nature genetics

Article

https://doi.org/10.1038/s41588-025-02325-9

Bryophytes hold a larger gene family space than vascular plants

Received: 11 March 2024

Accepted: 31 July 2025

Published online: 22 September 2025

Check for updates

Shanshan Dong[1,34], Sibo Wang[2,3,34], Linzhou Li[2,3,34], Jin Yu[2,3,34], Yongxia Zhang[4,34], Jia-Yu Xue[5,34], Hengchi Chen[6], Jianchao Ma[7], Yuying Zeng[2,3,8], Yuqing Cai[2,3,8], Wei Huang[2,3], Xuping Zhou[1], Jiayi Wu[1,5], Jianyou Li[1], Yifeng Yao[9], Ruoyang Hu[10], Tao Zhao[11], Juan Carlos Villarreal A[12], Leon Dirick[13], Li Liu[14], Michael Ignatov[15], Minghui Jin[16], Jue Ruan[16], Yikun He[10], Haifeng Wang[17], Bo Xu[9], Ricardo Rozzi[18,19], Jill Wegrzyn[20], Dennis William Stevenson[21], Karen S. Renzaglia[22], Hongfeng Chen[23], Li Zhang[1], Shouzhou Zhang[1], Roy Mackenzie[19,24], Javier E. Moreno[25], Michael Melkonian[26], Tong Wei[2,3], Ying Gu[2], Xun Xu[2], Stefan A. Rensing[27], Jinling Huang[7,28,29], Manyuan Long[30], Bernard Goffinet[20], John L. Bowman[31,35] ✉, Yves Van de Peer[5,6,32,35] ✉, Huan Liu[2,33,35] ✉ & Yang Liu[1,2,3,35] ✉

After 500 million years of evolution, extant land plants compose the following two sister groups: the bryophytes and the vascular plants. Despite their small size and simple structure, bryophytes thrive in a wide variety of habitats, including extreme conditions. However, the genetic basis for their ecological adaptability and long-term survival is not well understood. A comprehensive super-pangenome analysis, incorporating 123 newly sequenced bryophyte genomes, reveals that bryophytes possess a substantially greater diversity of gene families than vascular plants. This includes a higher number of unique and lineage-specific gene families, originating from extensive new gene formation and continuous horizontal transfer of microbial genes over their long evolutionary history. The evolution of bryophytes' rich and diverse genetic toolkit, which includes new physiological innovations like unique immune receptors, likely facilitated their spread across different biomes. These newly sequenced bryophyte genomes offer a valuable resource for exploring alternative evolutionary strategies for terrestrial success.

The colonization of land by plants marks a pivotal moment in the evolution of life on the Earth[1,2], ultimately giving rise to the complex modern-day terrestrial ecosystem[3]. Today, approximately 400,000 species of land plants (embryophytes) compose the following two major groups: bryophytes (that is, hornworts, liverworts and mosses) and vascular plants (tracheophytes)[4], with the latter being distinguished by their branched sporophyte and lignified water-conducting cells. The phylogenetic relationships between bryophytes and tracheophytes, as well as among bryophytes themselves, were subjects of debate, until recent analyses of transcriptomic and genomic data resolved bryophytes as a monophyletic group, which is sister to all living tracheophytes[5–9]. These findings suggest that extant land plants emerged from two distinct evolutionary lines that diverged around 500 million years ago (Mya). Bryophytes and vascular plants differ primarily in their dominant vegetative generation, with bryophytes dominated by the gametophyte and vascular plants dominated by the

A full list of affiliations appears at the end of the paper. ✉e-mail: john.bowman@monash.edu; yves.vandepeer@psb.ugent.be; liuhuan@genomics.cn; liuyang@szbg.ac.cn

sporophyte. Consequently, their diversification may be driven primarily, though not exclusively, by selection on traits associated with these respective generations[7]. However, whether the processes of genome evolution underlying these diversifications are similar or not have remained largely unexplored, primarily due to the limited availability of genomic data from a diverse range of bryophytes.

Bryophytes comprise approximately 22,000 species worldwide and exhibit a wide distribution, often thriving and even dominating in extreme habitats[10,11]. Their structure is rather simple, as they lack lignified water-conducting cells, roots[12] and their sporophyte remains unbranched and reliant on the maternal gametophyte for the nourishment[13]. The widespread ecological success of bryophytes may be attributed to their unique and divergent physiological adaptations, such as dehydration and desiccation tolerance, which differ substantially from those of tracheophytes[14], and may be concordant with the surprisingly divergent genomes[15–19]. Whether such genomic diversity is universal among bryophytes remains to be tested. Indeed, genomes remain to be assembled for the majority of bryophyte orders, which account for most of the approximately 70 land plant orders still lacking genomic information[20]. To date, only 8 of the 55 bryophyte orders have sequenced genomes, highlighting a substantial knowledge gap in furthering our understanding of genome evolution in land plants[21]. This study bridges this gap by presenting 123 newly sequenced bryophyte genomes, enabling a comprehensive 'pangenome' analysis of bryophytes. Our findings also uncover divergent gene family spaces and distinct evolutionary processes that have shaped the genomes of bryophytes and vascular plants.

Results

Bryophyte super-pangenomes: core, accessory, unique genes

We newly sequenced, assembled and annotated 123 high-quality and highly complete bryophyte genomes, and compiled these in a centralized platform (www.bryogenomes.org). The sampling includes 37 liverworts, 82 mosses and 4 hornworts, representing 47 of the 55 known bryophyte orders (Fig. 1, Extended Data Fig. 1 and Supplementary Data 1). The genomes of bryophytes are relatively small compared to other land plants, and hold on average fewer genes (that is, 27,959 overall, or specifically, 19,267 for hornworts, 27,768 for liverworts and 28,701 for mosses; Supplementary Data 2 and 3 and Supplementary Note 1) than those of vascular plants (that is, 34,794 genes). Our phylogenetic analyses place bryophytes as a sister group to extant vascular plants, with relationships within bryophyte lineages largely aligning with recent phylogenomic studies[4,8,22–24] (Fig. 1, Extended Data Fig. 2 and Supplementary Data 1 and 4), providing a robust framework for tracing the evolution of bryophyte genomes (Supplementary Notes 1 and 2).

The proteomes of 343 Archaeplastida species (138 bryophytes, 146 tracheophytes and 59 algae; Supplementary Data 5) were classified into 1,113,359 orthogroups (gene families). Bryophytes exhibit a considerably larger cumulative number of nonredundant gene families compared to vascular plants (that is, 637,597 versus 373,581; Fig. 2a), despite slightly larger taxon sampling for the latter. Bryophytes also have a higher average number of gene families unique to a single taxon (that is, 3,862 versus 2,223; Fig. 2a and Supplementary Data 6), and more accessory gene families, that is, those present in at least two but fewer than 80% of all samples of a lineage (4,021 versus 1,583), although they possess fewer (6,233 versus 6,647) core gene families (that is, those present in ≥80% of samples of a lineage). Whereas bryophytes have a lower percentage of unique or 'orphan' gene families (that is, 84% versus 87%; Fig. 2a), their absolute number far exceeds that of vascular plants (532,840 versus 324,552). Collectively, the total number of unique and accessory gene families in a single bryophyte genome averages 7,883 (56% of the total number of gene families) versus only 3,806 (36%) in vascular plants.

Bryophyte genomes show substantial gene family diversity, with a sharp increase in the number of orthogroups as more genomes are added (Fig. 2b, bottom). This contrasts with the stabilization of core gene families after sampling 30 genomes. Tracheophyte genomes exhibit similar patterns, though with a less steep accumulation curve (Fig. 2b), possibly reflecting higher rates of gene innovation in bryophytes or the result of deeper phylogenetic sampling of bryophytes (Extended Data Fig. 3). These findings suggest that, despite extensive sampling, many genes and gene families in land plants remain undiscovered.

The estimated numbers of gene family gains and losses in land plants (Extended Data Fig. 3 and Supplementary Data 7) are much higher than previously suggested[7,25], indicating the profound impact of sampling size on reconstructing gene family evolution (Supplementary Note 3). Bryophytes show a long history of gene family innovation, especially notable in mosses (for example, Bryidae; Fig. 2c and Extended Data Fig. 3) since the early Cretaceous (~100 Mya), a phenomenon perhaps linked to the cumulative effects of successive whole-genome duplications (WGDs) in mosses (Supplementary Fig. 1 and Supplementary Note 5). This contrasts with tracheophytes, wherein ancestral diversity of gene families is lower in lineages that arose over the last 65 million years and which are characterized by constant and small numbers of total gene families (Fig. 2c and Extended Data Fig. 3), possibly due to mass extinctions during the Cretaceous–Paleogene boundary[26]. Over time, bryophyte ancestral nodes are estimated to hold more gene family diversity than those of vascular plants (Fig. 2c), underscoring deep-rooted gene family innovation processes in bryophytes.

