nature plants

Article

https://doi.org/10.1038/s41477-024-01883-w

Pan-phylum genomes of hornworts reveal conserved autosomes but dynamic accessory and sex chromosomes

Received: 31 May 2024

Accepted: 27 November 2024

Published online: 03 January 2025

Check for updates

Peter Schafran[1] ✉, Duncan A. Hauser[1], Jessica M. Nelson[1], Xia Xu[1], Lukas A. Mueller[1,2], Samarth Kulshrestha[3], Isabel Smalley[4], Sophie de Vries[5], Iker Irisarri[5,6], Jan de Vries[5], Kevin Davies[3], Juan Carlos A. Villarreal[7] & Fay-Wei Li[1,8] ✉

Hornworts, one of the three bryophyte phyla, show some of the deepest divergences in extant land plants, with some families separated by more than 300 million years. Previous hornwort genomes represented only one genus, limiting the ability to infer evolution within hornworts and their early land plant ancestors. Here we report ten new chromosome-scale genomes representing all hornwort families and most of the genera. We found that, despite the deep divergence, synteny was surprisingly conserved across all hornwort genomes, a pattern that might be related to the absence of whole-genome duplication. We further uncovered multiple accessory and putative sex chromosomes that are highly repetitive and CpG methylated. In contrast to autosomes, these chromosomes mostly lack syntenic relationships with one another and are evolutionarily labile. Notable gene retention and losses were identified, including those responsible for flavonoid biosynthesis, stomata patterning and phytohormone reception, which have implications in reconstructing the evolution of early land plants. Together, our pan-phylum genomes revealed an array of conserved and divergent genomic features in hornworts, highlighting the uniqueness of this deeply diverged lineage.

Hornworts (Anthocerotophyta) are the deepest-diverging phylum among the extant land plants. They branched from setaphytes (that is, mosses and liverworts) around 450–479 million years ago (Ma), following the separation of haploid-dominant bryophytes and diploid-dominant vascular plants[1,2]. Far fewer extant species of hornworts exist compared with liverworts and mosses, with around 230, 5,000 and 12,000 species, respectively[3]. Despite their depauperate nature, hornworts are critical to reconstructing the evolution of land plants, considering their deep divergence and unique combination of morphological and physiological traits[4,5].

The recent publications of hornwort genomes from *Anthoceros agrestis*, *A. punctatus* and *A. angustus* not only filled a key gap in plant genomics but also provided new insights into hornwort biology as well as the early transitions in land plant evolution[6–8]. Furthermore, these genomic resources, coupled with the development of *A. agrestis* as a model species[9–14], are enabling investigations into the genetic basis of a

[1]Boyce Thompson Institute, Ithaca, NY, USA. [2]Plant Breeding and Genetics Section, Cornell University, Ithaca, NY, USA. [3]The New Zealand Institute for Plant and Food Research Limited, Palmerston North, New Zealand. [4]Department of Biology, University of Minnesota Duluth, Duluth, MN, USA. [5]Institute of Microbiology and Genetics, Göttingen Center for Molecular Biosciences, Campus Institute Data Science, Department of Applied Bioinformatics, University of Göttingen, Göttingen, Germany. [6]Section Phylogenomics, Centre for Molecular Biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change, Zoological Museum Hamburg, Hamburg, Germany. [7]Department of Biology, Laval University, Quebec City, Quebec, Canada. [8]Plant Biology Section, Cornell University, Ithaca, NY, USA. ✉e-mail: ps997@cornell.edu; fl329@cornell.edu

https://doi.org/10.1038/s41477-024-01883-w

Table 1 | Newly sequenced and/or scaffolded hornwort genomes from this study

Taxon	Sex	Assembly size (Mbp)	Scaffold N50 (Mbp)	Contig N50 (Mbp)	Scaffold number	Assembly in scaffolds (%)	LTR assembly index	Gene number	BUSCO complete (%)
Anthocerotaceae									
Anthoceros agrestis 'Oxford'	Monoicous	127	20	7	6	97.8	19.8	27,068	94
Anthoceros fusiformis	Monoicous	118	27.2	27.2	5	97.1	18.9	21,103	92
Anthoceros punctatus	Monoicous	130	33.4	13.4	4	99.3	15.0	21,280	92
Dendrocerotaceae									
Megaceros flagellaris	Monoicous	194	49.2	12.4	4	99.7	15.1	20,511	92
Phaeomegaceros chiloensis	Unknown	149	37.7	1.5	6	91	11.5	16,015	90
Leiosporocerotaceae									
Leiosporoceros dussii	Male	115	18.1	1.7	6	94.6	17.4	13,175	93
Notothyladaceae									
Notothylas orbicularis	Monoicous	108	28	28	5	99.8	33.0	21,399	92
Paraphymatoceros pearsonii	Monoicous	167	42.8	4.5	6	97.7	17.8	27,066	94
Phaeoceros carolinianus	Monoicous	156	37.1	8.1	4	89.3	18.2	25,343	94
Phaeoceros sp.	Monoicous	139	NA	11.9	NA	NA	23.0	28,604	95
Phymatocerotaceae									
Phymatoceros phymatodes	Male	166	43.6	15.8	5	99.5	14.1	26,031	93

NA, not applicable.

variety of unique hornwort features, such as cyanobacterial symbiosis[15] and the pyrenoid-based CO_2-concentrating mechanism[14,16,17].

Despite these recent advances, the hornwort genomes published so far all come from a single genus *Anthoceros* and, hence, cannot properly represent the entire hornwort lineage nor capture any meaningful diversity within. Although hornworts have a consistent number of autosomes ($n = 4–6$ (ref. 18)), they have undergone relatively dynamic trait evolution. The pyrenoid has been gained and lost across the hornwort phylogeny at least five times, mostly within the past 35 Ma (ref. 19). Stomata, thought to be homologous to those of vascular plants, have been independently lost in multiple hornwort lineages[20]. Likewise, at least 15 transitions between monoicous and dioicous sexual systems—and the evolution of U/V sex chromosomes—have occurred at shallow timescales, often within genera[21]. Finally, whereas the autosome number has remained relatively stable, up to three accessory chromosomes (also known as B or supernumerary chromosomes, which we treat as analogous with micro (m)-chromosomes[22]), were identified in many monoicous species[18], often varying in number within species and even among tissue types within individuals[23].

In this study, we set out to generate pan-phylum genomes of hornworts to better understand the variation across this deeply diverged lineage. We assembled ten new chromosome-scale genomes representing every hornwort family and around 80% of all genera in the phylum. By comparing protein-coding genes, repetitive elements and DNA methylation landscapes, we found a relatively slow rate of structural changes in autosomes. Such genomic stasis is probably due to the absence of whole-genome duplication (WGD) across the phylum. However, unlike autosomes, we found that accessory and putative sex chromosomes are highly dynamic in hornworts and have evolved repeatedly. We further uncovered unique gene retention and losses that help redefine the ancestral traits of land plants.

Results and discussions

Chromosome-level assemblies across the hornwort phylogeny

The previously sequenced hornwort genomes all came from the genus *Anthoceros* (Anthocerotaceae): *A. agrestis* ('Oxford' and 'Bonn' strains), *A. punctatus* and *A. angustus*[6,7], with only the *A. agrestis* 'Bonn' genome at chromosomal scale. To broaden the sampling, we established axenic cultures of nine other species: *A. fusiformis*, *Leiosporoceros dussii* (Leiosporocerotaceae), *Megaceros flagellaris* (Dendrocerotaceae), *Notothylas orbicularis* (Notothyladaceae), *Paraphymatoceros pearsonii* (Notothyladaceae), *Phaeoceros carolinianus* (Notothyladaceae), *Phaeoceros* sp. (Notothyladaceae), *Phaeomegaceros chiloensis* (Dendrocerotaceae) and *Phymatoceros phymatodes* (Phymatocerotaceae) (Table 1 and Supplementary Table 1). From these cultures, we generated high-quality genome assemblies using a combination of Illumina and Oxford Nanopore Technologies (ONT) sequencing. To further scaffold the draft genomes, Hi-C data were generated for all but one species (*Phaeoceros* sp.), as well as *A. agrestis* 'Oxford' and *A. punctatus* for which only contig-level assemblies were available. Four to six putative chromosomes were scaffolded representing 89–99% (median 98%) of each assembly, and the final genomes ranged in size from 108 to 194 Mbp (Table 1 and Supplementary Notes). In total, this study de novo assembled eight chromosomal genomes and one contig-level genome and scaffolded two existing genomes to chromosomes (Fig. 1).

Repeat annotation, gene prediction and 5mC profiling

Repetitive elements account for 22–47% of each genome (Supplementary Table 2), and their abundance does not correlate significantly with genome size ($P = 0.12$, $R^2 = 0.25$), contradicting what was previously proposed[24]. The relative proportions of repeat types vary among species, but common groups are *Copia* and *Gypsy* long terminal repeat retrotransposons (LTRs), *Cacta*, *Mutator*, and *hAT* terminal inverted repeat transposons (TIRs), and *Helitrons* (Supplementary Notes). Gene prediction identified between 13,175 and 28,604 genes per genome (Table 1). Despite the twofold difference in the number of genes, all predicted proteomes have 90–95% complete BUSCO (benchmarking universal single-copy orthologues) genes using the Viridiplantae dataset[25]. Interestingly, 13 BUSCO genes are missing from every species (accounting for ~3% of the Viridiplantae gene set; Supplementary Table 3), suggesting true gene loss in hornworts rather than absence due to assembly or annotation error.

