A, Holm L. PANNZER2: a rapid functional annotation web server. *Nucleic Acids Research* **46**, W84-W88 (2018).
52. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **44**, W54-57 (2016).
53. Tian F, Yang DC, Meng YQ, Jin J, Gao G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res*, (2019).
54. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* **20**, 238 (2019).
55. Edgar RC. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *bioRxiv*, 2021.2006.2020.449169 (2021).
56. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
57. Minh BQ*, et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530-1534 (2020).
58. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* **14**, 587-+ (2017).
59. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518-522 (2018).
60. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
61. Nei M, Gojobori T. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Molecular Biology and Evolution* **3**, 418-426 (1986).
62. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, proteomics & bioinformatics* **8**, 77-80 (2010).
63. Stenico M, Lloyd AT, Sharp PM. Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases. *Nucleic Acids Res* **22**, 2437-2446 (1994).
64. Wright F. The 'effective number of codons' used in a gene. *Gene* **87**, 23-29 (1990).
65. Sun P*, et al.* WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Molecular Plant* **15**, 1841-1851 (2022).
66. Wang Z*, et al.* A high-quality Buxus austro-yunnanensis (Buxales) genome provides new insights into karyotype evolution in early eudicots. *BMC Biology* **20**, 216 (2022).
67. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. *Science* **320**, 486-488 (2008).
68. Chanderbali AS*, et al.* Buxus and Tetracentron genomes help resolve eudicot genome history. *Nature Communications* **13**, 643 (2022).

69. Tang H, Lyons E, Pedersen B, Schnable JC, Paterson AH, Freeling M. Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinformatics* **12**, 102 (2011).

Supplementary Information

Supplementary Figures and Supplementary Table S1-S2 are not available with this version.