Accessory and unique gene families in bryophytes are integral to their genomes, with a guanine-cytosine (GC) content similar to core genes (that is, ~50%; Extended Data Fig. 4d and Supplementary Fig. 2) and a substantial proportion (50–80%) are expressed (Extended Data Fig. 4b and Supplementary Fig. 3), including in model species like *Physcomitrium patens* and *Marchantia polymorpha* (Extended Data Fig. 4a), and, further, a considerable fraction of them are differentially expressed under various stresses in *P. patens* (Extended Data Fig. 4c). However, in bryophytes, the function of most accessory and unique gene families remains ambiguous, with only a small fraction (27% and 16%) functionally annotated based on protein domains, compared to 91% in core families (Extended Data Fig. 4e). This highlights the substantial knowledge gap in understanding the functional roles of genes in bryophytes.

The de novo origination of orphan genes in bryophytes

In all sampled bryophytes (Supplementary Fig. 2), including the model species *P. patens* and *M. polymorpha* (Extended Data Fig. 4f,g), genes within accessory and unique orthogroups contain fewer introns and possess shorter coding regions compared to those within core orthogroups. These characteristics are also observed in newly evolved or 'young' genes in angiosperms such as *Oryza*[27] and *Arabidopsis*[28,29], suggesting that the majority of accessory and unique gene families in bryophytes may have originated recently. Additionally, genes belonging to unique gene families in *P. patens* and *M. polymorpha* are generally expressed at lower levels than those in accessory and core families (Extended Data Fig. 4h and Supplementary Fig. 3), a pattern consistent with observations of new genes in rice[27]. It is likely that most unique orthogroups in bryophytes arose as orphan genes through various molecular mechanisms, including rapid sequence evolution and de novo gene origination[27,30]. In *P. patens* and *M. polymorpha*, less than 15% of unique genes show some sequence similarity to genes in other orthogroups (Extended Data Fig. 4i), suggesting their derivation following gene duplication. Conversely, the 85% that exhibit no sequence similarity to existing orthogroups likely arose through different mechanisms, predominantly de novo from noncoding regions[31]. Identifying ancestral noncoding sequences poses challenges due to rapid sequence divergence[32] and the scarcity of genomic resources for closely related taxa. In our study, we selected *M. polymorpha* ssp. *ruderalis* for comparison with two conspecific subspecies and its sister species, all of

https://doi.org/10.1038/s41588-025-02325-9

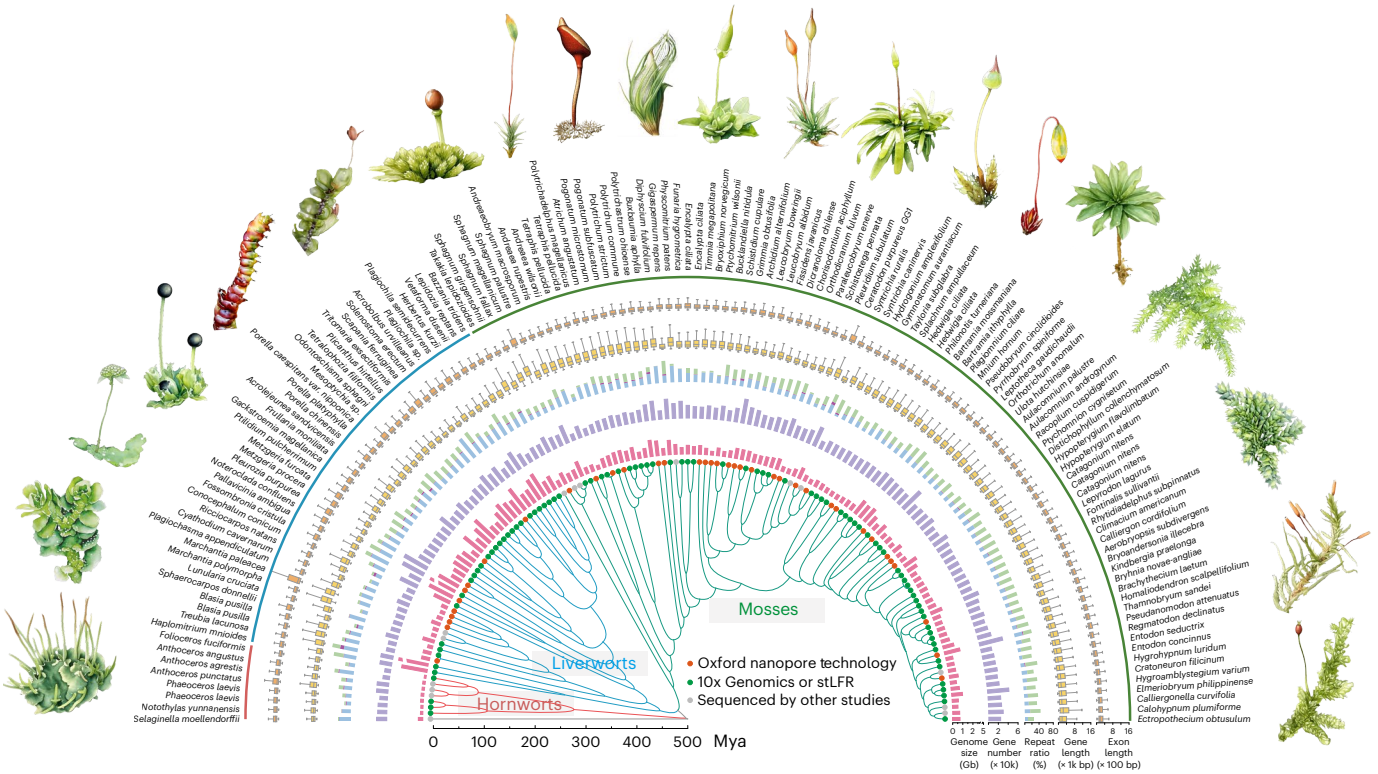


Fig. 1 | General features of the 138 bryophyte genomes analyzed. The phylogeny is based on a 190-taxon (see details in Supplementary Data 1–3) phylogenetic analysis of green plants (Extended Data Fig. 2) with divergent time based on the molecular dating analysis of bryophytes (Supplementary Fig. 19). The branches and species names are color-highlighted by the three major bryophyte groups. From the inside to outside, the five channels illustrated are genome size, gene number, repeat ratio (percentage of repeat content in the genomes, with blue, purple and green bars representing LTR, LINE and unknown repeats, respectively), gene length and exon length. The bryophyte images (not to scale) shown from left to right include the hornwort *Phaeoceros laevis*[1]; the liverworts *Apotreubia nana*[2], *Asterella syngenesica*[2], *Petalophyllum ralfsii*[2], *Pleurozia purpurea*[2] and *Scapania bolanderi*[2]; and the mosses *Sphagnum fallax*[2], *Polytrichum* sp.[1], *Buxbaumia aphylla*[3], *Diphyscium mucronifolium*[2], *Encalypta ciliata*[2], *Ptychomitrium* sp.[4], *Octoblepharum albidum*[2], *Rhacocarpus purpurascens*[2], *Imbribryum miniatum*[2], *Rhodobryum roseum*[1], *Racopilum cf. capens*[2], *Hypopterygium flavolimbatum*[2], *Antitrichia californica*[2] and *Exsertotheca intermedia*[2]. The illustrations were created based on photographs by S. S. Dong (*P. laevis*, *Polytrichum* sp. and *R. roseum*), D. Callaghan (*A. nana*, *A. syngenesica*, *P. ralfsii*, *P. purpurea*, *S. bolanderi*, *S. fallax*, *D. mucronifolium*, *E. ciliata*, *O. albidum*, *R. purpurascens*, *I. miniatum*, *Racopilum cf. capens*, *H. flavolimbatum*, *A. californica* and *E. intermedia*), Y. Liu (*B. aphylla*) and L. Zhang (*Ptychomitrium* sp.). LINE, long interspersed nuclear element.

which diverged within the last 5–7 million years[33,34]. Approximately 70–80% (3,120–3,583; Extended Data Fig. 4j) of genes in orphan gene families of *M. polymorpha* ssp. *ruderalis* aligns with noncoding regions in closely related species, indicating a likely origin from these regions (Extended Data Fig. 4j).

De novo gene origination, previously documented only in angiosperms such as rice[27] and bamboo[30], arises through gradual accumulation of mutations. For instance, the evolutionary trajectory of gene birth from proto-genes[35] derived from noncoding regions can be reconstructed for several unique genes in *M. polymorpha* ssp. *ruderalis*, demonstrating the successive elimination or conversion of stop codons in proto-genes (Supplementary Fig. 4). For example, the orphan gene *Mp7g02380* in *M. polymorpha* ssp. *ruderalis* shows increasing sequence similarity to orthologous sequences in noncoding regions of related *Marchantia* subspecies, with a corresponding decrease in the number of stop codons. Similarly, the orphan gene *Mp4g17190* may have originated from noncoding regions with high similarity across *Marchantia* subspecies, where stop codons were converted to sense codons through single-base changes. Among mosses, particularly the Bryidae, about 26% of unique genes exhibit a high degree of similarity to noncoding regions in at least two other species (Supplementary Fig. 5, with exemplar alignments). This represents the largest set of de novo gene candidates proposed for any major plant lineage, totaling 36,481 genes. These findings not only demonstrate the occurrence of de novo gene innovation outside angiosperms but also underscore its significance in bryophytes. This mechanism[36] has a crucial role in driving the adaptations and ecological diversification of bryophytes, highlighting the dynamic nature of their genomes and the evolutionary processes shaping them.