DNA methylation (5-methylcytosine modification) in the CpG context was called using the ONT signal profiles and benchmarked with the bisulfite data (Supplementary Notes and Supplementary Fig. 1). We found that, across all hornworts, gene-body methylation is generally low (Supplementary Fig. 2), which is similar to other bryophytes but different from vascular plants[8,26,27]. Notable exceptions (that is, heavily

https://doi.org/10.1038/s41477-024-01883-w

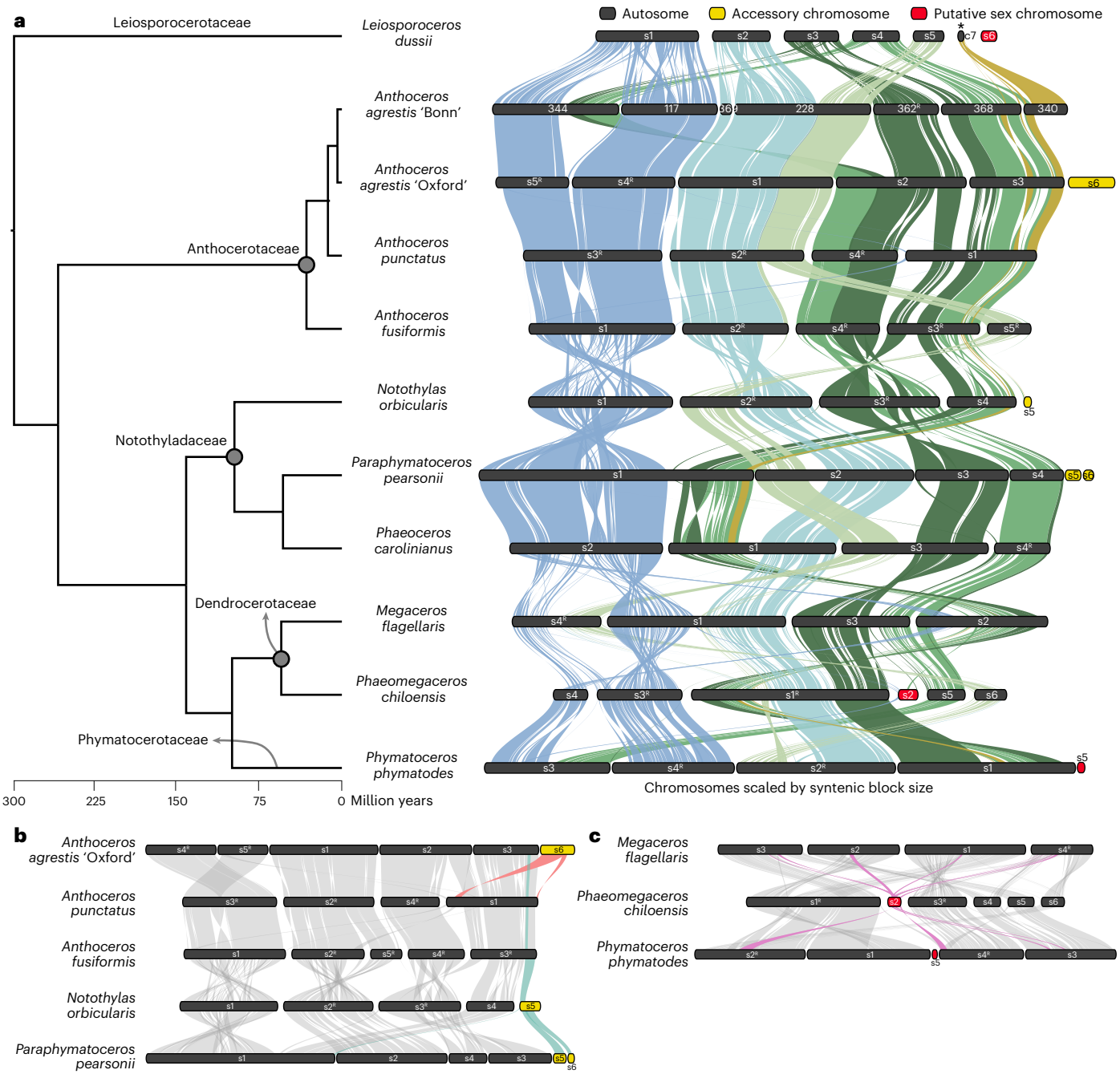

Fig. 1 | Pan-phylum genomes of hornworts. a, Synteny across hornwort genomes. Syntenic blocks are shown as ribbons of different colours. Not all relationships are displayed because synteny was drawn in reference to *Leiosporoceros dussii*. Accessory and putative sex chromosomes (coloured in yellow and red, respectively) have multiple independent origins and are generally not in synteny with one another. The phylogeny was drawn on the basis of Peñaloza-Bojacá et al.[81]. **b**, Detailed view centred around two accessory chromosomes highlighting their relationships with autosomes. **c**, Synteny between the putative sex chromosome of *Phaeomegaceros chiloensis* and multiple autosomes in related species. All the chromosomes depicted here were scaled by syntenic block sizes and therefore do not correlate with physical length. The asterisk denotes an unscaffolded contig having synteny with other chromosomes. Chromosomes marked by a superscript R were inverted for a better view.

methylated genes) can be found on the accessory and putative sex chromosomes (see below; Supplementary Fig. 3). Genes with no or very low expression levels tend to be heavily methylated (Supplementary Fig. 4). Repeat and methylation density are positively correlated (mean $R^2 = 0.53$, $P < 2.2 \times 10^{-16}$; Supplementary Fig. 5).

Putative centromeric regions are variable across hornworts

Previous analyses on the *A. agrestis* genome indicated that gene and repeat content, as well as euchromatic and heterochromatic histone marks, are interspersed throughout the chromosomes[26]. This pattern is shared with mosses and liverworts[8,28] but distinct from angiosperms where transposons are concentrated around the centromeres and protein-coding genes enriched along chromosome arms[29]. Here, we found that the genome organization in *A. agrestis* is not ubiquitous across hornworts, with some having rather large swaths of putative centromeric regions. While centromere locations cannot be predicted by DNA sequences alone, other genomic features, such as increased repeat density, DNA methylation and interchromosomal contact,

https://doi.org/10.1038/s41477-024-01883-w

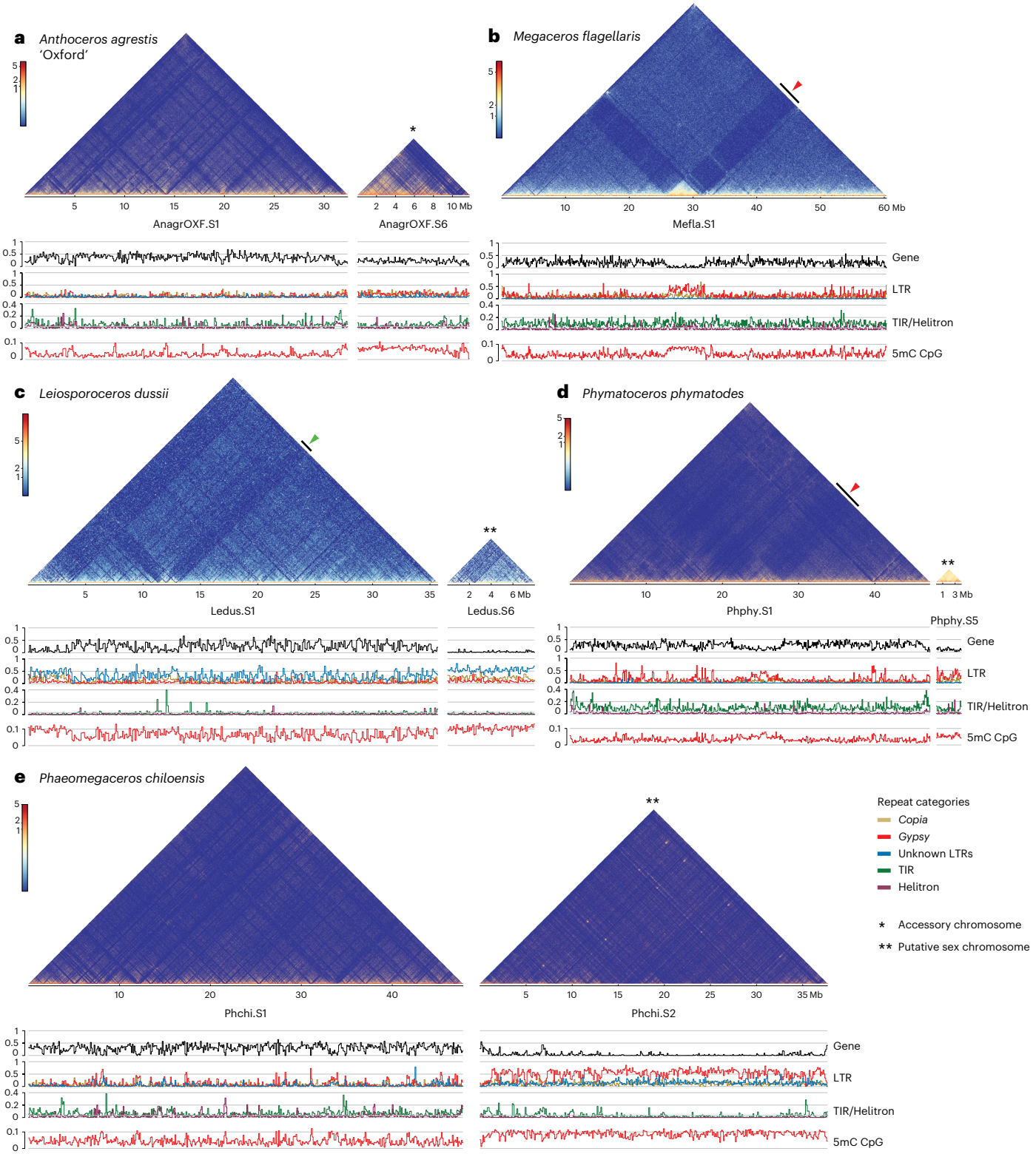

Fig. 2 | Structures of hornwort chromosomes. a–e, Selected chromosomes of *Anthoceros agrestis* 'Oxford' (**a**), *Megaceros flagellaris* (**b**), *Leiosporoceros dussii* (**c**), *Phymatoceros phymatodes* (**d**), and *Phaeomegaceros chiloensis* (**e**), highlighting densities of genes, repeats, CpG methylation and Hi-C contact intensities (as triangular heatmaps). In autosomes of *M. flagellaris* and *P. phymatodes*, large genomic blocks with prominent dips in gene and rise in repeat densities can be seen (marked by red arrowheads) and may correspond to centromeric regions. A different organization was found in *L. dussii* (green arrowhead). Accessory and putative sex chromosomes are denoted by one asterisk and two asterisks, respectively, both characterized by high repeat and low gene contents.

and also decreased gene density and intrachromosomal contact, are often used to identify likely centromere locations[30–33]. Based on these criteria, large putative centromeric regions are clearly discernible in Notothyladaceae, Phymatocerotaceae and one of the genomes in Dendrocerotaceae (Fig. 2b,d, red arrowheads, and Supplementary Figs. 6–15). The repeat composition in the putative centromeres varies

https://doi.org/10.1038/s41477-024-01883-w

between species. In most cases, the dominant repeat is from either the *Gypsy* or *Copia* LTR superfamily, although in some cases both are elevated (*Paraphymatoceros pearsonii*; Supplementary Fig. 11), or DNA transposons instead of LTRs (Noorb.S4 in *Notothylas orbicularis*; Supplementary Fig. 10). This type of chromosome organization with large centromeric regions is notably absent in Anthocerotaceae (Fig. 2a) and *Phaeomegaceros chiloensis* (Dendrocerotaceae; Fig. 2e). In *Leiosporoceros dussii* (Leiosporocerotaceae), blocks of reduced intrachromosomal contact are present and often associated with increased TIR repeats (Fig. 2c, green arrowhead), but the boundaries between putative centromeric and non-centromeric regions are less clear than in *Megaceros* (Fig. 2b) and *Phymatoceros* (Fig. 2d). None of the hornwort genomes contains the specific RLC5 *Copia* elements that constitute centromeres in moss genomes[28,30,31,33]. Altogether, within the hornwort phylum, a high diversity of chromosome organizations exists, with some species being more similar to angiosperms than to other bryophytes. Future histone-profiling studies in species with large putative centromeric regions should provide critical insights.