Figures

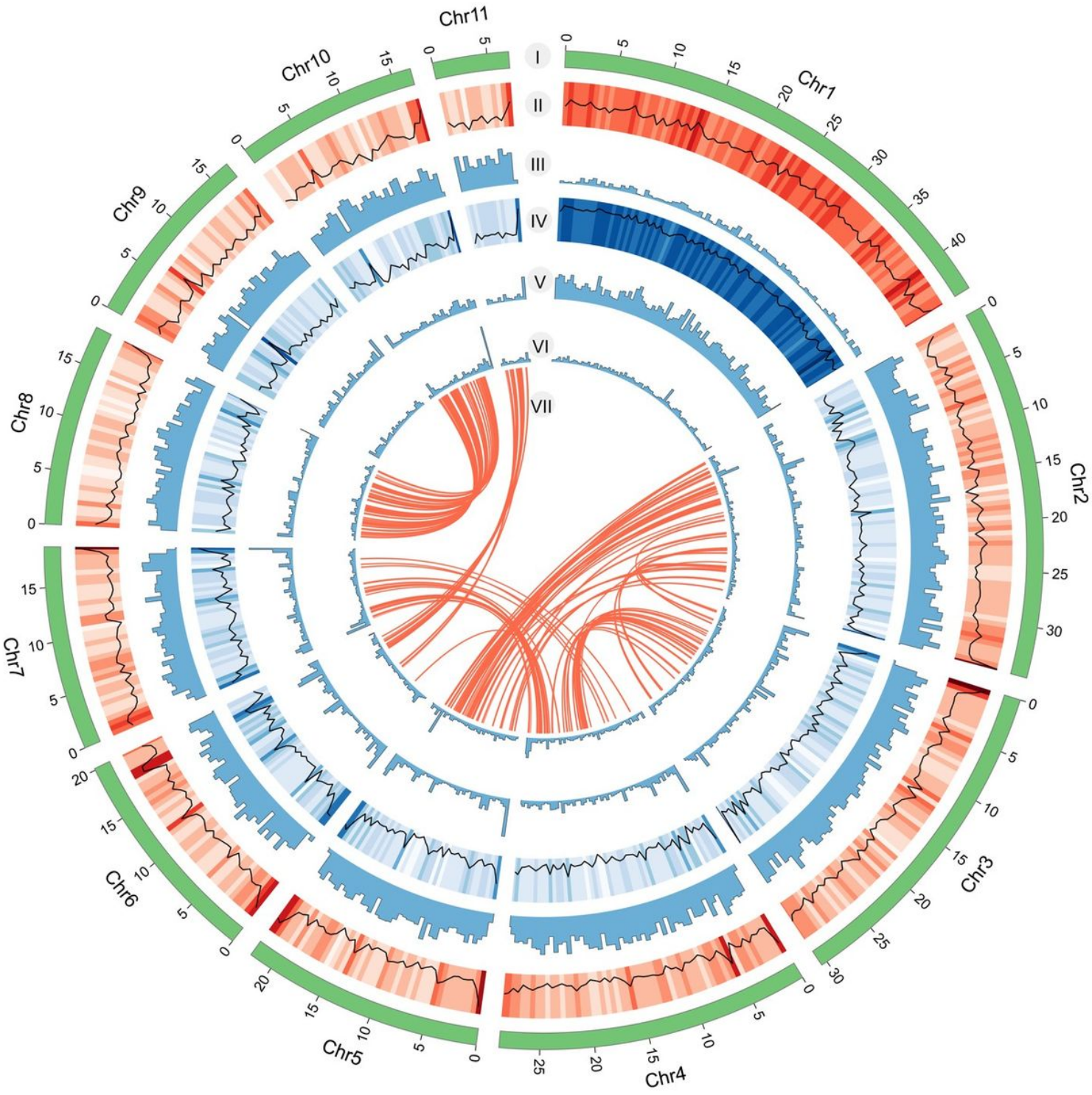
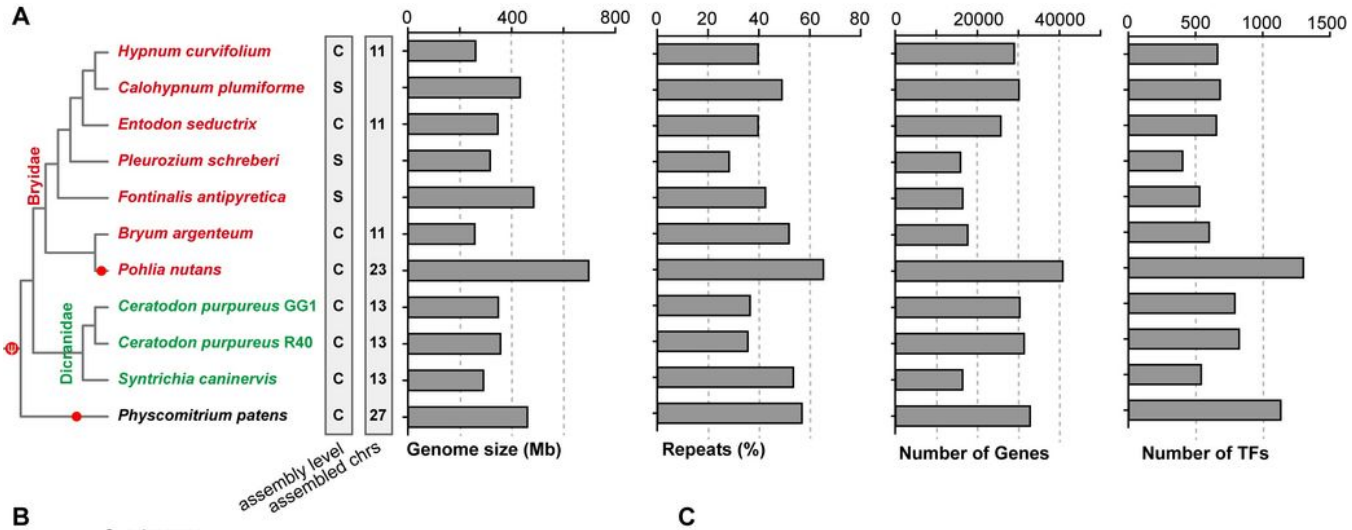
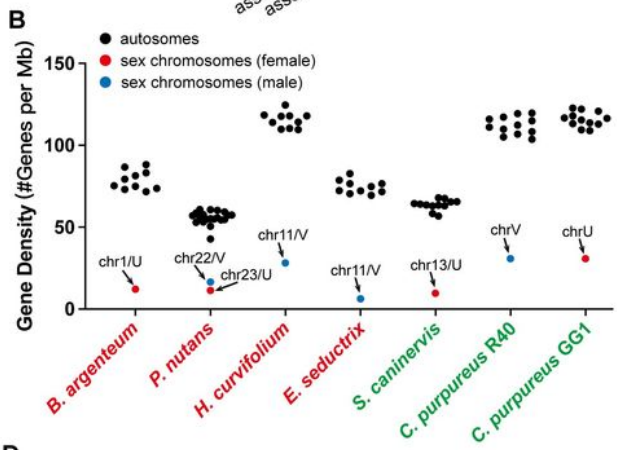
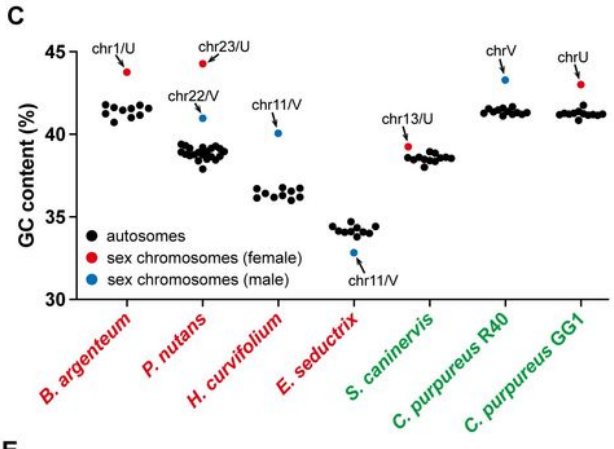


Figure 1

A circos illustration depicting the genomic features of *Bryum argenteum*. The seven tracks depicted are: (I) the eleven chromosome-level assemblies of the *B. argenteum* genome, (II) GC content of 400-kb sliding windows along the chromosomes, (III) gene density in the 400 kb-sliding windows, (IV) repeat coverage in the 400 kb-sliding windows, (V) Gypsy LTR retrotransposon element coverage and (VI) Copia LTR element coverage in the 400 kb-sliding windows, (VII) genomic synteny connections within the whole genome.