Frequent horizontal gene transfer (HGT) shapes bryophyte evolution

The genomes of 49 streptophyte species encompass 1,809 orthologous gene families with origins in prokaryotes, viruses, fungi or animals, and of these 1,306 are newly identified ones (Supplementary Data 8–11, Fig. 3 and Supplementary Figs. 6–13). The most substantial burst of HGT occurred in the ancestral streptophyte, with 115 HGT events, followed by 90 HGT events in the ancestral embryophyte (Fig. 3a). On average, individual bryophyte species have acquired more horizontally transferred genes than tracheophytes (229 versus 163; Fig. 3b), and our thorough phylogenetic analysis of bryophyte genomes indicates a higher number of species-specific HGT events (Fig. 3b), as also recently revealed for a relative of *P. patens*[37]. This suggests that HGT may have been a consistent feature throughout the bryophyte evolution, unlike the pattern observed during the diversification of vascular plants (Fig. 3a). HGT genes in Viridiplantae tend to have fewer introns and shorter lengths compared to core genes, reflecting the genes of their largely bacterial or fungal donor organisms[36] (Supplementary Fig. 14). In bryophytes, HGT-acquired genes are often found in regions with a higher density of long terminal repeats (LTRs) than core genes,

https://doi.org/10.1038/s41588-025-02325-9

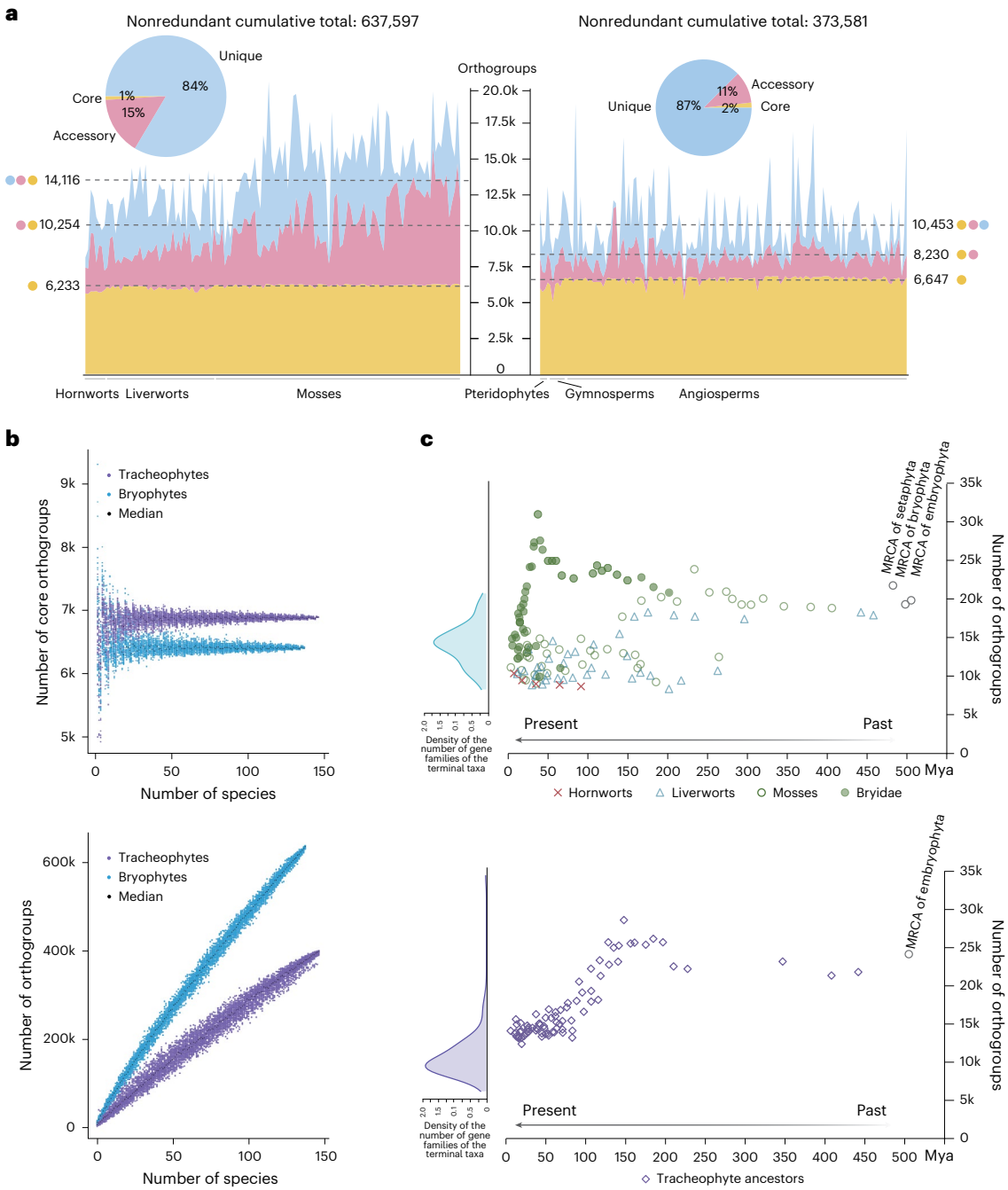


Fig. 2 | Pangenome analyses of bryophytes and tracheophytes. a, The number of gene families of each composition (core, accessory and unique) in individual genomes of 138 bryophytes (left) and 146 tracheophytes (right). The taxa on *x* axis are ordered based on the phylogeny (Extended Data Fig. 3). The added-up average number of the three types of gene families are marked by dash lines. The pie charts show the total number of three types of gene families of all sampled bryophytes and tracheophytes. **b**, Simulations of the decrease of core genome size (top), and the increase of the pangenome size (bottom) in bryophytes (blue dot) and tracheophytes (purple dot). In total, 100 random combinations were used for each given number of species. **c**, Dot plot indicates the total number of gene families (*y* axis) reconstructed at the ancestral nodes, dated based on Extended Data Fig. 3 (*x* axis), across the bryophyte (top), and tracheophyte (bottom) phylogeny. MRCA, most recent common ancestor.

implying that LTR elements facilitate the integration or proliferation of HGT genes within their genomes (Supplementary Fig. 14).

A substantial portion of the genes that bryophytes have acquired through HGT are of unknown function, with the remainder having roles in replication, carbohydrate metabolism, amino acid metabolism and secondary metabolism (Supplementary Data 12). In the moss *P. patens*, a large fraction of HGT genes were differentially expressed under heat, drought and UV stress, with 151 (60%), 148 (59%) and 190 (76%) HGT-derived orthogroups exhibiting altered expression, respectively (Supplementary Data 13–22). In total, expression of 233 HGT orthogroups was associated with abiotic stress responses, accounting for 93% of all HGT orthogroups in *P. patens*, implying they may enhance bryophyte ecological adaptability in diverse environments. Other HGT genes have been previously described as stress responsive. For example, the bacterial *HPA3* gene, acquired by mosses, encodes an N-acetyltransferase, detoxifying toxic D-amino acids through N-acetylation, a crucial adaptation for cellular homeostasis and survival in nutrient-deprived or stressed ecosystems[38,39]. Finally, the fungal *RDS1*