Hornwort genomes are highly collinear and lack WGD

We next investigated gene collinearity across the 11 chromosome-level genome assemblies of hornworts (10 of which were newly generated here). We found that the chromosome structure is highly conserved despite over 300 million years of evolution (Fig. 1 and Supplementary Fig. 16). About 78–94% of the genomic space was contained within syntenic blocks, containing 16–54% of genes between the focal species and at least one other species. Compared with mosses, liverworts and angiosperms, hornworts have a much slower rate of synteny breakdown (Fig. 3). The relatively stable genome structure over deep time could be attributed to the absence of WGD, as previously shown in *Anthoceros*[6,7]. Looking across all the genomes, distributions of synonymous substitutions (Ks) did not show any peaks indicative of WGD (Supplementary Fig. 17). Likewise, the one-to-one relationship of syntenic blocks among all genomes and the absence of within-genome synteny further support a lack of WGD (Fig. 1) and refute a WGD event previously inferred from transcriptomes[34]. Hornworts are thus so far the only plant phylum conclusively lacking ancient WGD anywhere in their evolutionary history. Liverworts and Selaginellales might be the other two major lineages without WGD based on transcriptomic evidence[8,35], although more genomes across the phylogeny are needed to confirm or reject this claim.

Accessory chromosomes are gene-poor and heavily methylated

Our assemblies identified several gene-poor, repeat-rich chromosomes consistent with early karyotypes that reported small, heterochromatic m-chromosomes (Rink 1935 and Proskauer 1957). We adopt the term accessory chromosome based on evidence suggesting that these can vary in number within species (Fritsch 1991). These accessory chromosomes share the typical features of heterochromatin, including decreased gene density and gene expression levels, while having elevated repeat and methylation densities as well as increased intrachromosomal contact relative to the rest of the genome (Fig. 2 and Supplementary Figs. 6–15 and 18). The heterochromatic nature is also supported by the enrichment of histone marks H3K9me1 and H3K27me1 recently reported on the accessory chromosome of *A. agrestis* 'Oxford' (AnagrOXF.S6)[26].

Unlike autosomes, accessory chromosomes in hornworts appear to be highly dynamic and labile. We found that the accessory chromosomes lack any of the phylogeny-wide synteny seen in the autosomes. There is just one instance of accessory chromosomes from two species being syntenic with each other (that is, Noorb.S5 of *N. orbicularis* versus Papea.S5 and Papea.S6 of *Paraphymatoceros pearsonii*; Fig. 1b, cyan ribbons), pointing to a shared origin around 100 Ma. There is also a small portion of these accessory chromosomes that is syntenic with autosomal regions in *Anthoceros* spp., which suggests a possible ancestral location within an autosome. In addition, accessory chromosomes appear to have a high turnover rate. For example, comparing two sister species *A. agrestis* 'Oxford' and *A. punctatus*[36], only the former possesses an accessory chromosome (Fig. 1b, red ribbons).

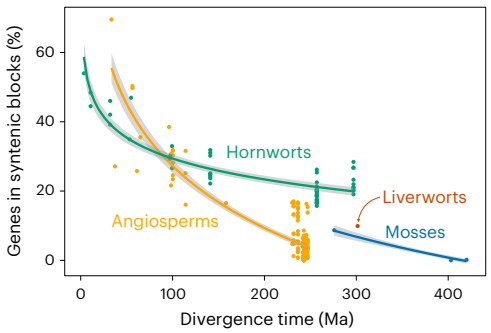

Fig. 3 | Comparison of rates of synteny breakdown. Hornwort genes tend to stay in synteny longer than those in angiosperms, liverworts and mosses. The data points represent pairwise comparisons of percent of syntenic genes with natural logarithmic-fitted lines. The grey bands represent 95% confidence intervals.

To determine whether accessory chromosomes may play a common functional role, we performed a Gene Ontology (GO) enrichment analysis. No GO terms were significantly enriched, but several were significantly depleted ($P < 0.05$ after Bonferroni correction). The depleted terms are often related to housekeeping functions, including those associated with organelles and several metabolic processes (Supplementary Table 4). This finding is consistent with the suppression of gene expression in highly methylated accessory chromosomes in hornworts, suggesting that the presence of critical functional genes was selected against. An opposite pattern has been observed in angiosperms, pointing to different evolutionary origins and selection pressures in accessory chromosomes in other lineages[37]. Hornwort accessory chromosomes are also distinct from the microchromosomes described from vertebrates, which are gene-rich and have widely preserved synteny[38]. It was previously proposed that accessory chromosomes are more likely to occur in lineages with frequent chromosomal rearrangements and unstable chromosome numbers[37]. However, the genomic stability and conserved chromosome numbers in hornworts counter this hypothesis. It is clear that there is a wide spectrum of evolutionary trajectories for accessory chromosomes across the tree of life, and our work provides a preliminary look at such chromosomes in bryophytes.

Sex chromosomes were not derived from accessory chromosomes

Accessory chromosomes of monoecious or monoicous species and sex chromosomes of dioecious or dioicous species have long been linked due to their similar heterochromatin content, meiotic behaviour, size and repeat content[22,23,39,40]. Recently assembled bryophyte sex chromosomes reflect these characteristics[30,33,41]. To test whether accessory and sex chromosomes have shared evolutionary origins, we focused on three dioicous species from three different families, *L. dussii*, *Phaeomegaceros chiloensis* and *Phymatoceros phymatodes*. These species are supposed to have independently evolved U/V sex chromosomes[18,21,23], and indeed, each possessed a scaffold that shared the characteristics of accessory chromosomes mentioned above (Ledus.S6, Phphy.S5 and Phchi.S2). To identify putative sex chromosomes, we resequenced five to six individuals from each of the three species and compared read mapping rates (Supplementary Notes). From each species, we found only one chromosome exhibiting significantly varied read depth compared with all other chromosomes ($P < 2 \times 10^{-16}$), with certain individuals having much reduced mapping rates (Fig. 4).

https://doi.org/10.1038/s41477-024-01883-w

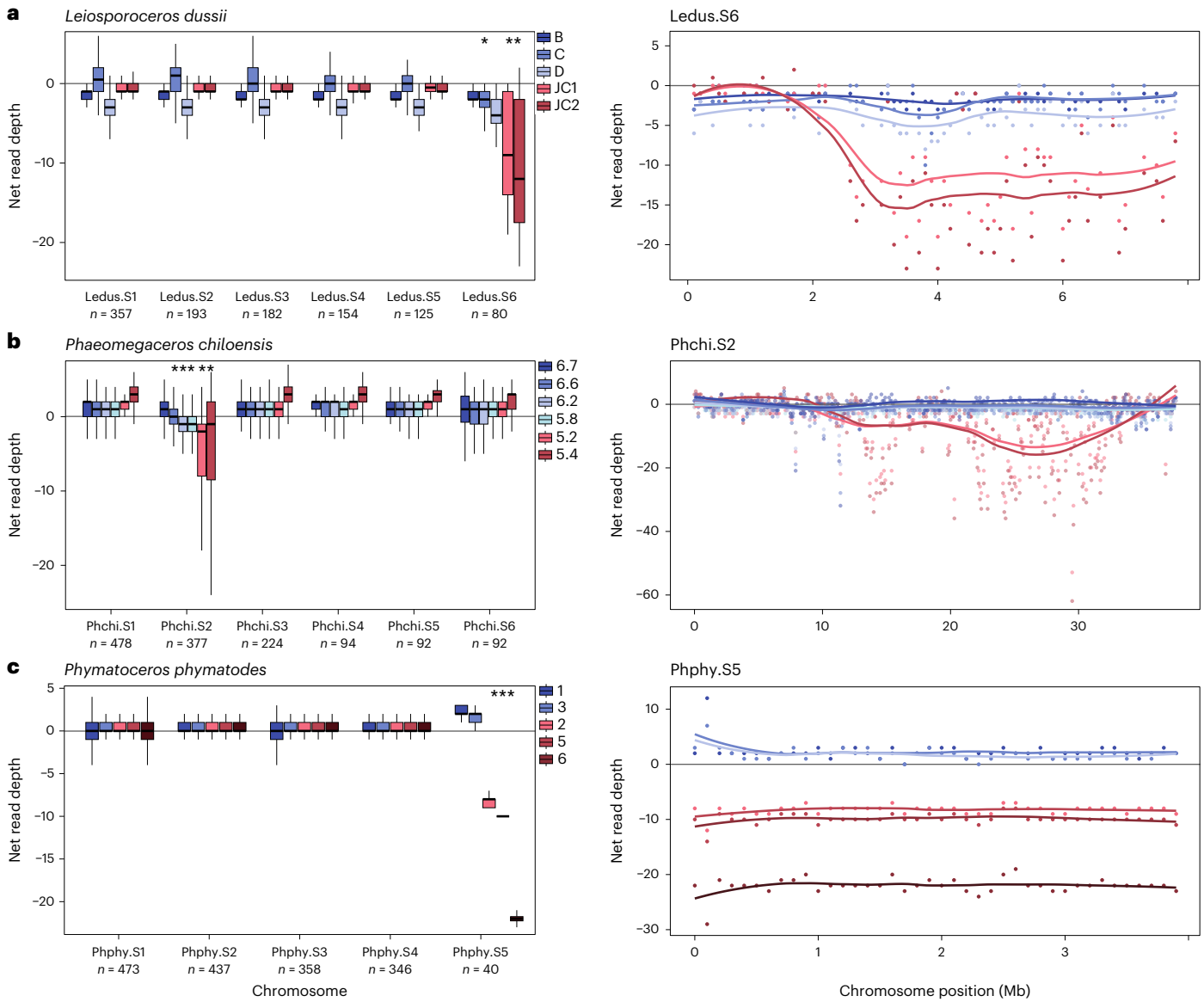

Fig. 4 | Identification of putative sex chromosomes. a–**c**, Resequencing data were generated in three dioicous species: *Leiosporoceros dussii*, (**a**), *Phaeomegaceros chiloensis* (**b**) and *Phymatoceros phymatodes* (**c**). Left: normalized net read mapping depth by chromosome for each individual compared with the reference genome. For each species, some individuals exhibit significantly lower read mapping rates at the putative sex chromosomes based on pairwise Wilcoxon rank-sum tests (indicated by asterisks; *P* values listed in Source data). The box plots represent the range of values calculated in 1 kb windows along each chromosome. The box divisions are the first, second and third quartiles, and the whiskers show ±1.5× the interquartile range. Red versus blue colouration indicates opposite sexes inferred by read mapping patterns over the putative sex chromosomes. Right: a comparison of individual read depth along the putative sex chromosomes. Each dot represents a 100 kb window, connected with locally estimated scatterplot smoothing lines.