D

Homologues on sex chrs	Sex-linked UV genes*	K_A	K_S
E. seductrix chr11 (jg23771.t1)	U-linked	0.376	1.337
	V-linked	**0.306**	**0.780**
H. curvifolium chr11 (jg26576.t1)	U-linked	0.390	1.201
	V-linked	**0.323**	**0.800**
P. nutans chr22 (Poh0072050.1)	U-linked	0.375	0.860
	V-linked	**0.275**	**0.798**
P. nutans chr23 (Poh0129450.1)	**U-linked**	**0.258**	**0.667**
	V-linked	0.384	0.960
B. argenteum chr1 (BA01G04160)	**U-linked**	**0.320**	**0.853**
	V-linked	0.428	1.268
S. caninervis chr13 (Sc_g00011)	**U-linked**	**0.173**	**0.481**
	V-linked	0.400	1.048

*U-linked: CepurGG1.UG068700; V-linked: CepurR40.VG026200

E

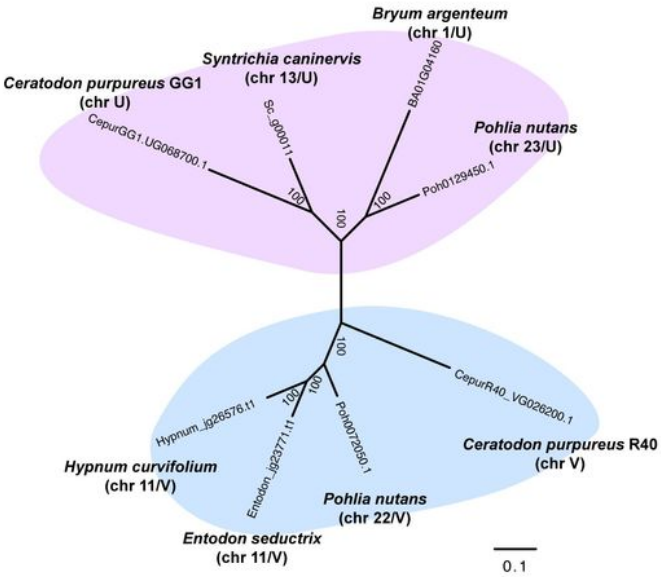


Figure 2

Phylogenetic relationships of sequenced mosses, genomic comparisons and UV-linked marker genes on moss sex chromosomes. (A) Phylogenetic relationships for 11 Bryopsida mosses resolved by 132 single-copy homologous clusters (all branches with 100% support) were consistent with the conventional moss taxonomy. Moss species of Dicranidae and Bryidae were labelled as red and green fonts, respectively. Genomes assembled into scaffold and chromosomal levels were labelled as “S” and “C”, respectively. Number of chromosomes, genome sizes, percentage of repetitive sequences, number of genes and transcription factors were compared. **(B)** Gene densities on each assembled chromosomes of the sequenced Bryidae and Dicranidae genomes, putative sex chromosomes were highlighted and labelled. **(C)** GC contents of each assembled chromosomes of the sequenced Bryidae and Dicranidae genomes**,** putative sex chromosomes were highlighted and labelled. **(D)** Non-synonymous and synonymous rates between the conserved *Ceratodon purpureus* female (U-linked) and male (V-linked) marker genes and the top-hit homologous genes found on each of the putative moss sex chromosomes. Lower K_A and K_S values suggested higher similarity to the U or V linked marker genes and were highlighted in bold fonts. **(E)** A maximum-likelihood gene tree constructed using the sex chromosome maker genes. Moss species and corresponding sex chromosomes were labelled beside the maker genes. The male (light blue) and female (light purple) clades were conspicuously identified and are constituted of V and U chromosome linked maker genes, respectively.