https://doi.org/10.1038/s41588-025-02325-9

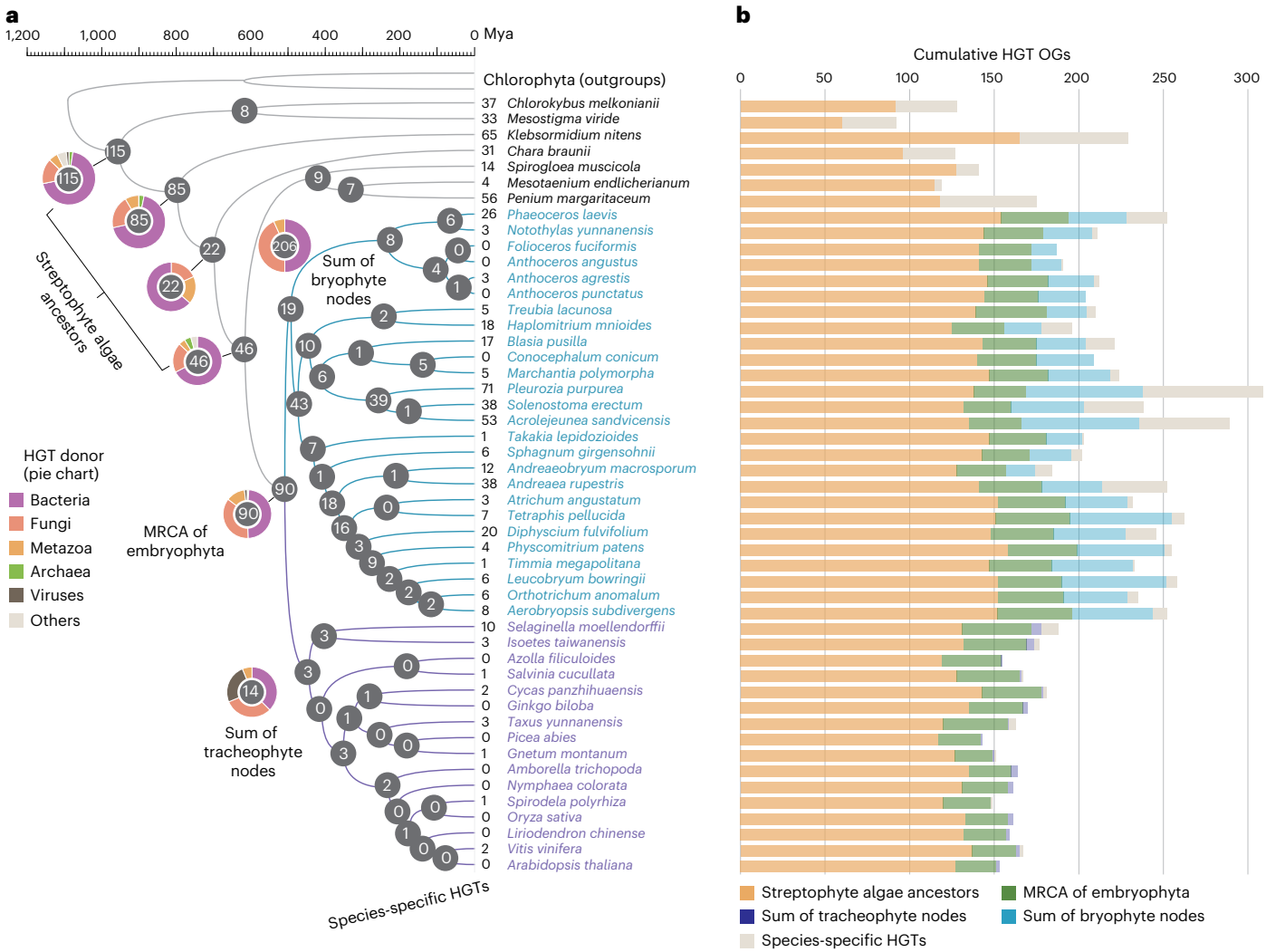


Fig. 3 | The landscape of HGTs in streptophytes. a, A summary of the number of HGT events (Supplementary Data 10) on the chronogram of streptophytes (bryophytes in blue and tracheophytes in purple), with the numbers in the gray disks indicating total HGTs reconstructed at a node. The pie chart illustrates the distribution of putative donors to HGT among major categories, at select ancestral nodes and for specific clades, that is, including numbers from all internal bryophyte or tracheophyte nodes. Number next to terminal branch refers to HGT unique to the terminal taxon in our dataset. **b**, Bar chart of the cumulative numbers of HGT orthogroups in each sampled species, with each bar broken down into the proportion of gene families acquired from ancestral HGTs. Sum of tracheophyte nodes = summary of all tracheophyte internal nodes. Sum of bryophyte nodes = summary of all bryophyte internal nodes. OGs, orthologous groups.

gene acquired by liverworts, potentially linked to stress responses[40], underscores the contribution of HGT-derived genes to plants' environmental resilience. Collectively, these genes highlight the pivotal role of HGT in shaping the adaptive capacities of bryophytes, enabling them to thrive in challenging ecological conditions.

Newly identified HGT genes in this study may also have various adaptive roles in interactions with other organisms. For example, the bacteria-derived *LEC* gene in mosses, encoding a mannose-specific lectin, exhibits antifungal activity against *Alternaria alternata* and *Colletotrichum* species[41], underscoring its potential role in pathogen defense[42]. Similarly, the fungal-derived *YANB* gene in liverworts, encoding a decarboxylase involved in yanuthone D biosynthesis, provides antibiotic activity against fungi and bacteria, and may protect bryophytes from microbial competitors[43]. Another fungal-derived gene in liverworts, *EUPF*, biosynthesizes eupenifeldin, a secondary metabolite with broad-spectrum antifungal, antimalarial and anthelmintic properties[44], likely enhancing bryophyte survival under diverse ecological pressures.

HGT-acquired genes may have critical roles in bryophytes' defense against herbivores. For example, in *M. polymorpha*, HGT-acquired terpene synthases (TPSs) are suggested to act mainly as antiherbivory agents[45,46]. Additionally, the *FBT* gene, encoding a protein with an FB lectin domain (PF07367), acquired from fungi and found in many liverworts and mosses, has an insecticidal effect (Supplementary Fig. 15). The protein synthesized by *M. polymorpha* causes 80% mortality in individuals of the moth *Plutella xylostella* injected with 5.41 µg FB lectin[47]. The antiherbivory function of this gene is strengthened here, even at a lower dose (0.45 µg mg^{-1}, corresponding to 2.86 µg per individual) of the FB lectin derived from the leafy liverwort *Acrolejeunea sandvicensis*, causing 97.62% mortality in larvae of the cotton bollworm *Helicoverpa armigera* (Extended Data Fig. 5b and Supplementary Data 23). This potency approaches that of the Bt toxin secreted by *Bacillus thuringiensis*[48]. The FB lectin-containing diet also substantially reduces the growth rate of the fall armyworm, *Spodoptera frugiperda* (Extended Data Fig. 5c and Supplementary Data 24). Thus, bryophytes possess a diverse array of gene families originating from microbial HGT, many of which are unique (Fig. 3), and offer a vast pool of genes for screening potential resistances to bacterial or fungal diseases (for example, *KP4* (ref. 49)), or for exploring their roles in plant development[50] and

plant–fungi interactions[51]. This highlights the importance of HGT in the evolution and ecological diversification of bryophytes.

Discussion

Bryophytes exhibit a substantially larger number of unique and accessory gene families compared to vascular plants (Fig. 2), with these families emerging throughout their evolutionary history (Fig. 2). Although bryophyte genomes contain a greater diversity of gene families (Fig. 2), they possess, on average, fewer genes than those of vascular plants (that is, 27,937 versus 34,813), where gene families expand primarily through gene duplications (Supplementary Fig. 16)[52]. Unique gene families in bryophytes may originate from de novo genes[53], a process of gene innovation previously documented in flowering plants such as rice and bamboo[27,30] and now also demonstrated in bryophytes. De novo origination appears to be a crucial mechanism for the diversification of bryophyte genomes. Our comprehensive genomic sampling across the plant kingdom substantially expands the phylogenetic distribution of de novo genes, highlighting the vital role of de novo gene origination in plant gene evolution and functional innovation. The patterns in new gene diversity imply their continuous emergence and fixation throughout bryophyte evolution. Fixation of 'new' genes may be more efficient in bryophytes than in tracheophytes[54], as these genes are immediately subjected to selection in the haploid vegetative body of bryophytes. Additionally, the totipotent nature of bryophyte gametophyte tissue may further enable the propagation of genetic newness, thus facilitating the fixation of new genes.

Many new functions in plants have been acquired through HGT, and our dense sampling indicates that HGT is not a sporadic but a continuous source of new genes in plants, particularly in bryophytes. Notably, nearly 70% of HGT gene families in bryophytes are categorized as accessory or unique gene families (Supplementary Data 25), and most sampled taxa carry unique HGTs (Fig. 3). Bryophytes show a differential accumulation of HGTs compared to seed plants, suggesting a possible link between the frequency of HGTs and the likelihood of germ line cells' direct exposure to microbes[55].

Given the fundamentally simple architecture of bryophytes, the diversification of their genetic repertoire, leading to a vast number of accessory and unique gene families (Fig. 2a), may have been driven by the selection for new ecophysiological adaptations and their optimization[14]. In *P. patens*, for instance, 63% of accessory and 27% of unique gene families respond to abiotic stresses such as UV, heat and drought (Extended Data Fig. 4c). The *PpARDT* gene, involved in drought tolerance[56], was initially identified as an 'orphan' gene in *P. patens* but found to be common among more mosses in this study. Other physiological innovations may include adaptations to ultracold environments, enabling bryophytes to thrive in polar ecosystems[57]. Bryophytes are also distinguished by their diverse secondary metabolites, particularly terpenoids[58], encoded by a rich assortment of *TPS* genes, notably the *TPS*-c type (Supplementary Fig. 17). More than 70% of bryophyte *TPS* genes belong to the accessory and unique gene families (Supplementary Data 25), and have been acquired through HGT multiple times independently across various land plant lineages[59]. In bryophytes, terpenoids are involved in numerous critical physiological processes, including allelopathy[60], herbivore deterrence[46,61], pathogen defense[62] and hormonal activities[58]. Additionally, the diversity and variance in immune receptors (Extended Data Fig. 6, Supplementary Note 4 and Supplementary Fig. 18) between bryophytes and angiosperms also suggest distinct evolutionary defense strategies in bryophytes.