Furthermore, these putative sex chromosomes are invariably the ones that resemble the accessory chromosomes from monoicous species. However, contrary to our hypothesis, the putative sex chromosomes lacked any syntenic relationships with accessory chromosomes. Our results thus imply that both accessory and sex chromosomes have evolved independently and repeatedly, and such dynamic evolutionary history is in stark contrast to the relatively stable autosomes.

Based on the presence of gametangia or the size of the thalli, we inferred the reference strains of *L. dussii* and *P. phymatodes* to be male and, thus, Ledus.S6 and Phphy.S5 to be putative V chromosomes (Supplementary Notes). We have not been able to observe *P. chiloensis* gametangia in culture and, hence, cannot determine the sex. In the liverwort *M. polymorpha*, sex determination is controlled by the transcription factor basic pentacysteine on the U chromosome (BPCU)[42]. *BPCU* has a gametologue on V, *BPCV*, which is not required for sex determination. We found that *BPCU*/*V* is a single-copy gene in hornworts (Supplementary Fig. 19) and does not locate on the putative sex chromosomes. Downstream of *BPCU* is a MYB transcription factor encoded by *FGMYB* that controls female development[43]. We found that *FGMYB* is single-copy in most hornwort genomes but absent in *L. dussii*, *Paraphymatoceros pearsonii* and *Phymatoceros phymatodes* (Supplementary Fig. 25). The presence or absence of *FGMYB*, however, does not strictly correspond to species being monoicous or dioicous. These results imply that a different sex determination mechanism exists in hornworts.

Comparing across the putative sex chromosomes, we found 11 shared orthogroups including multiple kinase and transcription factor gene families (Supplementary Table 5) that warrant

https://doi.org/10.1038/s41477-024-01883-w

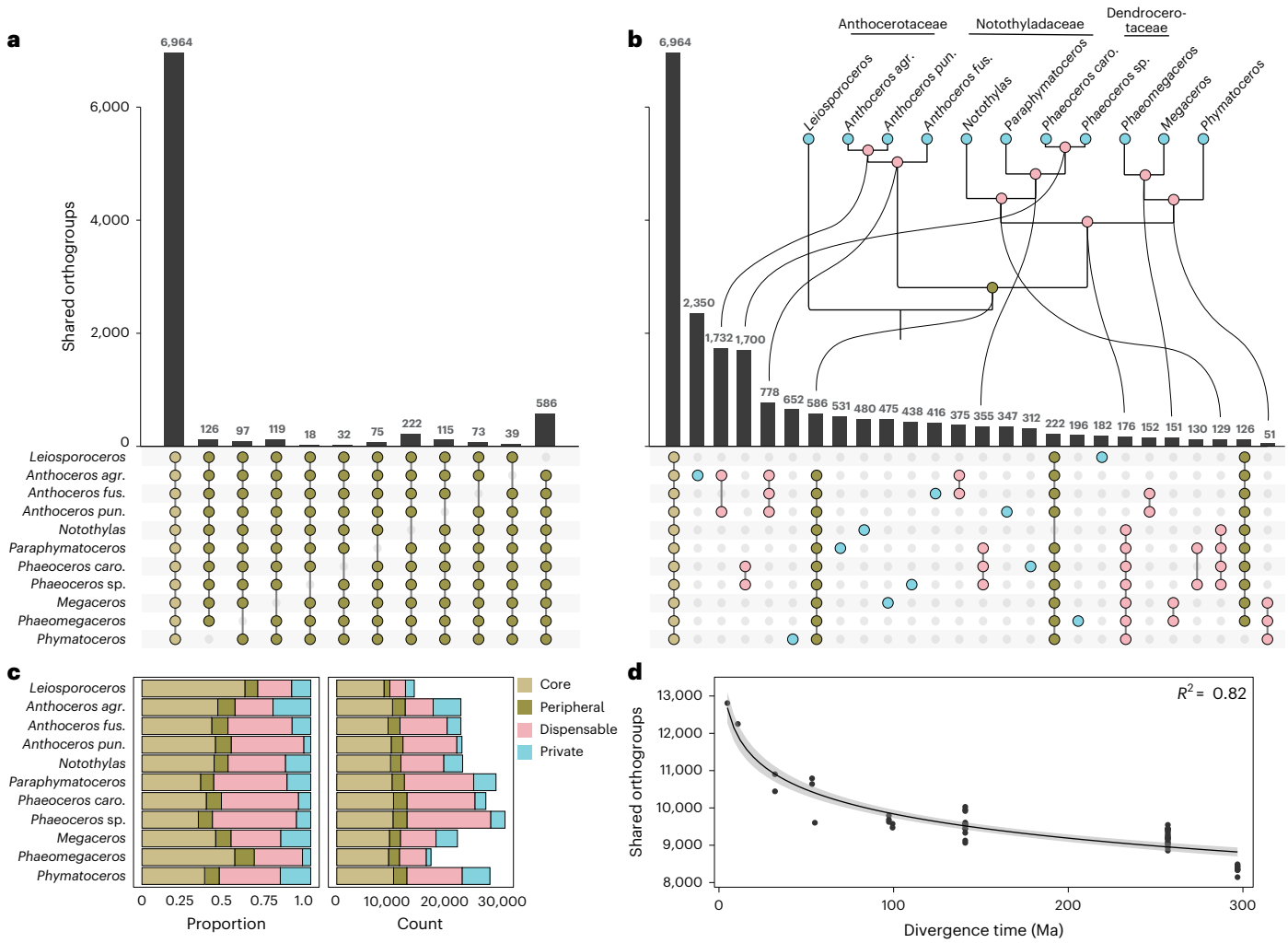

Fig. 5 | Shared and unique gene families across hornwort genomes. a, The distribution of peripheral orthogroups shows that the largest category is the one that has an absence in *Leiosporoceros dussii*. **b**, Categories ranked by the number of shared gene families, highlighting lineage-specific groups. **c**, The proportion and the number of core, peripheral, dispensable and private gene families in each genome. **d**, Negative correlation between the number of shared orthogroups and divergence time across all the pairwise species comparisons. The logarithmic-fitted line is bounded by 95% confidence intervals in grey.

future investigations. None of these shared orthogroups is sex-linked in *Marchantia polymorpha*, *Ceratodon purpureus* and *Lunularia cruciata*, suggesting that hornwort sex chromosomes evolved independently from other bryophytes.

Gene family evolution across hornworts

In addition to comparing broad genomic features across the hornwort phylogeny, we also zoomed into gene family-level dynamics. We used OrthoFinder[44] to infer the 'core', 'peripheral', 'dispensable' and 'private' orthogroups, defined as having copies in all 11 genomes, in 10, between 2 and 9, and in just one genome, respectively. About 27% of all orthogroups were classified as core and 25% as private. Peripheral orthogroups, at 6%, were commonly missing in just *L. dussii*, which is the supposed sister taxon to the rest of hornworts[2] (Fig. 5a). Out of the dispensable orthogroups, most showed clade-specific presence–absence patterns (Fig. 5b). Within each genome, an average of 43% of genes were in core orthogroups, 37% dispensable, 9% peripheral and 11% private (Fig. 5c). As expected, the number of shared orthogroups between any two genomes is negatively correlated with the species divergence time (Fig. 5d). Analysis using CAFE[45] revealed contraction in an unusually large number of orthogroups in *L. dussii*, corresponding with its low number of predicted genes (Supplementary Fig. 20). By contrast, the low number of genes in *P. chiloensis* is not associated with a change in the number of contracting orthogroups. The reason for the difference in gene number between *P. chiloensis* and its closest relative, *M. flagellaris*, is not clear but is partially accounted for by the orthogroups that could not be analysed by CAFE (Methods). The evolution of a few noteworthy orthogroups is discussed below.

PPR proteins and RNA editing

Hornworts are one of the few plant lineages having exceptional high numbers of RNA editing sites in organellar genomes[46]. For example, in *A. agrestis* at least 1,100 C-to-U and 1,300 U-to-C editing sites can be detected in mitogenome and plastome combined[46]. Because RNA editing at each site is typically carried out by a specific pentatricopeptide repeat (PPR) protein[47], one of the largest gene families in most hornwort genomes is the *PPR*s. We found that *L. dussii* has the fewest among all investigated hornworts with 216 *PPR* genes (compared with an average of 914; Supplementary Fig. 21), which partially accounts for its much reduced proteome size compared with other hornwort genomes. Interestingly, *L. dussii* is also known to have very limited RNA editing[48], consistent with its small *PPR* repertoire in the genome. Given the phylogenetic position of *L. dussii*, it is likely that RNA editing, and the concurrent expansion of the *PPR* gene family, took place after *L. dussii* diverged >300 Ma.

Flavonoid biosynthetic and regulatory genes are absent

Flavonoids originated early during plant terrestrialization, possibly as a protective mechanism against the stronger ultraviolet B light exposures in the terrestrial environment[49]. Flavonoid biosynthesis, downstream of the phenylpropanoid pathway, is thought to be ubiquitous in land plants, although, interestingly, flavonoids have yet to be reported from hornworts. The genetic basis for this absence is unknown. We found that most hornworts lack the pathway to flavone or flavonol glycosides. There was no chalcone isomerase (*CHI*) candidate sequence in any of the genomes, except for *L. dussii*. The gene Ledus.5G067800 from *L. dussii* grouped with characterized *CHI* genes from other land plants with strong support (Supplementary Fig. 22 and Supplementary Notes). Analysis of the intron and exon structure of *CHI* and chalcone isomerase-like (*CHIL*) genes also strongly suggested that Ledus.5G067800 is a *CHI* orthologue (Supplementary Fig. 22).