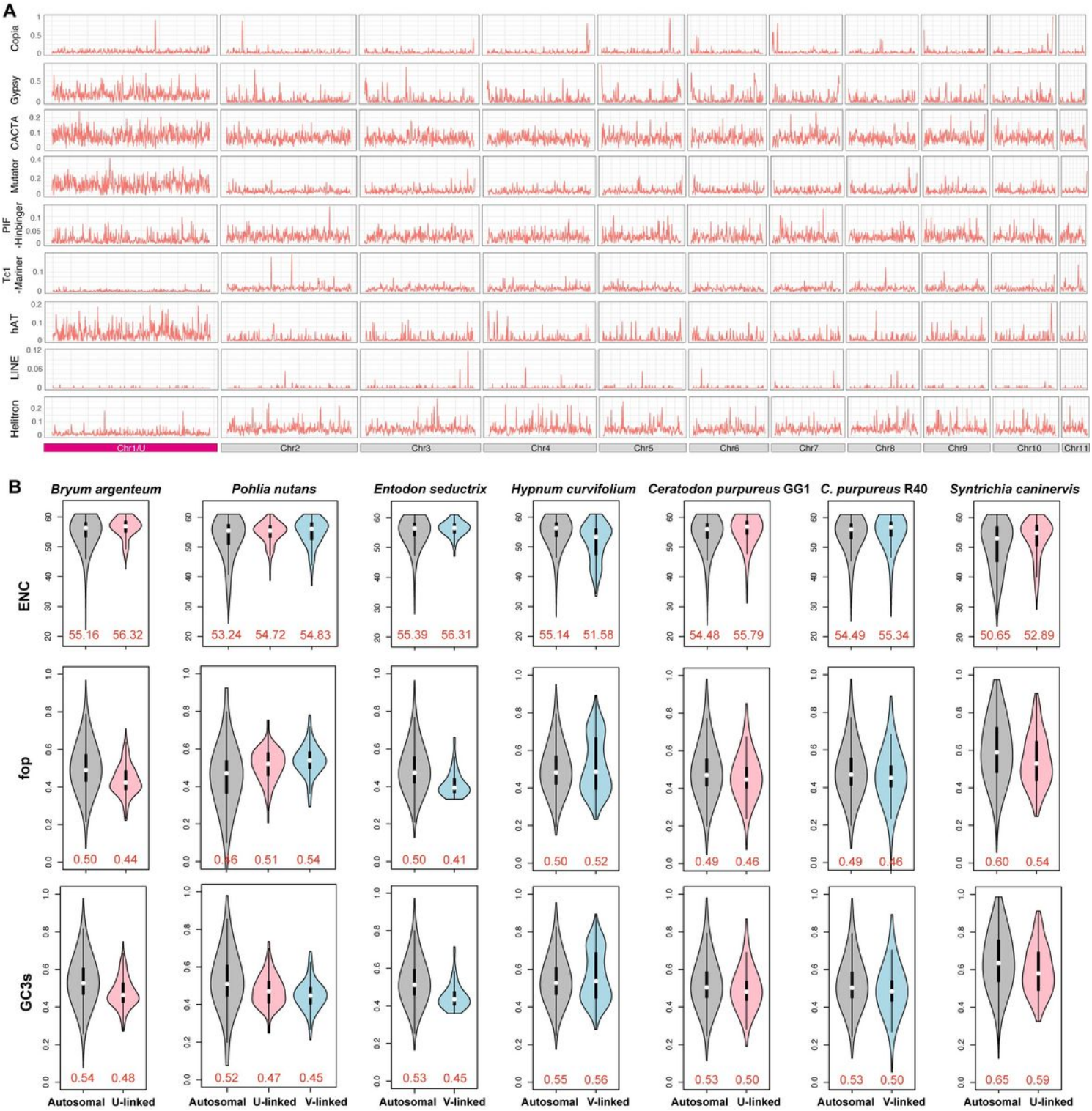


Figure 3

Densities of repetitive elements across *B. argenteum* chromosomes and distinct evolutionary signatures of autosomal and sex-linked genes. (A) Density plots of distinct classes of repetitive elements across *B. argenteum* chromosomes, showing the proportion of 100-kb sliding window with 90-kb jump of each feature. The density peaks of copia LTR elements on each chromosome represented candidate centromeric regions, similar to *P. patens* and *C. purpureus*. The *B. argenteum* sex chromosome U accumulate higher densities of Gypsy, CACTA, Mutator and hAT elements than autosomes, similar to that

in *C. purpureus*; but lower densities of Tc1-Mariner and Helitron elements. (B) Codon usage properties in terms of effective number of codons (ENC), frequency of optimal codons (fop) and GC contents on the 3rd positions of codon (GC3s) were compared between autosomal and sex-linked genes in the seven Bryopsida genomes. Numbers below each violin plots show the corresponding mean values.