In conclusion, our comprehensive analyses of genomes from across the land plant tree of life indicated that the evolution and diversification of bryophyte genomes—unlike those of vascular plants—are driven in substantial part by ongoing de novo gene innovation and the acquisition of microbial genes. Although bryophyte genomes, on average, contain fewer genes, these genes belong to a wider variety of functional gene families. This genetic diversity likely underpins the resilience of these structurally simple and haploid dominant organisms. It may also explain their enduring presence on land over the past 500 million years, their widespread ecological distribution, and their success in colonizing extreme environments worldwide.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-025-02325-9.

References

1. Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. *Nature* **389**, 33–39 (1997).
2. Bowman, J. L. The origin of a land flora. *Nat. Plants* **8**, 1352–1369 (2022).
3. Graham, L. E., Cook, M. E. & Busse, J. S. The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proc. Natl Acad. Sci. USA* **97**, 4535–4540 (2000).
4. Puttick, M. N. et al. The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr. Biol.* **28**, 733–745 (2018).
5. Leebens-Mack, J. H. et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
6. Su, D. et al. Large-scale phylogenomic analyses reveal the monophyly of bryophytes and neoproterozoic origin of land plants. *Mol. Biol. Evol.* **38**, 3332–3344 (2021).
7. 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).
8. 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).
9. Sousa, F., Foster, P. G., Donoghue, P. C. J., Schneider, H. & Cox, C. J. Nuclear protein phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). *New Phytol.* **222**, 565–575 (2019).
10. Christenhusz, M. & Byng, J. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201–217 (2016).
11. Degola, F., Sanità di Toppi, L. & Petraglia, A. Bryophytes: how to conquer an alien planet and live happily (ever after). *J. Exp. Bot.* **73**, 4267–4272 (2022).
12. Ligrone, R., Duckett, J. G. & Renzaglia, K. S. Major transitions in the evolution of early land plants: a bryological perspective. *Ann. Bot.* **109**, 851–871 (2012).
13. Shaw, J. & Renzaglia, K. Phylogeny and diversification of bryophytes. *Am. J. Bot.* **91**, 1557–1581 (2004).
14. Glime, J. M. Primitive or advanced? *Bryological* **55**, 5–7 (1990).
15. Bowman, J. L. et al. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287–304 (2017).
16. Rensing, S. A. et al. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64–69 (2008).
17. Zhang, J. et al. The hornwort genome and early land plant evolution. *Nat. Plants* **6**, 107–118 (2020).
18. 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).
19. Hu, R. et al. Adaptive evolution of the enigmatic *Takakia* now facing climate change in Tibet. *Cell* **186**, 3558–3576 (2023).

20. Marks, R. A., Hotaling, S., Frandsen, P. B. & VanBuren, R. Representation and participation across 20 years of plant genome sequencing. *Nat. Plants* **7**, 1571–1578 (2021).
21. Szövényi, P., Gunadi, A. & Li, F. W. Charting the genomic landscape of seed-free plants. *Nat. Plants* **7**, 554–565 (2021).
22. Renzaglia, K. S., Villarreal Aguilar, J. C. & Garbary, D. J. Morphology supports the setaphyte hypothesis: mosses plus liverworts form a natural group. *Bryophyt. Divers. Evol.* **40**, 11–17 (2018).
23. Frangedakis, E. et al. The hornworts: morphology, evolution and development. *New Phytol.* **229**, 735–754 (2021).
24. Liu, Y. et al. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nat. Commun.* **10**, 1485 (2019).
25. Bowles, A. M. C., Bechtold, U. & Paps, J. The origin of land plants is rooted in two bursts of genomic novelty. *Curr. Biol.* **30**, 530–536 (2020).
26. McElwain, J. C. & Punyasena, S. W. Mass extinction events and the plant fossil record. *Trends Ecol. Evol.* **22**, 548–557 (2007).
27. Zhang, L. et al. Rapid evolution of protein diversity by de novo origination in *Oryza*. *Nat. Ecol. Evol.* **3**, 679–690 (2019).
28. Cui, X. et al. Young genes out of the male: an insight from evolutionary age analysis of the pollen transcriptome. *Mol. Plant* **8**, 935–945 (2015).
29. Wu, D. D. et al. 'Out of pollen' hypothesis for origin of new genes in flowering plants: study from *Arabidopsis thaliana*. *Genome Biol. Evol.* **6**, 2822–2829 (2014).
30. Jin, G. et al. New genes interacted with recent whole-genome duplicates in the fast stem growth of bamboos. *Mol. Biol. Evol.* **38**, 5752–5768 (2021).
31. Long, M., VanKuren, N. W., Chen, S. & Vibranovski, M. D. New gene evolution: little did we know. *Annu. Rev. Genet.* **47**, 307–333 (2013).
32. Xu, X. et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* **30**, 105–111 (2011).
33. Villarreal, A. J., Crandall-Stotler, B. J., Hart, M. L., Long, D. G. & Forrest, L. L. Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. *New Phytol.* **209**, 1734–1746 (2016).
34. Linde, A. M., Eklund, D. M., Cronberg, N., Bowman, J. L. & Lagercrantz, U. Rates and patterns of molecular evolution in bryophyte genomes, with focus on complex thalloid liverworts, Marchantiopsida. *Mol. Phylogenet. Evol.* **165**, 107295 (2021).
35. Carvunis, A. R. et al. Proto-genes and de novo gene birth. *Nature* **487**, 370–374 (2012).
36. Xia, S., Chen, J., Arsala, D., Emerson, J. J. & Long, M. Functional innovation through new genes as a general evolutionary process. *Nat. Genet.* **57**, 295–309 (2025).
37. Vuruputoor, V. S. et al. Crossroads of assembling a moss genome: navigating contaminants and horizontal gene transfer in the moss *Physcomitrellopsis africana*. *G3 (Bethesda)* **14**, jkae104 (2024).
38. Kolukisaoglu, Ü. *D*-amino acids in plants: sources, metabolism, and functions. *Int. J. Mol. Sci.* **21**, 5421 (2020).
39. Yow, G.-Y., Uo, T., Yoshimura, T. & Esaki, N. Physiological role of D-amino acid-N-acetyltransferase of *Saccharomyces cerevisiae*: detoxification of D-amino acids. *Arch. Microbiol.* **185**, 39–46 (2006).
40. Ludin, K. M., Hilti, N. & Schweingruber, M. E. *Schizosaccharomyces pombe rds1*, an adenine-repressible gene regulated by glucose, ammonium, phosphate, carbon dioxide and temperature. *Mol. Gen. Genet* **248**, 439–445 (1995).
41. Sudmoon, R., Sattayasai, N., Bunyatratchata, W., Chaveerach, A. & Nuchadomrong, S. Thermostable mannose-binding lectin from *Dendrobium findleyanum* with activities dependent on sulfhydryl content. *Acta Biochim. Biophys. Sin. (Shanghai)* **40**, 811–818 (2008).
42. Sattayasai, N. et al. *Dendrobium findleyanum agglutinin*: production, localization, anti-fungal activity and gene characterization. *Plant Cell Rep.* **28**, 1243–1252 (2009).
43. Holm, D. K. et al. Molecular and chemical characterization of the biosynthesis of the 6-MSA-derived meroterpenoid yanuthone D in *Aspergillus niger*. *Chem. Biol.* **21**, 519–529 (2014).
44. Ayers, S. et al. Noreupenifeldin, a tropolone from an unidentified ascomycete. *J. Nat. Prod.* **71**, 457–459 (2008).
45. Romani, F. et al. Oil body formation in *Marchantia polymorpha* is controlled by *MpC1HDZ* and serves as a defense against arthropod herbivores. *Curr. Biol.* **30**, 2815–2828 (2020).
46. Kanazawa, T. et al. The liverwort oil body is formed by redirection of the secretory pathway. *Nat. Commun.* **11**, 6152 (2020).
47. Ma, J. et al. Major episodes of horizontal gene transfer drove the evolution of land plants. *Mol. Plant* **15**, 857–871 (2022).
48. Wu, K., Guo, Y. & Head, G. Resistance monitoring of *Helicoverpa armigera* (Lepidoptera: Noctuidae) to bt insecticidal protein during 2001–2004 in China. *J. Econ. Entomol.* **99**, 893–898 (2006).
49. Gage, M. J., Bruenn, J., Fischer, M., Sanders, D. & Smith, T. J. *KP4* fungal toxin inhibits growth in *Ustilago maydis* by blocking calcium uptake. *Mol. Microbiol.* **41**, 775–785 (2001).
50. Guan, Y. et al. Horizontally acquired fungal killer protein genes affect cell development in mosses. *Plant J.* **113**, 665–676 (2023).
51. Sun, G. et al. Are fungi-derived genomic regions related to antagonism towards fungi in mosses? *New Phytol.* **228**, 1169–1175 (2020).
52. Qiao, X. et al. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* **20**, 38 (2019).
53. McLysaght, A. & Hurst, L. D. Open questions in the study of de novo genes: what, how and why. *Nat. Rev. Genet.* **17**, 567–578 (2016).
54. Kondrashov, A. S. & Crow, J. F. Haploidy or diploidy: which is better? *Nature* **351**, 314–315 (1991).
55. Huang, J. Horizontal gene transfer in eukaryotes: the weak-link model. *Bioessays* **35**, 868–875 (2013).
56. Dong, X. M. et al. Orphan gene *PpARDT* positively involved in drought tolerance potentially by enhancing ABA response in *Physcomitrium* (*Physcomitrella*) *patens*. *Plant Sci.* **319**, 111222 (2022).
57. Liu, S. et al. The Antarctic moss *Pohlia nutans* genome provides insights into the evolution of bryophytes and the adaptation to extreme terrestrial habitats. *Front. Plant Sci.* **13**, 920138 (2022).
58. Chen, F. et al. Terpenoid secondary metabolites in bryophytes: chemical diversity, biosynthesis and biological functions. *Crit. Rev. Plant Sci.* **37**, 210–231 (2018).
59. Jia, Q., Kollner, T. G., Gershenzon, J. & Chen, F. *MTPSL*s: new terpene synthases in nonseed plants. *Trends Plant Sci.* **23**, 121–128 (2018).
60. Xie, C. & Lou, H. Secondary metabolites in bryophytes: an ecological aspect. *Chem. Biodivers.* **6**, 303–312 (2009).
61. Romani, F. et al. Liverwort oil bodies: diversity, biochemistry, and molecular cell biology of the earliest secretory structure of land plants. *J. Exp. Bot.* **73**, 4427–4439 (2022).
62. Asakawa, Y., Ludwiczuk, A. & Nagashima, F. Chemical constituents of bryophytes. Bio- and chemical diversity, biological activity, and chemosystematics. *Prog. Chem. Org. Nat. Prod.* **95**, 1–796 (2013).