In addition to the general lack of key flavonoid biosynthetic genes, none of the hornwort genomes (including *L. dussii*) possesses orthologues of the transcription factors required for flavonoid pathway activation in other land plants (namely, specific subclades of R2R3MYBs and bHLH; Supplementary Notes and Supplementary Figs. 22–26). In summary, hornworts uniquely lack the complete flavonoid biosynthetic machinery (with the possible exception of *L. dussii*). In addition to their response to ultraviolet B light, it would be interesting to examine whether hornworts differ in the varied flavonoid-mediated pathways identified in other land plants, such as interactions with pathogenic or beneficial microorganisms. Because both liverworts and mosses produce flavonoids, it is likely that the pathway was secondarily lost in hornworts. This hypothesis is further supported by the single, and possibly relict, orthologous flavonoid gene (*LdCHI*) in *L. dussii*. Future work should focus on analysis of *L. dussii* to investigate the presence of flavonoid-like compounds.

Stomatal patterning genes in species that lost stomata

Stomata probably have a deep homology across land plants[50]. In hornworts, stomata are restricted to sporophytes and hypothesized to function as air pores to desiccate spores for dispersal[20]. We previously showed that the orthologues of genes encoding transcription factors involved in stomatal patterning in flowering plants, including *SMF*, *SCRM*, *TMM* and *EPF*, were upregulated in *A. agrestis* sporophytes, supporting their conserved functions in hornworts[6]. Here, we found that *EPF* orthologues were, however, absent in most hornworts, suggesting it does not have a core role in stomatal development.

Stomata have independently been lost twice during hornwort evolution, once in *Notothylas* and another time in the clade containing *Megaceros*, *Nothoceros* and *Dendroceros*, offering a unique opportunity to examine the fate of genes when the corresponding trait is lost. We showed that *TMM* homologues were uniquely absent in both *N. orbicularis* and *M. flagellaris*. In *N. orbicularis*, orthologues of *SMF* and *SCRM* are present but often contain unique in-frame insertions making the predicted proteins much longer (Supplementary Fig. 27). The expression levels of these orthologues were also low and did not significantly differ between sporophyte and gametophyte tissues ($P > 0.3$; Supplementary Fig. 28). We thus hypothesize that the stomatal patterning genes in *N. orbicularis* might be in the early stages of degeneration.

Meanwhile, the *M. flagellaris SMF* and *SCRM* orthologues appear to be conserved at the sequence level and we could not find any evidence of pseudogenization. It is possible that stomata were only recently lost in *Megaceros* and not enough time has lapsed for pseudogenization to occur. Indeed, 'unspecified pores' have been observed in *Megaceros* sporophytes[51], which implies that partially degenerated stomata might still develop. Alternatively, Fortin and Friedman[52] recently advocated that the pores present on hornwort gametophytes are homologous to stomata. If true, the conservation of *SMF* and *SCRM* orthologues in *M. flagellaris* could reflect the fact that gametophytic pores are retained in this species. Indeed, while *TMM* is absent in both hornwort species that lack stomata, knocking out *TMM* alone in *Physcomitrium* did not eliminate stomata[53]. In summary, our data suggest that orthologues of stomatal genes have not completely lost or pseudogenized following the loss of sporophytic stomata.

In addition to the candidate gene approach, we also carried out an untargeted search, looking for orthogroups that are convergently lost in both *N. orbicularis* and *M. flagellaris*. Only six orthogroups were found, and none contained genes known to be associated with stomata development (Supplementary Table 6).

Gibberellin receptor is retained in a hornwort lineage

Gibberellin (GA) is a key phytohormone that regulates a suite of plant developmental processes[54], but its role in hornworts is unclear. Whereas the gene encoding the GA receptor GID1 was found to be absent in bryophyte genomes including those of *Anthoceros*[6], transcript fragments resembling a *GID1* gene were reported in the transcriptomes of *Phaeoceros carolinianus* and *Paraphymatoceros halli*[55]. Here, we were able to identify full-length *GID1* sequences in the genomes of *Notothylas orbicularis, Phaeoceros* sp., *P. carolinianus*, *Paraphymatoceros pearsonii, Phaeomegaceros chiloensis* and *Phymatoceros phymatodes*. Phylogenetic reconstruction further confirmed that these sequences are indeed *GID1* orthologues (Supplementary Fig. 29a). No *GID1* was found in any other hornwort genome. Based on its phylogenetic distribution, *GID1* was most likely present in the last common ancestor of land plants and experienced multiple losses in bryophytes. The finding here highlights the importance of having phylodiverse sets of genomes to infer ancestral states. Interestingly, at most *GID1* residues important for binding GA in vascular plants[56], hornworts have alternative amino acids uncharacteristic of *GID1*, *GID1*-like or closely related carboxylesterase (*CXE*) genes (Supplementary Fig. 29b). By contrast, these sites are highly conserved over ca. 125 Ma within hornworts, potentially indicating an affinity for a different substrate.

As in previously published *Anthoceros* genomes[6], we found no orthologues of genes encoding GA20ox in any species of hornwort, which is essential in the biosynthesis of active GA. Other components of the pathway, including orthologues of genes encoding CPS, KS, KO, KAO, GA13ox, GA2ox and GA3ox, were present and transcribed in species both with and without *GID1*. The moss *Physcomitrium patens* and the liverwort *Marchantia polymorpha* likewise lack a complete GA biosynthetic pathway but produce other *ent*-kaurenoic acid (KA) derivatives that are biologically active[57,58]. *Physcomitrium patens* also contains a *GA3ox* orthologue (*PpKA2ox*) that inactivates KA[59]. We hypothesize that hornworts have similarly evolved alternative active compounds derived from KA or GA_{12}, which their GA oxidase orthologues are interacting with. Whether GID1 is the receptor for such compounds awaits future research.

Abscisic acid receptor occupies an unusual phylogenetic position

The phytohormone abscisic acid (ABA) is one of the major signals that regulate the response of plants to stress[60]. The canonical ABA receptor belongs to the pyrabactin resistance1-like family (*PYL*), and a single *PYL* is thought to be present in the last common ancestor of land plants. Based on characterizations of angiosperm proteins, *PYL*s have been divided into three subfamilies[61], with subfamily I containing *PYL* homologues from across tracheophytes, whereas the liverwort and moss *PYL* radiated independently[62]. We found that *PYL* is single-copy in hornworts, and surprisingly, all hornwort *PYL*s fell into the (formerly) tracheophyte-specific subfamily I (Supplementary Fig. 30) and are distant from the other bryophyte *PYL*s. This unique phylogenetic position raises the possibility that subfamily I might represent the ancestral *PYL* clade and was present in the last ancestor of extant land plants.

Conclusion

We present chromosome-level genome assemblies for ten hornwort species, covering all families, and over 300 million years of divergence. Our efforts now make hornworts the most densely sampled plant phylum in terms of proportion of the species diversity sequenced. With this phylum-wide dataset, we demonstrated that the previously published genomes do not adequately capture the genomic diversity of hornworts. For instance, we found that one major lineage contains large repeat-dense blocks of enriched interchromatin contact consistent with putative centromeric regions. This feature is not shared with other sequenced bryophyte genomes. Our study assembles and characterizes the elusive accessory chromosomes from bryophytes, showing that hornwort accessory chromosomes have recurrent origins and are characterized by high repeat content and CpG methylation as well as reduced gene presence and expression. Putative sex chromosomes were identified in three species, and we found no evidence that they were derived from accessory chromosomes. The evolutionarily labile nature of accessory and sex chromosomes is in stark contrast to autosomes, which have remained largely syntenic over deep time. We hypothesize that the relatively stable autosomes might be due to the absence of any detectable ancient WGD, an extremely rare phenomenon among land plants. Pan-phylum genomic comparison revealed that roughly 40% of the genes in each hornwort genome can be classified as 'core' hornwort genes. This number will serve as a key reference point for future comparisons with other plant phyla. Lastly, the detailed analyses of genes involved in RNA editing, flavonoid biosynthesis, stomatal patterning, and GA and ABA receptors helped refine the reconstructions of the early evolution of land plants. Altogether, by greatly expanding the available genomes in hornworts—and concomitantly enabling genetic transformation across the phylogeny[14]—we anticipate that our resources will facilitate future comparative analyses to gain further insights into the origin and early evolution of land plants.

Methods

Axenic hornwort culturing

Axenic cultures for most species were established by sterilizing mature spores in 500 µl of 2% (v/v) bleach for 2 min, after which the reaction was halted with an equal volume of filter-sterilized 0.1 M sodium thiosulfate. The liquid containing the spores was dispersed onto Petri plates containing AG medium[11] lacking sucrose and supplemented with 200 mg l^{-1} timentin, 20 mg l^{-1} rifampicin and 50 mg l^{-1} vancomycin. The plates were sealed and placed in a Percival Plant Tissue Culture Chamber with a 16:8 day:night cycle at 22 °C. Spores generally germinated within 1–3 months, after which they were transferred to plates containing standard AG medium supplemented with 0.1% (w/v) sucrose. For *Phymatoceros phymatodes*, the axenic culture was obtained by sterilizing tubers instead of spores, following the same protocol as above. Plants were subdivided and transferred to new plates as needed, approximately every 1–2 months.