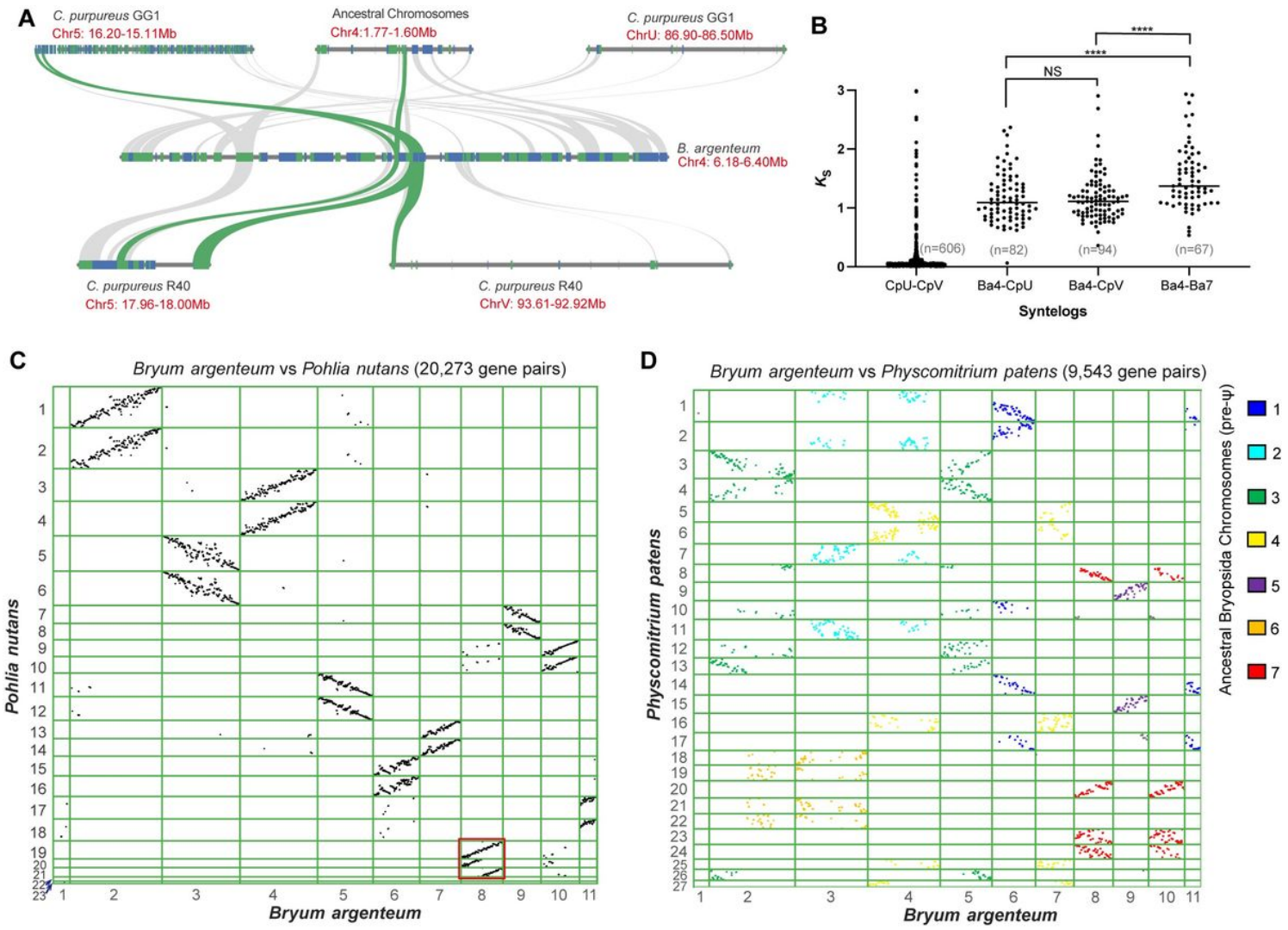


Figure 4

Chromosomal synteny alignments provided multi-faceted clues for illuminating Bryopsida genome evolution. (A) Exemplar local synteny plot depicting micro-syntenies between *B. argenteum* chr. 4 and *C. purpureus* autosomes and sex chromosomes. The fragment of *B. argenteum* chr. 4 represented a modern chromosomal piece of the ancestral Bryopsida chromosome 4 (see methods), harboring genomic syntenies with *C. purpureus* chr. 5 and sex chromosomes. Syntenies suggested ancestral chromosome 4 (or ancestral element D) may have contributed gene contents to sex chromosomes. **(B)** Distribution of K_S of syntenic homologue pairs suggested fusion of genes from ancestral chromosome 4 to non-recombining sex chromosomes postdates the ancestral psi polyploidy event. K_S were estimated using the Nei-Gojobori approach implemented in KaKs_calculator and values larger than 3 were excluded in the analyses. CpU-CpV: syntelog pairs between *C. purpureus* U and V chromosomes; Ba4-CpU: syntelog pairs

between *B. argenteum* chr. 4 and *C. purpureus* U chromosome; Ba4-CpV: syntelog pairs between B. argenteum chr. 4 and C. purpureus V chromosome; Ba4-Ba7: syntelog pairs between *B. argenteum* chrs. 4 and 7, which constituted a part the psi event. **(C)** Genome-wide synteny comparison between *B. argenteum* and *P. nutans* revealed an extra round of WGD and an evident chromosome break of chrs. 20 and 21 in *P. nutans*. **(D)** Genome-wide synteny comparison between *B. argenteum* and *P. patens* recaptured the seven ancestral Bryopsida chromosomes, the seven ancestral Bryopsida chromosomes were depicted as distinct colors and projected onto the synteny dot plots, tracing the corresponding modern chromosomal regions.