[1]Key Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden & Chinese Academy of Sciences, Shenzhen, China. [2]State Key Laboratory of Genome and Multi-omics Technologies, Key Laboratory of Genomics, Ministry of Agriculture, BGI Research, Shenzhen, China. [3]BGI Research, Wuhan, China. [4]College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China. [5]College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing, China. [6]Department of Plant Biotechnology and Bioinformatics, Ghent University and VIB-UGent Center for Plant Systems Biology, Gent, Belgium. [7]State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng, China. [8]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. [9]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, Beijing, China. [10]College of Life Sciences, Beijing Key Laboratory of Plant Gene Resources and Biotechnology for Carbon Reduction and Environmental Improvement, Capital Normal University (CNU), Beijing, China. [11]State Key Laboratory of Crop Stress Biology for Arid Areas/Shaanxi Key Laboratory of Apple, College of Horticulture, Northwest A&F University, Yangling, China. [12]Department de Biologie, Université Laval, Québec City, Québec, Canada. [13]IPSIM, Université Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France. [14]School of Life Sciences, Hubei University, Wuhan, China. [15]Lomonosov Moscow State University, Moscow, Russia. [16]State Key Laboratory of Genome and Multi-omics Technologies, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [17]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi Key Lab of Sugarcane Biology, College of Agriculture, Guangxi University, Nanning, China. [18]Sub-Antarctic Biocultural Conservation Program, Department of Philosophy and Religion, and Department of Biological Sciences, University of North Texas, Denton, TX, USA. [19]Cape Horn International Center (CHIC), Omora Ethnobotanical Park, Universidad de Magallanes, Puerto Williams, Chile. [20]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA. [21]New York Botanical Garden, Bronx, NY, USA. [22]Department of Plant Biology, Southern Illinois University, Carbondale, IL, USA. [23]South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China. [24]Instituto de Ciencias Aplicadas, Universidad Autónoma de Chile, Santiago, Chile. [25]Instituto de Agrobiotecnología del Litoral, Universidad Nacional del Litoral—CONICET, Facultad de Bioquímica y Ciencias Biológicas, Santa Fe, Argentina. [26]Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research (MPIPZ), Cologne, Germany. [27]Faculty of Chemistry and Pharmacy, University of Freiburg, Freiburg, Germany. [28]Department of Biology, East Carolina University, Greenville, NC, USA. [29]Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China. [30]Department of Ecology and Evolution, The University of Chicago, Chicago, IL, USA. [31]ARC Centre of Excellence for Plant Success in Nature and Agriculture, School of Biological Sciences, Monash University, Melbourne, Victoria, Australia. [32]Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa. [33]BGI Life Science Joint Research Center, Northeast Forestry University, Harbin, China. [34]These authors contributed equally: Shanshan Dong, Sibo Wang, Linzhou Li, Jin Yu, Yongxia Zhang, Jia-Yu Xue. [35]These authors jointly supervised this work: John L. Bowman, Yves Van de Peer, Huan Liu, Yang Liu. ✉e-mail: john.bowman@monash.edu; yves.vandepeer@psb.ugent.be; liuhuan@genomics.cn; liuyang@szbg.ac.cn

Methods

Taxon sampling, and DNA and RNA extraction

Bryophyte samples were collected from the field in Antarctica, Belgium, Canada, Chile, China, Germany, New Zealand, Russia, South Africa and the United States, and then maintained in environmental growth chambers. Voucher specimens were deposited in the Herbarium of Shenzhen Fairy Lake Botanical Garden in Shenzhen, China (SZG) and George Safford Torrey Herbarium at the University of Connecticut (CONN). Except for the three hornworts that used axenic cultures (DNA 639, 899 and 902; Supplementary Data 1), all the other bryophytes used field collection materials for DNA extraction and genome sequencing. Bryophytes were cleaned at least thrice with distilled water, and sampled under the dissecting microscope to avoid potential contamination. Genomic DNA and RNA were extracted using the FastPure Plant DNA Isolation Mini Kit (Vazyme) and RNA-easy Isolation Reagent (Vazyme), respectively. DNA and RNA quantification and qualification were performed using 1% agarose gel electrophoresis, Qubit fluorometer (Thermo Fisher Scientific), and NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). Our sampling spreads across the bryophyte phylogeny, representing all but 8 (that is, Catoscopiales, Dendrocerotales, Disceliales, Leiosporocerotales, Neohodgsoniales, Oedipodiales, Phymatocerotales and Scouleriales) of the 55 currently recognized orders of extant bryophytes[24,63].

Genome sequencing and *k*-mer analysis

DNA libraries for short-read whole-genome sequencing were constructed using the MGIEasy FS DNA Library Prep Set (Item 1000006988) or the Illumina TruSeq DNA PCR-free library preparation kit (Illumina) with 300–500 bp fragment sizes, and sequenced on the MGI-SEQ or Illumina NovaSeq 6000 platform to generate 150 bp paired-end reads. Transcriptome libraries were constructed with a TruSeq RNA Library Prep Kit v2 (Illumina) with an insert size of 200–400 bp, after polyA selection, and were sequenced for 150 bp paired-end reads on the MGI-SEQ or Illumina NovaSeq 6000 platform. For long-read sequencing, DNA fragments over 30 kb were selected for library construction, and the library was first end-repaired with NEBNext FFPE Repair Mix (New England Biolabs) and the NEBNext Ultra II End Repair/dA-Tailing Module (NEB), and then prepared for sequencing using the SQK_LSK109 Ligation Sequencing Kit. We used the Nanopore PromethION sequencer to generate long reads for 38 species. We generated single-tube long fragment reads (stLFR) and 10x Genomics reads for 85 species. The stLFR library was prepared with the MGIEasy stLFR Library Prep Kit (Item 1000005622) and sequenced with MGI-SEQ. The 10x Genomics Chromium Genome libraries were prepared with Chromium Genome Reagent Kit (v2 Chemistry) following the manufacturer's protocols and sequenced on an MGI-SEQ platform. Hi-C libraries were generated using DpnII restriction enzymes following in situ ligation protocols, and sequenced on the Illumina NovaSeq 6000 platform to generate 150-bp reads. We also prepared MethylC-seq libraries according to the protocol described in ref. 64 and sequenced on an Illumina NovaSeq 6000 platform to generate 150-bp reads.

The whole-genome sequencing short paired reads were filtered with TRIMMOMATIC v0.39 (ref. 65) to remove duplicates, low-quality reads and adaptors. The quality-controlled reads were further filtered with GC content (Supplementary Fig. 20) to remove aberrant reads, and the filtered reads were used for *k*-mer analyses to estimate the genome size. JELLYFISH v2.3.0 (ref. 66) and KMERFREQ[67] were used to count the *k*-mer frequencies with parameters '-m -C'. GENOMESCOPE v2.0 (ref. 68) was used to estimate genome size with '-p 1' and other parameters with default settings.

Genome assembly, Hi-C scaffolding and decontamination

Nanopore long reads were assembled with NEXTDENOVO v2.4.0 (https://github.com/Nextomics/NextDenovo) with the parameter 'seed_cutoff = 45'. RACON v1.4.7 (ref. 69) and PILON v1.23 (ref. 70) were then used to polish the assembled contigs with nanopore long reads and short reads, respectively. For stLFR data, STLFR2SUPERNOVA pipeline was used to assemble the stLFR reads into scaffolds with default parameters. SUPERNOVA v2.1.1 (ref. 71) was used to assemble the 10× genomics reads with the default parameters (Supplementary Fig. 21).