DNA and RNA extraction

DNA was isolated with a modified cetyltrimethylammonium bromide (CTAB) precipitation method to reduce the coprecipitation of polysaccharides that are abundant in hornwort species. To preserve the maximum length of DNA fragments, all transfer steps were done with wide-bore pipette tips and all mixing was done by gentle inversion. Approximately 1 g of fresh thallus tissue was ground to a fine powder in liquid nitrogen, then transferred into 20 ml of lysis buffer containing 2% CTAB, 0.1% (v/v) β-mercaptoethanol, 100 mM Tris, 20 mM ethylenediaminetetraacetic acid, 1.4 M NaCl and 1% (w/v) polyvinylpyrrolidone (PVP), preheated to 65 °C. After a minimum incubation of 1 h at 65 °C, the solution was washed with an equal volume of 24:1 chloroform:isoamyl alcohol and centrifuged for 10 min at 4,000*g*. The supernatant was transferred to a new tube and mixed with 1/10 volume of 10× CTAB buffer (10% CTAB and 0.7 M NaCl), and the chloroform:isoamyl extraction was repeated. After the second extraction, the supernatant was moved to a new tube with an equal volume of CTAB precipitation buffer (1% CTAB, 50 mM Tris and 10 mM ethylenediaminetetraacetic acid). The solution was gently mixed while incubating at 55 °C for 30 min, after which it was centrifuged at 4,000*g* for 30 min. All liquid was discarded, leaving a pellet of CTAB–DNA complexes. The pellet was resuspended in 1 ml of high-salt TE buffer containing 1 M NaCl by mixing for at least 1 h. The tube was centrifuged at 4,000*g* for 10 min to pellet any insoluble material, and the liquid was transferred to a new 1.5 ml centrifuge tube. An equal volume of isopropanol was added and mixed for 30 s, immediately followed by centrifuging at 10,000*g* for 30 min. The liquid was carefully removed without disturbing the DNA pellet. The DNA was washed with 1 ml cold 70% (v/v) ethanol, centrifuged for 5 min at 10,000*g*, and the liquid was discarded. The 70% ethanol wash was repeated once. After the second wash, the tube was left open to dry in a laminar flow hood, approximately 1 h. DNA was resuspended in 50 µl of 10 mM Tris. The quantity was measured by a Qubit 1× dsDNA HS kit and quality assessed by a NanoDrop One spectrophotometer.

RNA was extracted with a Sigma Spectrum Plant Total RNA kit (STRN250; Sigma-Aldrich) using 50 mg of fresh sterile thallus tissue ground in liquid nitrogen. Sporophytes and thallus tissue of four individuals of *Notothylas orbicularis* (collected in the Roy H Park Preserve, 42.42570 N, 76.33571 W, 12 August 2020) were also extracted using the same method.

Sequencing

Total genomic DNA was sequenced on the ONT MinION with R9 flowcells. Libraries were prepared from 2 µg gDNA with a ligation sequencing kit (SQK-LSK109). Sequencing was performed for 72 h or until <1% of pores were still sequencing. If any library was left over following the initial loading, the flowcell was washed with nuclease buffer and a second sequencing run was performed. Signal files were basecalled in Guppy v6.4.2 with the super accuracy (SUP) r9.4.1 v3.4 model. Short-read DNA sequence data were generated from 100 ng of gDNA prepared with the sparQ DNA library prep kit (Quantabio). Libraries were sequenced to approximately 50× genome depth on the Illumina NovaSeq 5000 platform by Novogene. Eukaryotic mRNA libraries were prepared by Novogene and sequenced on the NovaSeq 5000 to 7 Gbp per sample for use in genome annotation and 15 Gbp per sample for *Notothylas* gene expression analysis.

Genome assembly

Draft genomes were assembled with Flye 2.8 (ref. 63) using ONT reads filtered to remove reads shorter than 5 kb and with adapter sequence trimmed using porechop 0.2.4 (https://github.com/rrwick/Porechop), which yielded 50–60× genome depth for most species. Contigs were error-corrected with Illumina DNA sequences using three iterations of Pilon 1.24 (ref. 64). Draft genomes were scaffolded using Hi-C libraries prepared and processed by Phase Genomics. TGS-Gapcloser 1.1.1 (ref. 65) was used to fill gaps between scaffolds with ONT reads and polish filled gaps with Illumina data. Scaffolds and contigs were ordered by decreasing length and numbered incrementally. Any likely contaminant sequences were removed after visualizing genomes in BlobTools2 (ref. 66). Using a combination of Illumina and ONT read depth, contig GC content, BLAST + 2.10.0 hits against the National Center for Biotechnology Information (NCBI) nt database, and Hi-C contact patterns, contigs that showed clear differences from the rest of the genome in at least two of these criteria were removed. Genome completeness was estimated with BUSCO 5.2.1 using the Viridiplantae dataset[25].

DNA methylation analyses

5mC bases in CpG context sites were called from ONT DNA sequencing data using Megalodon 2.5.0 (https://github.com/nanoporetech/megalodon) with the R9.4.1_e8_HAC model. The methylation calls

from the ONT data for *Anthoceros agrestis* 'Oxford' were compared against those based on bisulfite sequencing performed on the same accession[26]. Bisulfite sequence data for *A. agrestis* 'Oxford' were downloaded from ENA (SRR22428133) and processed with bismark 0.24.1 (ref. 67). Based on a strong bimodal distribution of modification frequency observed in all species, data were subset into CG sites with either greater or less than 50% modification frequency. To evaluate the correlation between modified sites detected by ONT and bisulfite data, read depth and frequency of modifications at each site were calculated from the output of either Megalodon or bismark, respectively. 5mC CG modification over gene bodies and adjacent non-genic regions within each genome was calculated from gene models and 5mC modification GFF files using custom code and plotted with R.

Genome annotation

Custom repeat libraries were constructed for each species using EDTA 2 (ref. 68). The transposable element (TE) library output by EDTA was filtered by extracting LTRs labelled as unknown and searching their nucleotide sequences against a database of transposases by BLAST. Any sequences with an *e*-value $\leq 1 \times 10^{-10}$ were retained as part of the transposable element library, and the others were removed. The transposable element library was then searched against known plant proteins in Uniprot by BLAST to remove any protein potentially misidentified as transposable element. RepeatMasker v4.1.0 (ref. 69) was used with this final transposable element library to soft mask each genome. The LTR assembly index was calculated for each genome with the LTR_retriever script included with EDTA[70]. Additional screening for tandem repeats was done with Tandem Repeats Finder 4.09.1 (ref. 71).

Gene models were predicted using BRAKER2 (ref. 72), with input consisting of Illumina RNA reads mapped to the soft-masked genome using HISAT2 (ref. 73) and predicted hornwort proteins from the published *Anthoceros* genomes[6,7]. BRAKER2 output files were screened for genes with in-frame stop codons, which were marked as pseudogenes in the corresponding GTF file. Genes were renamed to contain their respective scaffold/contig name plus a number incremented by 100, restarting at the beginning of each scaffold/contig. Subsets of primary transcripts were created by selecting the longest transcript associated with each gene. Gene functional annotation was performed with EggNOG-mapper 2 (ref. 74) using DIAMOND 2.0.15 (ref. 75) as the search engine.

Densities of genomic features across each genome were calculated from their respective GFF and GTF files converted to BED format and analysed with bedtools 2.29.2 (ref. 76), taking the average number of base pairs within each feature over 100 kb windows. Results were plotted with HiCExplorer 3.7.2 (ref. 77) and pyGenomeTracks 3.8 (ref. 78).

Synteny, chromosome evolution and WGD

Synteny among hornwort genomes was inferred and plotted using GENESPACE 1.3 (ref. 79). To calculate the rate of synteny breakdown, we selected approximately chromosome-level assemblies from mosses, liverworts and angiosperms that capture similar ranges of divergence time and phylogenetic diversity, and analysed each phylum in GENESPACE (Supplementary Table 7). The number of genes in all syntenic blocks was extracted from GENESPACE output files and divided by the average number of genes in each pairwise comparison. Divergence times between species pairs were from Bechteler et al.[2], Zuntini et al.[80] and Peñaloza-Bojacá et al.[81]. Distributions of synonymous substitution (Ks) were constructed for each species using the primary transcripts with wgd 1.0 (ref. 82). GO term enrichment and purification analysis comparing autosomes versus accessory/sex chromosomes was performed using goatools 1.2.3 (ref. 83), with a cut-off of Bonferroni-corrected *P* value <0.05.

Sex chromosome inference

For dioicous species, five to ten individuals from single populations were sequenced to approximately 30× depth on Illumina NovaSeq 5000. Organellar reads were removed by mapping onto chloroplast and mitochondrial genome assemblies with BWA 0.7.17 (ref. 84), and PCR duplicates were removed using samblaster 0.1.26 (ref. 85). The cleaned reads were mapped to their respective reference genomes, and read depth at each site was calculated with samtools 1.18 (ref. 86). Illumina whole-genome sequencing data from the reference genomes were filtered in the same way and subsampled to the same amount of data as the other individuals of the same species; these datasets were mapped to the reference genomes to represent read depth control. The resulting read depth data were averaged over 100 kbp windows and filtered to remove outliers with >10× of the expected read depth. Net read depth difference was calculated by subtracting each sample's site depths from the respective reference site depths. Between-chromosome differences in read depth were analysed by pairwise Wilcoxon rank-sum tests with Bonferroni-corrected *P* values.

Gene expression analysis

Gene expression was quantified by mapping RNA reads to the respective genome with HISAT2 (ref. 73) and analysed with StringTie 2.1.1 (ref. 87). HISAT2 indexes were built with splice site and exon information generated from the genome's GTF annotation file. The HISAT2 script 'hisat2_extract_exons.py' was modified to work with GTF files produced by BRAKER. Expression levels were qualitatively classified based on the standard deviation of $\log_{10}$ transformation of transcript per million, which had a roughly normal distribution. More than 2 standard deviations (s.d) above the mean, very high; more than 1 s.d. above the mean, high; between 1 and −1 s.d. from the mean, average; less than 1 s.d. below the mean, low; less than 2 s.d. below the mean, very low.

Differential expression in *Notothylas orbicularis* (sporophytes versus gametophytes) followed the HISAT2–StringTie–Ballgown pipeline[88]. In brief, each replicate RNAseq dataset was mapped onto the genome with HISAT2 and the alignment was used to assemble transcripts with StringTie. Assemblies were merged with StringTie to create a non-redundant set of transcripts, and then transcript abundance was estimated for each individual replicate.

Gene family evolution

Orthogroups were inferred from all the hornwort genomes generated in this project, plus other Viridiplantae representatives (Supplementary Table 8). Orthogroup inference was performed with OrthoFinder 2.5.4 (ref. 44) using DIAMOND for amino acid sequence comparisons, MAFFT[89] for alignment and fasttree[90] for phylogenetic tree construction. The hornwort-only orthogroups including *A. agrestis* 'Bonn'[6] and *A. angustus*[7] found many orthogroups only missing in these two genomes, which probably reflect the incompleteness in gene prediction. For this reason, we excluded *A. agrestis* 'Bonn' and *A. angustus* from pan-genome classification. Following orthogroup inference, each orthogroup was aligned with Clustal Omega 1.2.3 (ref. 91), alignments were trimmed to remove sites containing >90% gaps with TrimAl 1.4 (ref. 92) and phylogenetic trees were inferred with IQ-TREE 2.0.3 (ref. 93) using ModelFinder[94] to identify the best evolutionary model for each orthogroup and 5,000 ultrafast bootstrap replicates to determine branch support values. Trees were visualized in FigTree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and alignments in Geneious Prime 2023.0.4 (www.geneious.com).