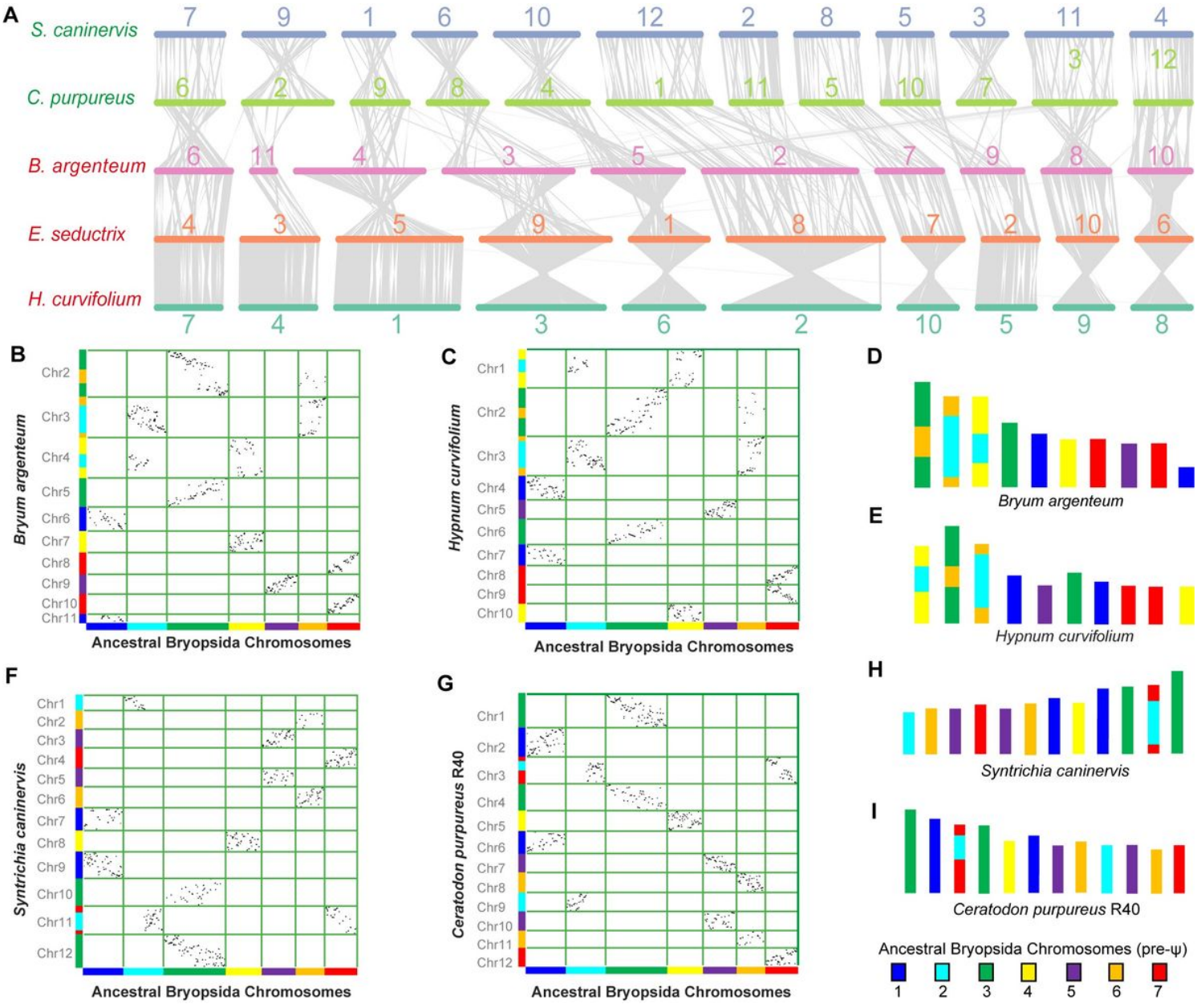


Figure 5

Genomic synteny and inference of modern chromosomal architectures by synteny alignment to the seven ancestral chromosomes. Only autosomes of the moss genomes were analyzed in this figure. The seven

ancestral chromosomes were color-coded as shown in the bottom right. **(A)** Genomic macrosynteny alignments of the five moss genomes including two Dicranidae species (*Syntrichia caninervis* and Ceratodon purpureus R40) and three Bryidae species (*Bryum argenteum*, *Entodon seductrix* and *Hypnum curvifolium*), demonstrating intensive inter-genomic syntenies. **(B)** Dot plot depicting syntelog clusters between *B. argenteum* and the seven ancestral chromosomes. **(C)** Dot plot depicting syntelog clusters between *H. curvifolium* and the seven ancestral chromosomes. The ten autosomes in both species demonstrated similar chromosomal compositions with three chromosomal fusions **(D-E)** Schematic karyotypes demonstrating architecture for each of the autosomes in *B. argenteum* (**D**) and *H. curvifolium* (**E**) based on syntenic alignments to the seven ancestral chromosomes. **(F)** Dot plot depicting syntelog clusters between *S. caninervis* genome and the seven ancestral chromosomes and **(G)** Dot plot depicting syntelog clusters between *C. purpureus* R40 genome and the seven ancestral chromosomes. **(H-I)** Schematic karyotypes demonstrating the architecture for each of the autosomes in *B. argenteum* (H) and *H. curvifolium* (I) based on syntenic alignments to the seven ancestral chromosomes. The twelve autosomes in both species demonstrated similar chromosomal compositions with one conspicuous chromosomal fusion.