Raw Hi-C reads were first filtered by TRIMMOMATIC v0.39 (ref. 65) with default parameters (Supplementary Fig. 22). JUICER v1.6 (ref. 72) was then used to extract valid data. The 3D-DNA pipeline v180922 (ref. 73) was used to correct, anchor, order and orient the misassembled contigs based on the cross-linked Hi-C data (Supplementary Fig. 22). The JUICEBOX ASSEMBLY TOOLS v1.11.08 (ref. 74) was used to manually adjust the mis-joins, translocations and inversions (for results, see Supplementary Fig. 23).

The draft genome assemblies were filtered for potential contaminants. We first performed BLASTN[75] search for the sequences with an *e*-value threshold of 1×10^{-5} (Supplementary Fig. 22). The National Center for Biotechnology Information's (NCBI's) nucleotide collection and nonredundant protein sequences (nt/nr database, released in May 2020) were used as the BLAST reference databases. The taxonomic affiliations were assigned to each BLAST hit. In detail, sequences from the draft assemblies were queried (BLAST) against the nt database, and scaffolds with resultant sites with more than 50% hits on non-Embryophyta (calculated by sites getting hits on non-Embryophyta/sites getting any hits in nt database) were removed. Then, the retained scaffolds were further queried against the nr database with DIAMOND v0.9.25 (ref. 76) using BLASTX, and scaffolds with resultant sites with more than 75% hit on non-Embryophyta (calculated by sites getting hits on non-Embryophyta/sites getting any hits in nr database) were removed. Mitochondrial and plastid sequences were removed by performing BLASTN search against organellar sequences from three model bryophytes (that is, *M. polymorpha*, *P. patens* and *Anthoceros angustus*). To evaluate the reliability of the decontamination pipeline, we compared the GC-depth distribution of the draft and the clean assemblies with a sliding window size of 10 kb, and the results indicated that the decontamination pipeline is effective (Supplementary Fig. 20).

Genome annotation

For repeat annotation, we customized a de novo repeat library with homology-based approach. PILER v0.4.1 (ref. 77), LTR_FINDER v1.0.5 (ref. 78), REPEATSCOUT v1.0.5 (ref. 79) and REPEATMODELER v2.0 (ref. 80) were used to create a custom repeat library, which was then used for de novo repeat identification. REPEATMASKER v4.1.2 was then used for homology-based repeat identification against the Repbase database[81]. To calculate the insertion time of LTRs, LTR_STRUC[82] was used to extract the complete 5′ and 3′ ends of the LTR elements. DISTMAT from EMBOSS v6.5.7.0 (ref. 83) was then used to calculate the *K* value of the retrotransposons' 5′-LTR and 3′-LTR sequences. Finally, the insertion time (*T*) of LTRs was calculated by the formula $T = K/2r$, where *r* is the average substitution rate of 9.4×10^{-9} substitutions per year per synonymous site[84]. The BRAKER v2.1.5 (ref. 85) pipeline was applied for gene structure annotation after repeat masking, by integrating the transcriptome-based, homology-based and ab initio evidence. Protein sequences of six green plants (that is, *Arabidopsis thaliana*, *Azolla filiculoides*, *M. polymorpha*, *P. patens*, *Salvinia cucullata* and *Selaginella moellendorffii*) retrieved from the Phytozome v13 database (https://phytozome-next.jgi.doe.gov/) were used as the homology-based evidence. To evaluate the completeness of the assemblies, we conducted BUSCO v5.3.2 (ref. 86) assessment on the assemblies against the Viridiplantae_odb10 database.

For gene functional annotation, the annotated protein sequences were aligned against the KEGG (https://www.genome.jp/kegg/) and SWISSPROT (https://www.uniprot.org/) databases by NCBI BLASTP[75] with an *e* value of 1×10^{-5}. Protein domains were predicted with INTERPROSCAN v5.51-85.0 (ref. 87). Cluster of orthologous groups functional categories were used for functional annotation and classification of genes using the EGGNOG-MAPPER online server[88,89] (Supplementary Fig. 24).

Methylation analysis

To identify the methylation pattern of bryophytes, the pair-end reads generated by whole-genome bisulfite sequencing (WGBS) were mapped to the corresponding representative genomes using BISMARK v0.23.1 (ref. 90). Methylated cytosines (C) of CG, CHG and CHH were extracted using bismark_methylation_extractor under '–CX_context –comprehensive' parameters, and only the methylated cytosines covered by at least five reads were used for calculation of the methylation level. The methylation level of CG, CHG and CHH was calculated for each candidate cytosine as the number of mapped methylated reads divided by the number of total mapped reads, that is, mC/(mC + T). For gene methylation analyses, the gene body and gene flanking region (±2 kb) were equally divided into 20 bins, respectively. The weighted methylation level of each bin was calculated using the 'summed number of mapped methylated reads' divided by the 'summed number of total mapped reads', that is, (mC1 + mC2 + … + mCn)/(mC1 + mC2 + … + mCn + T1 + T2 + … + T*n*).

WGD analysis

JCVI v.1.1.8 (ref. 91) was used to generate the dot plot and show the relationship of intragenomic collinear blocks with a cscore cutoff of 0.99 (for results, see Supplementary Fig. 25). Synonymous substitutions per synonymous site (K_s) estimates for pairwise comparisons of paralogues located on syntenic blocks were obtained using the Nei–Gojobori method[92] implemented in yn00 program of the PAML package[93] (for results, see Supplementary Fig. 26).

KSRATES v1.1.1 (ref. 94) was also used for K_S-age distributions of paralogues and orthologues for representative bryophyte species as shown in Supplementary Figs. 27–31. The correction of synonymous substitution rates was realized in KSRATES. The maximum number of trios was set to 14, 8 and 7 for the moss, liverwort and hornwort datasets, respectively. An exponential-lognormal mixture model was fitted upon each whole paranome K_S-age distribution. I-ADHORE v.3.0.01 (ref. 95) was used to infer anchor pairs for the subsequent K_S analysis as required by KSRATES. An anchor K_S clustering was performed to categorize anchor pairs into distinct groups, potentially indicating different WGDs. Ten representative moss species were selected to test all candidate WGDs while avoiding redundant species sampling, as shown in Supplementary Figs. 32–36 (for method details, see Supplementary Note 5).

Phylogenetic and dating analyses

Phylogenetic analyses were performed based on a 54-taxon (Supplementary Data 8), a 190-taxon (Supplementary Data 4 and Extended Data Fig. 2) and a 343-taxon dataset (Supplementary Data 5), respectively. ORTHOFINDER v2.4.0 (ref. 96) was used to generate gene families with all the proteins with default settings. KINFIN v1.0.3 (ref. 97) was used to select the mostly single-copy gene families with default settings. The protein sequences from the selected gene families aligned with MAFFT v5.0 (ref. 98), trimmed for ambiguous portions by GBLOCKS v0.91b[99] with -b3 (the least stringent settings with maximum number of contiguous nonconserved positions) set to 8, -b4 (minimum length of a block) set to 5 and -b5 (allowed gap positions) set to half and -b6 (use similarity matrices) set to yes, and subjected to maximum likelihood tree optimization with RAxML v8.2.12 (ref. 100) and ASTRAL v5.7.3 (ref. 101) with the concatenated and coalescent methods. The PROTOGAMMAAUTO model and 200 fast bootstrap replicates were used in the RAxML tree inference method (for alignments and phylogenetic trees, see Figshare repository, https://doi.org/10.6084/m9.figshare.23528667). PHYPARTS v0.0.1 (ref. 102) was used to map all individual genes to view gene tree incongruences on each node.

For molecular dating analysis, genes with concordant evolutionary histories were selected by SORTADATE v1.0 (ref. 103) and the resulting alignments were concatenated to generate the XML file in BEAUti as implemented in BEAST v2.4.6 (ref. 104). Then we performed the dating analysis with fossil calibrations (Supplementary Data 26) by using the uncorrelated lognormal relaxed molecular clock model and LG + G model with four site-rate categories in BEAST v2.4.6 (ref. 104). Different loci were partitioned, and model parameters were unlinked across different partitions. A starting tree with branch lengths satisfying all fossil prior constraints was generated with TREEPL[105]. Two independent Markov chains of Monte Carlo analyses of 10,000 M generations with 25% burn-in and sampling every 50,000 generations were carried out to evaluate the credibility of posterior distributions of parameters. Convergence and mixing were assessed to make sure that the effective sample size values of all parameters exceeded 200 using Tracer v.1.5 (ref. 106).