With 95% highest probability density intervals for phylum and order levels from the time-calibrated tree in Bechteler et al.[2] as constraints, we generated a chronogram in r8s (ref. 95) from the Viridiplantae OrthoFinder results pruned to represent only bryophytes. The chronogram and OrthoFinder results were input to CAFE5 (ref. 45) to estimate the number of orthogroups with expansions or contractions

throughout the phylogeny. Forty-seven orthogroups with more than 100 members in any individual taxon had to be removed to generate non-infinite log-likelihood scores. Figures were generated with CafePlotter v0.2.0 (https://github.com/moshi4/CafePlotter).

The phylogenetic analysis of the *PYR/PYL/RCAR* family was conducted on the basis of the dataset from Sun et al.[62], with additional homologues from *Isoetes taiwanensis*[96] and hornworts. To identify hornwort homologues, we used BLASTP with an *e*-value cutoff of 10^{-10}. Sequence alignment was done using MAFFT[89], and a maximum likelihood phylogeny was computed using IQ-TREE v1.5.5 (ref. 93) with the best-fit model (JTT + Γ4 according to Bayesian information criterion) determined by ModelFinder[94] and 1000 UFBoot[97] pseudoreplicates. The tree was visualized using Interactive Tree of Life (iTOL)[98].

PPR gene identification followed Zhang et al.[7], with pfam_scan v1.6 (https://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/) searching all predicted protein sequences for the *PPR* Pfam profile PF01535 and retaining any sequences with more than one occurrence of the *PPR* repeat motif. These were combined with sequences containing the DYW deaminase motif (PFAM profile PF14432).

HornwortBase

All new hornwort genomes are hosted on HornwortBase (http://www.hornwortbase.org), a website designed on the Breedbase platform[99]. Genomes are available for download along with associated files, including gene model annotations, repeat annotations, gene sequences, coding sequences, translated coding sequences, Hi-C contact matrices and 5mC modified bases. Users can BLAST-search any of the full-length genomes or gene model sequences, with connections to orthogroups from the Viridiplantae dataset, gene functional information and gene expression. JBrowse2 (ref. 100) is incorporated to visualize genome tracks and gene models, Hi-C heatmaps and intergenomic synteny.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequence data are deposited at NCBI SRA under BioProject PRJNA996135 (see Supplementary Table 9 for individual SRA accessions). Previously published data utilized in this study are available in BioProjects PRJNA574453 and PRJNA574424. Hornwort genome sequence and annotation files used in this study are v1.0 versions available via HornwortBase at https://hornwortbase.org/ftp/ and via Figshare at https://doi.org/10.6084/m9.figshare.26053480.v2 (ref. 101). Source data are provided with this paper.

Code availability

Custom R and Python codes and bioinformatic steps used in this study are available via GitHub at https://pschafran.github.io/jekyll/update/2024/04/30/hornwort-genomes/.

References

1. Harris, B. J. et al. Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nat. Ecol. Evol.* **6**, 1634–1643 (2022).
2. Bechteler, J. et al. Comprehensive phylogenomic time tree of bryophytes reveals deep relationships and uncovers gene incongruences in the last 500 million years of diversification. *Am. J. Bot.* **110**, e16249 (2023).
3. Söderström, L. et al. World checklist of hornworts and liverworts. *PhytoKeys* **59**, 1–828 (2016).
4. Renzaglia, K. S., Villarreal, J. C. & Duff, R. J. in *Bryophyte Biology* (ed. Goffinet, B.) Vol. 2, 139–171 (Cambridge Univ. Press, 2008).
5. Frangedakis, E. et al. The hornworts: morphology, evolution and development. *New Phytol.* **229**, 735–754 (2021).
6. Li, F.-W. et al. Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* **6**, 259–272 (2020).
7. Zhang, J. et al. The hornwort genome and early land plant evolution. *Nat. Plants* **6**, 107–118 (2020).
8. Szövényi, P., Gunadi, A. & Li, F.-W. Charting the genomic landscape of seed-free plants. *Nat. Plants* **7**, 554–565 (2021).
9. Frangedakis, E. et al. An Agrobacterium-mediated stable transformation technique for the hornwort model Anthoceros agrestis. *New Phytol.* **232**, 1488–1505 (2021).
10. Szövényi, P. et al. Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biol.* **15**, 98 (2015).
11. Gunadi, A., Li, F.-W., & van Eck, J. Accelerating gametophytic growth in the model hornwort *Anthoceros agrestis*. *Appl. Plant Sci.* **10**, e11460 (2022).
12. Neubauer, A. et al. Step-by-step protocol for the isolation and transient transformation of hornwort protoplasts. *Appl. Plant Sci.* **10**, e11456 (2022).
13. Waller, M. et al. An optimized transformation protocol for *Anthoceros agrestis* and three more hornwort species. *Plant J.* **114**, 699–718 (2023).
14. Lafferty, D. J. et al. Biolistics-mediated transformation of hornworts and its application to study pyrenoid protein localization. *J. Exp. Bot.* **75**, 4760–4771 (2024).
15. Chatterjee, P., Schafran, P., Li, F.-W. & Meeks, J. C. Nostoc talks back: temporal patterns of differential gene expression during establishment of *Anthoceros–Nostoc* symbiosis. *Mol. Plant. Microb. Interact.* **35**, MPMI05220101R (2022).
16. Li, F.-W., Villarreal Aguilar, J. C. & Szövényi, P. Hornworts: an overlooked window into carbon-concentrating mechanisms. *Trends Plant Sci.* **22**, 275–277 (2017).
17. Robison, T. A. et al. Hornworts reveal a spatial model for pyrenoid-based CO_2-concentrating mechanisms in land plants. *Nat. Plants* (in press).
18. Fritsch, R. Index to bryophyte chromosome counts. *Bryophyt. Bibl.* **40**, 1–352 (1991).
19. Villarreal, J. C. & Renner, S. S. Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc. Natl Acad. Sci. USA* **109**, 18873–18878 (2012).
20. Renzaglia, K. S., Villarreal Aguilar, J. C., Piatkowski, B. T., Lucas, J. R. & Merced, A. Hornwort stomata: architecture and fate shared with 400-million-year-old fossil plants without leaves. *Plant Physiol.* **174**, 788–797 (2017).
21. Villarreal, J. C. & Renner, S. S. Correlates of monoicy and dioicy in hornworts, the apparent sister group to vascular plants. *BMC Evol. Biol.* **13**, 239 (2013).
22. Camacho, J. P., Sharbel, T. F. & Beukeboom, L. W. B-chromosome evolution. *Philos. Trans. R. Soc. Lond. B* **355**, 163–178 (2000).
23. Proskauer, J. Studies on Anthocerotales V. *Phytomorphology* **7**, 113–135 (1957).
24. Bainard, J. D. & Villarreal, J. C. Genome size increases in recently diverged hornwort clades. *Genome* **56**, 431–435 (2013).
25. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
26. Hisanaga, T. et al. The ancestral chromatin landscape of land plants. *New Phytol.* **240**, 2085–2101 (2023).
27. Bewick, A. J. et al. On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl Acad. Sci. USA* **113**, 9111–9116 (2016).

28. Bi, G. et al. Near telomere-to-telomere genome of the model plant *Physcomitrium patens*. *Nat. Plants* **10**, 327–343 (2024).
29. Sequeira-Mendes, J. et al. The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **26**, 2351–2366 (2014).
30. Carey, S. B. et al. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. *Sci. Adv.* **7**, eabh2488 (2021).
31. Lang, D. et al. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533 (2018).
32. Diop, S. I. et al. A pseudomolecule-scale genome assembly of the liverwort *Marchantia polymorpha*. *Plant J.* **101**, 1378–1396 (2019).
33. Healey, A. L. et al. Newly identified sex chromosomes in the *Sphagnum* (peat moss) genome alter carbon sequestration and ecosystem dynamics. *Nat. Plants* **9**, 238–254 (2023).
34. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
35. Dong, S., Yu, J., Zhang, L., Goffinet, B. & Liu, Y. Phylotranscriptomics of liverworts: revisiting the backbone phylogeny and ancestral gene duplications. *Ann. Bot.* **130**, 951–964 (2022).
36. Dawes, T. N. et al. Extremely low genetic diversity in the European clade of the model bryophyte *Anthoceros agrestis*. *Plant Syst. Evol.* **306**, 49 (2020).
37. Martis, M. M. et al. Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *Proc. Natl Acad. Sci. USA* **109**, 13343–13346 (2012).
38. Waters, P. D. et al. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc. Natl Acad. Sci. USA* **118**, e2112494118 (2021).
39. Mcclung, C. E. The accessory chromosome—sex determinant? *Biol. Bull.* **3**, 43–84 (1902).
40. Rink, W. Zur Entwicklungsgeschichte, Physiologie und Genetik der Lebermoosgattungen Anthoceros und Aspiromitus Erschienen als Dissertation der Philosophischen Fakultät der Universität Würzburg. *Flora Allg. Bot. Ztg.* **130**, 87–130 (1935).
41. Montgomery, S. A. et al. Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. *Curr. Biol.* **30**, 573–588.e7 (2020).
42. Iwasaki, M. et al. Identification of the sex-determining factor in the liverwort *Marchantia polymorpha* reveals unique evolution of sex chromosomes in a haploid system. *Curr. Biol.* **31**, 5522–5532.e7 (2021).
43. Hisanaga, T. et al. A cis-acting bidirectional transcription switch controls sexual dimorphism in the liverwort. *EMBO J.* **38**, e100240 (2019).
44. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
45. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2021).
46. Gerke, P. et al. owards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort *Anthoceros agrestis*. *New Phytol.* **225**, 1974–1992 (2020).
47. Knoop, V. C-to-U and U-to-C: RNA editing in plant organelles and beyond. *J. Exp. Bot.* **74**, 2273–2294 (2023).
48. Villarreal, J. C. et al. Genome-wide organellar analyses from the hornwort Leiosporoceros dussii show low frequency of RNA editing. *PLoS ONE* **13**, e0200491 (2018).
49. Davies, K. M. et al. The evolution of flavonoid biosynthesis: a bryophyte perspective. *Front. Plant Sci.* **11**, 7 (2020).
50. Harris, B. J., Harrison, C. J., Hetherington, A. M. & Williams, T. A. Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. *Curr. Biol.* **30**, 2001–2012 (2020).
51. Crandall-Stotler, B. in *Advances in Bryology* (ed. Schultze-Motel, W.) Vol. 1, 315–398 (J. Cramer, 1981).
52. Fortin, J. P. & Friedman, W. E. A stomate by any other name? The open question of hornwort gametophytic pores, their homology, and implications for the evolution of stomates. https://doi.org/10.1111/nph.20094 (2024).
53. Caine, R. S. et al. An ancestral stomatal patterning module revealed in the non-vascular land plant *Physcomitrella patens*. *Development* **143**, 3306–3314 (2016).
54. Shani, E., Hedden, P. & Sun, T.-P. Highlights in gibberellin research: a tale of the dwarf and the slender. *Plant Physiol.* **195**, 111–134 (2024).
55. Hernández-García, J., Briones-Moreno, A., Dumas, R. & Blázquez, M. A. Origin of gibberellin-dependent transcriptional regulation by molecular exploitation of a transactivation domain in DELLA proteins. *Mol. Biol. Evol.* **36**, 908–918 (2019).
56. Yoshida, H. et al. Evolution and diversification of the plant gibberellin receptor GID1. *Proc. Natl Acad. Sci. USA* **115**, E7844–E7853 (2018).
57. Sun, R. et al. Biosynthesis of gibberellin-related compounds modulates far-red light responses in the liverwort *Marchantia polymorpha*. *Plant Cell* **35**, 4111–4132 (2023).
58. Miyazaki, S. et al. An ancestral gibberellin in a moss *Physcomitrella patens*. *Mol. Plant* **11**, 1097–1100 (2018).
59. Miyazaki, S., Kawaide, H. & Nakajima, M. Inactivation pathway of diterpenoid regulator in the moss *Physcomitrium patens*. *J. Plant Growth Regul.* **43**, 2937–2943 (2024).
60. Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R. & Abrams, S. R. Abscisic acid: emergence of a core signaling network. *Annu. Rev. Plant Biol.* **61**, 651–679 (2010).
61. Ma, Y. et al. Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324**, 1064–1068 (2009).
62. Sun, Y. et al. A ligand-independent origin of abscisic acid perception. *Proc. Natl Acad. Sci. USA* **116**, 24892–24899 (2019).
63. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
64. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
65. Xu, M. et al. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *Gigascience* **9**, giaa094 (2020).
66. Challis, R. J., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. BlobToolKit—interactive quality assessment of genome assemblies. *G3* **10**, 1361–1374 (2019).
67. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
68. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
69. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-4.0. *ISB* (2013–2015); http://www.repeatmasker.org
70. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
71. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
72. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinf.* **3**, lqaa108 (2021).

73. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
74. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
75. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
76. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
77. Ramírez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
78. Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).
79. Lovell, J. T. et al. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* **11**, e78526 (2022).
80. Zuntini, A. R. et al. Phylogenomics and the rise of the angiosperms. *Nature* **629**, 843–850 (2024).
81. Peñaloza-Bojacá, G. F. et al. Ancient reticulation and incomplete lineage sorting at the dawn of hornwort diversification and origin of the pyrenoid in the Carboniferous. Preprint at https://www.authorea.com/doi/full/10.22541/au.172168442.26511381/v1 (2024).
82. Zwaenepoel, A., & van de Peer, Y. wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* **35**, 2153–2155 (2019).
83. Klopfenstein, D. V. et al. GOATOOLS: a Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 10872 (2018).
84. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
85. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
86. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
87. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
88. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
89. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
90. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
91. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
92. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
93. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
94. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
95. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution, divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
96. Wickell, D. et al. Underwater CAM photosynthesis elucidated by *Isoetes* genome. *Nat. Commun.* **12**, 6348 (2021).
97. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
98. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
99. Morales, N. et al. Breedbase: a digital ecosystem for modern plant breeding. *G3* **12**, jkac078 (2022).
100. Diesh, C. et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol.* **24**, 74 (2023).
101. Schafran, P. & Li, F.-W. Pan-phylum genomes of hornworts. *Figshare* https://doi.org/10.6084/m9.figshare.26053480.v2 (2024).

Acknowledgements

We thank C. Rothfels, E. Sigel, C. Tribble, J. Meeks, P. Szövényi and B. Goffinet for sharing hornwort cultures and sporophytes used to establish axenic cultures, J. Hernández-García for discussion on GID1 evolution, D. Wickell for advice on WGD analysis and Z. Fei, E. Richards and Li lab members for commenting on the earlier paper version. This work is supported by National Science Foundation grants DEB-1831428, ISO-1923011 and MCB-2139576 to F.-W.L. and IOS-2109789 to P.S. J.C.V.A. acknowledges support from the Canada Research Chair (950-232698), and the Canadian Foundation for Innovation (projects 36781 and 39135) and N. S. Allen for help in logistics during fieldwork. K.M.D. and S.K. acknowledge financial support by the Marsden Fund of New Zealand/Te Pūtea Rangahau a Marsden (PAF2002) and a James Cook Research Fellowship (JCF-PAF2001). I.I. acknowledges support by the Spanish Ministry of Science and Innovation (MCIN) through the Ramón y Cajal programme (RyC2022-038245-I).

Author contributions

P.S. and F.-W.L. conceived the project. P.S., J.M.N., X.X., J.C.A.V. and F.-W.L. established the hornwort cultures. P.S., D.A.H., X.X. and F.-W.L. sequenced the genomes. P.S. assembled and annotated the genomes and conducted synteny and other comparative analyses. P.S. and I.S. identified the putative sex chromosomes. P.S. and L.A.M. built the HornwortBase. P.S., S.K., S.d.V., I.I., J.d.V. and K.D. analysed gene families. P.S. and F.-W.L. wrote the paper with contributions and comments from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-024-01883-w.

Correspondence and requests for materials should be addressed to Peter Schafran or Fay-Wei Li.

Peer review information *Nature Plants* thanks Pierre-Marc Delaux, Susanne Renner, Stefan Rensing and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Corresponding author(s): Peter Schafran, Fay-Wei Li

Last updated by author(s): Nov 2, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Guppy 6.4.2
Data analysis	porechop 0.2.4, Flye 2.8, Pilon 1.24, TGS-Gapcloser 1.1.1, BlobTools2, BLAST+ 2.10.0, BUSCO 5.2.1, Megalodon 2.5.0, bismark 0.24.1, EDTA 2.0.1, RepeatMasker 4.1.0, Tandem Repeats Finder 4.09.1, HISAT2 2.2.1, BRAKER 2.1.5, EggNOG mapper 2.1.9, DIAMOND 2.0.15, bedtools 2.29.2, HiCExplorer 3.7.2, pyGenomeTracks 3.8, GENESPACE 1.3, wgd 1.0, goatools 1.2.3, BWA 0.7.17, samblaster 0.1.26, samtools 1.18, Stringtie 2.1.1, OrthoFinder 2.5.4, Clustal Omega 1.2.3, TrimAl 1.4, IQ-TREE 2.0.3 and 1.5.5, FigTree 1.4.4, Geneious Prime 2023.0.4, CAFE 5, CafePlotter 0.2.0, pfam_scan 1.6, Breedbase sgn-389.0, Jbrowse 2.16, PhyML Geneious Plugin 2.2.4, MUSCLE (Geneious Prime 2023 implementation), MrBayes Geneious Plugin 2.2.4. Custom R and python codes and bioinformatic steps used in study can be found at pschafran.github.io.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw sequence data are deposited at NCBI SRA under BioProjects PRJNA996135 (See Supplementary Table 9 for individual SRA accessions). Previously published data utilized in this study are in BioProjects PRJNA574453 and PRJNA574424. Genome sequence and annotation files used in this study are v1.0 versions on the HornwortBase FTP page (https://hornwortbase.org/ftp/) and FigShare (https://doi.org/10.6084/m9.figshare.26053480.v1).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We created 11 new high quality genomes and annotations for species of hornworts spanning the phylum Anthocerotophyta. We used these genomes to infer structural evolution of their genomes over 300 Ma, and correlated genetic components associated with various traits.
Research sample	Single spores or tubers were used to create axenic culture lines for each species, propagated on modified Hatcher's medium.
Sampling strategy	Each genome represents a single individual. For population-level sequencing of dioicous species, at least 5 individuals were used to provide >95% chance of capturing individuals of both sexes, assuming equal germination and establishment rates of males and females. For RNA expression data in Notothylas, the number of biological replicates was limited by the number of sporophytes that could be collected.
Data collection	DNA was isolated from axenic cultures in all cases. For RNA used for gene prediction, whole, non-reproductive plants were sampled from the axenic cultures. For RNA expression data in Notothylas, RNA was isolated from separated gametophytic and sporophytic tissue from wild collected plants.
Timing and spatial scale	Spores were collected whenever sporophytes were mature, which varied among species and ecoregions. Sporophytes for gene expression analysis were collected while the sporophytes were developing, before reaching maturity.
Data exclusions	Contigs identified as bacterial contaminants were removed from the assemblies based on criteria described in the methods.
Reproducibility	Gene expression analysis of the sporophytes was performed with four biological replicates of both sporophytes and control

Reproducibility	(gametophyte) tissue. Phylogenetic inference was done with bootstrap replication.
Randomization	No experiments were subsampled. All biological replicates generated for sex chromosome identification and gene expression were included in each analysis.
Blinding	No blinding was performed.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	Mature sporophytes were collected from their respective locations, desiccated, and brought back to the lab where spores were germinated. Some spores and axenic cultures were provided by collaborators as noted in the acknowledgments.
Novel plant genotypes	This study did not generate any novel genotypes
Authentication	Species identifications were confirmed by experts in hornwort and bryophyte taxonomy (J.C. Villarreal and B. Goffinet)

Anthoceros agrestis (Bonn, Switzerland)
Anthoceros agrestis (Berwickshire, United Kingdom)
Anthoceros fusiformis (Butte County, California, United States)
Anthoceros punctatus (Humboldt County, California, United States)
Leiosporoceros dussii (Coclé, Panama)
Megaceros flagellaris (Honolulu County, Hawaii, United States)
Notothylas orbicularis (Tompkins County, New York, United States)
Paraphymatoceros pearsonii (Contra Costa County, California, United States)
Phaeoceros carolinianus (Louisiana, United States)
Phaeoceros sp. (Sarasota County, Florida, United States)
Phaeomegaceros chiloensis (Aysén, Chile)
Phymatoceros phymatodes (Humboldt County, California, United States)