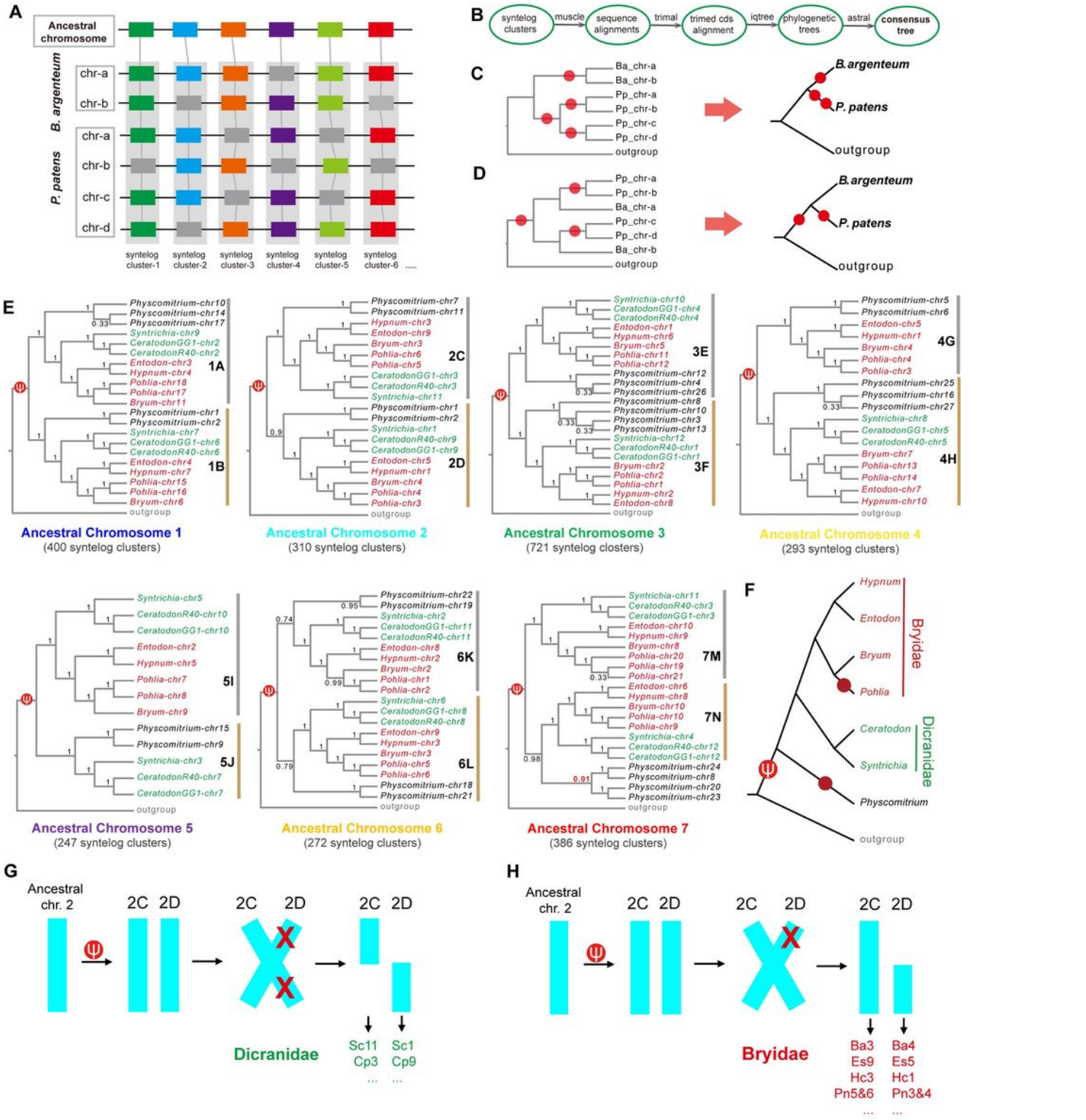


Figure 6

Phylogenomic resolution of the moss subgenomes derived from the ancestral psi polyploidy event. (A) A cartoon depicting the generation of rigorous homologous syntelog clusters by aligning moss genomes (taking *B. argenteum* and *P. patens*as examples) to the reconstructed ancestral chromosomes. Two synteny depths between ancestral chromosomes and analyzed moss genomes (1:4 for *P. patens*and *P. nutans*; 1:2 for others) were considered in the synteny alignment procedure. **(B)** General bioinformatic

pipeline of phylogenomic analyses for each of the ancestral chromosomal descendants. **(C)** A coalescence-based consensus tree that support the first hypothesized scenario assuming multiple independent species-specific whole genome duplications (WGDs). **(D)**A coalescent-based consensus tree that lead to a second hypothesized scenario assuming an ancestral WGD (i.e. the psi event) in the ancestor and and second species-specific WGD in *P. patens*. **(E)** Coalescence-based consensus phylogenies of syntelog clusters derived from the ancestral psi polyploidy event for each of the seven chromosomes. Subgenomes derived from the psi event were labeled to the right of the trees. Branch labels were posterior probabilities and the numbers of syntelog cluster trees were specified under the ancestral chromosomes. **(F)** Evolutionary history of moss chromosomes inferred from the seven consensus phylogenies, which rejected the first hypothesis illustrated in C and confidently support the second hypothesis depicted in D. Phylogenetic position of the psi event and two independent WGDs (red dots) in *P. patens*and *P. nutans* were labeled on the tree. **(G-H)** Proposed evolutionary trajectories of ancestral Bryopsida chromosome 2 in Dicranidae (G) and Bryidae (H) genomes inferred from karyotypic projection coupled with subgenome analyses, demonstrating distinct chromosomal loss patterns in the two lineages. Corresponding modern chromosomes of 2C and 2D subgenomes were also indicated as species abbreviations followed by chrmosome number. Sc11: *S. caninervis* chromosome 11; Cp: *C. purpureus*; Ba: *B. atgenteum*; Es: *E. seductrix*; Hc: *H. curvifolium*; Pn: *P. nutans*.

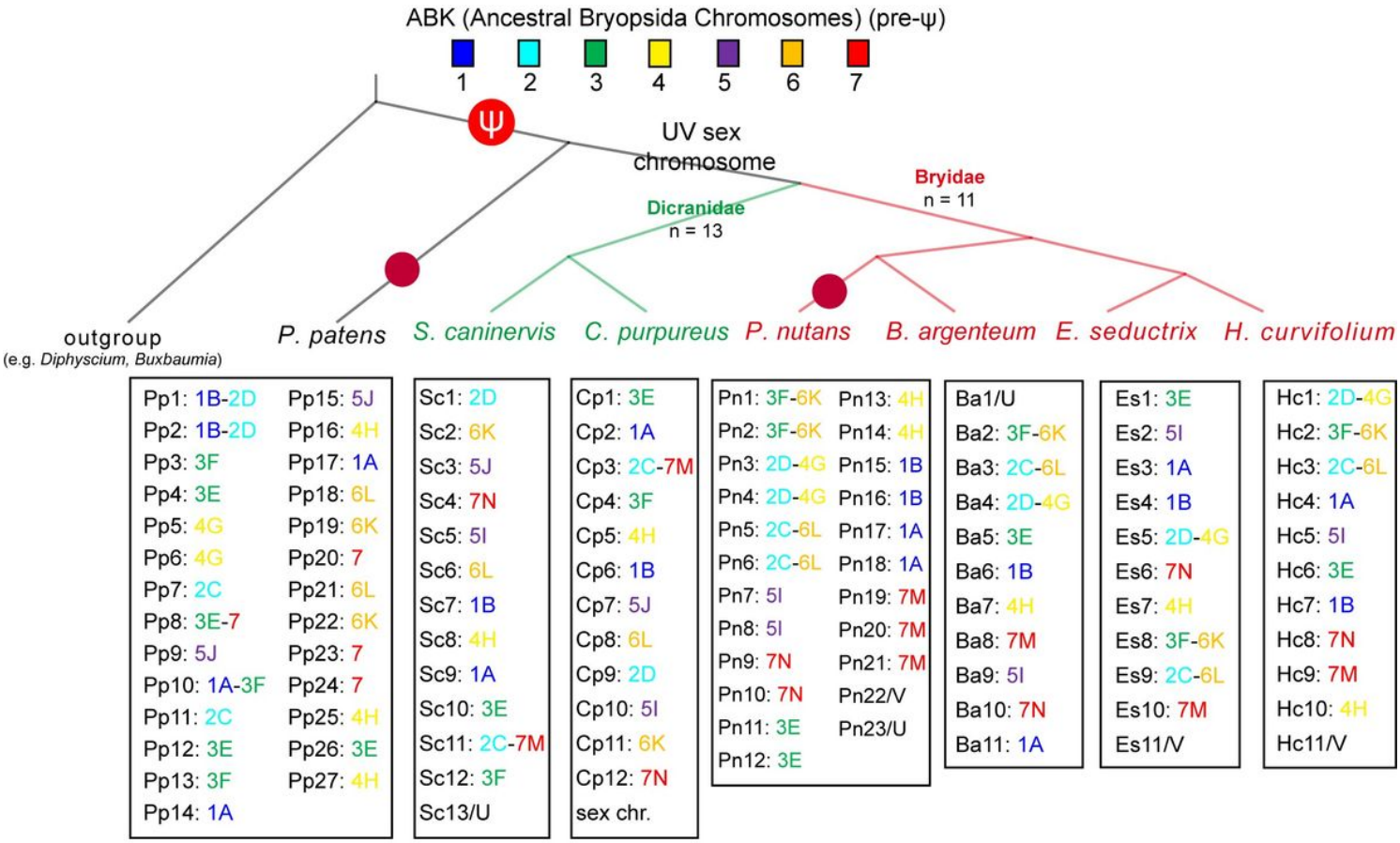


Figure 7

Simplified illustration of the modern genomic architecture including autosomes duplicated the seven ancestral chromosomes and sex chromosomes characterized by UV-linked marker genes. Phylogenetic positions of the polyploidy events are depicted as red dots, inferred ancestral chromosome number for Dicranidae and Bryidae were also indicated. Subgenomic compositions for each of the Bryopsida chromosomes were demonstrated, “-” connecting the two sungenomes indicated corresponding ancestral chromosomal fusions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.xlsx