Gene-based pangenome analysis

Gene-based pangenome analyses were conducted based on the dataset of 343 Archaeplastida genomes (see Supplementary Data 5 for details). ORTHOFINDER v2.4.0 (ref. 96) was used to generate gene families (orthogroups) with the default inflation value ($I = 1.5$). 'Core' gene families were defined as those presented in >80% samples of a lineage, 'accessory' gene families were defined as those presented in at least two but fewer than 80% of the total samples of a lineage and 'unique' gene families were defined as those presented in one individual. COUNT[107] was used to infer the gene family gains and losses along the tree with the DOLLO PARSIMONY model. A VENN diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to identify unique and shared gene families among the five Archaeplastida groups, that is, algae, tracheophytes, hornworts, liverworts and mosses. BLASTN was used to search the unique genes of *M. polymorpha* subsp. *ruderalis* against the noncoding genomic sequences of *M. polymorpha* subsp. *ruderalis* and the other three *Marchantia* (*M. polymorpha* ssp. *polymorpha* BR5, *M. polymorpha* ssp. *montivagans* SA2 and *M. paleacea* ssp. *diptera* MPA) with an *e*-value threshold setting to 1×10^{-10}. For each unique gene of *M. polymorpha* subsp. *ruderalis*, those contigs containing the corresponding best hit in the other three *Marchantia* genomes were extracted and aligned with MAFFT v5.0. The resultant alignments were analyzed in Geneious v10.0.2 (www.geneious.com) to trace the evolutionary history of the unique genes of *M. polymorpha* subsp. *ruderalis*. The three *M. polymorpha* genomes (that is, *M. polymorpha* subsp. *ruderalis*, BR5 and SA2) were downloaded from https://marchantia.info/, and the genome of MPA was downloaded from the NCBI database.

Identification of key functional genes

Transcription factors of the newly sequenced bryophyte genomes were identified with iTAK v1.6 (ref. 108), and manually curated using phylogenetic and domain information. Genes related to phytohormone-related genes were blasted ($e \leq 1 \times 10^{-5}$) against the phytohormone genes of *Arabidopsis*, and manually verified by phylogeny and functional annotation. Other important genes related to the developmental and physiological/biochemical networks that may have facilitated adaptation of plants to the terrestrial environment were mined from the literature and public databases. Genes related to signaling and development of roots/rhizoids, root hairs, lateral roots, vascular tissue and stomata followed ref. 25, genes related to biosynthesis and regulation of the cuticle followed ref. 109, genes related to 3D growth followed ref. 110, genes related to symbiosis followed ref. 111, genes related to leaf development followed ref. 112, genes related to desiccation tolerance followed ref. 113, genes related to cell wall biosynthesis and regulation followed refs. 114–116, genes related to regulation and sporopollenin synthesis were mined from ref. 117, genes related to UV protection were mined from the GO database (GO:0009650; UV protection) and from ref. 118, genes related to the biosynthesis of phenylpropanoid, coumarin, flavonoid, lignin, as well as genes related to plant pathogen interaction, and rhythm were mined from the KEGG pathway and KO database (https://www.genome.jp/kegg/). The corresponding *Arabidopsis* protein sequences were used as queries and searched against

our 190 species dataset with NCBI BLASTP[75] with an *e*-value cutoff of 1×10^{-5} to retrieve homologs. The recovered hits were further filtered with the characteristic PFAM domains of the specific query, with an *e* value of $\leq 1 \times 10^{-5}$ and a coverage >60% of the queries. The gene family of each pathway was imported into COUNT[107] to infer the gains and losses of these genes along the phylogenetic tree.

The *TPS* genes were identified following the method described in ref. 119. The protein sequences were searched against the Pfam-A database locally using HMMER 3.0 (ref. 120) hmmsearch with an *e* value of 1×10^{-5}. Only sequences with the best hits from four HMM profiles that were considered as putative *TPS* genes are as follows: Terpene_synth_C (PF03936) and TPS N-terminal domain (PF01397), *TRI5* (PF06330) and *SmMTPSL*s (a profile created by using 48 microbial type *TPS*s identified from *S. moellendorffii*). All the putative TPS protein sequences were subjected to a BLASTP search against the NCBI's nonredundant database using the default parameters. A TPS was annotated as 'Microbial TPS-like' (MTPSL) protein if all the top ten best hits were from bacteria and/or fungi or highly similar to SmMTPSLs. Then all the annotated TPS protein sequences were aligned with MAFFT v5.0 (ref. 98), trimmed with TRIMAL v1.4 (ref. 121). The resulting alignment was imported into the program FASTTREE[122] to generate a maximum likelihood tree using the JTT model. The plant-type TPSs and MTPSLs were further confirmed and identified on the phylogenetic tree.

The identification of NOD-like receptor genes was described in Supplementary Note 6. Stress experiments and differential gene expression analysis for *Physcomitrella patens* were detailed in Supplementary Note 7.

HGT analysis

A set of 54 representative species of Viridiplantae (Supplementary Data 8) were used for HGT identification. Although we carried out a strict pipeline to decontaminate the bryophyte assemblies ('Genome assembly, Hi-C scaffolding and decontamination') to avoid potential contaminations, we also applied strict strategies and rigorous analyses to identify HGTs, and remove candidate HGTs residing on small scaffolds, and confirmed that flanking genes are genuine plant genes (Supplementary Fig. 37).

We roughly divided the taxa in the NCBI nr database into the following seven categories according to NCBI taxonomy lineage: bacteria, fungi, archaea, viruses, metazoans, Viridiplantae and other (eukaryote species except fungi, Metazoa and Viridiplantae). We set donor groups to be species from bacteria (excluding cyanobacteria), fungi, archaea, viruses or metazoans. The putative recipient was set to be groups of various taxonomic ranks represented by each sampled species (for example, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, etc.) and was listed in Supplementary Data 9.

HGT identification in our analyses included the following three main steps: whole-genome screening to identify candidate acquired genes, elimination of contamination and high-confidence HGT filtering. A detailed flowchart of the analyses can be found in Supplementary Fig. 37. To identify HGTs in each species, the protein sequences of each species were used for searching against NCBI nr database (downloaded in December 2021) with the parameter '-e-value 1e-5 –outfmt 6 –max_target_seqs 50,000' using NCBI BLASTP[75], and then potential HGTs were identified by the methods described in ref. 47. In brief, if the top BLASTP hit belonged to the putative donor group and the number of hits from other groups (excluding the recipient group) exceeded five, the query gene was classified as an HGT candidate. Then, we examined vertical inheritance of neighborhood genes, location of the candidate gene on a large scaffold and distribution of orthologs in closely related taxa to reduce the possibility of a candidate acquired gene being derived from sequence contamination. Finally, to obtain high-confidence HGTs, we used a series of parameters to filter those candidate HGTs which are as follows: the sequence identity between putative donor group and receipt group should be >20%; the species number in donor groups should be >50; except for the recipient groups, the top 50 species of the BLAST result should mainly belong to putative donor groups (bacteria or fungi or archaea or viruses or metazoans); candidate transferred genes with sequence identity over 80 were also carefully checked for homologous genes in closely related taxa; HGTs originating from metazoan should distribute only in Metazoa and the recipient group; and candidate transferred genes identified to be of cyanobacterial origin or acquired before the origin of the recipient group were excluded. The curated HGTs were mapped against the gene family evolution inferred by COUNT based on the 54-species tree.

For each gene potentially acquired through HGT, homologs identified through BLAST searches against our customized database were divided into the following seven categories according to NCBI taxonomy database: bacteria, fungi, archaea, viruses, metazoans, Viridiplantae and other eukaryotic species. These seven categories were then further divided into subgroups of lower taxonomic ranks (Supplementary Data 9). At least one sequence of each subgroup was selected to ensure sufficient and balanced sampling. Subgroups with fewer than three species were manually inspected to exclude sequencing contamination, such as mislabeled sequences in the nr database. Multiple sequence alignments were performed using MAFFT v5.0 (ref. 98) with default settings. Poorly aligned regions and gaps were removed manually or using TRIMAL v1.4 (ref. 121) with parameters '-automated1'. Maximum likelihood trees were constructed using IQTREE v2.2.0 (ref. 123), with automatically selected best-fit amino acid substitution model. Branch support values were estimated with 1,000 replicates using ultrafast bootstrapping in IQTREE (-bb 1,000).

Experimental verification of the function of *FBT* gene

To further confirm the function of the *FBT* gene (fungal body lectin domain PF07367 containing gene), we expressed this gene in vitro and added the FBT protein into the diet of the insects and calculated the growth and survival rates. The full-length coding sequence of the *FBT* gene from *A. sandvicensis* was optimized for its codons to improve the expression efficiency in the prokaryotic system, and the optimized sequence was synthesized and ligated to the pET-30a vector. The pET-30a-*FBT* plasmid was transformed into *Escherichia coli* BL 21 (DE3) pLysS cells; the resulting strain was used for the expression of recombinant proteins under the control of IPTG-inducible T7 promoter. Overnight grown cultures were diluted 20-fold with fresh 200 ml LB medium and further grown at 37 °C and 220 rpm rotation till the OD_{600} reached 0.5. The culture was induced by adding 1 mM final concentration of IPTG and incubated at 28 °C for 3–4 h. Cells were then collected and suspended in 20 ml of 50 mM Tris–HCl cold buffer with pH 8, containing 200 mM NaCl, then disrupted by sonication at 4 °C. The cell lysate was centrifuged at 12,000*g* for 40 min at 4 °C in a centrifuge. The supernatant from the previous step was loaded on a Ni-NTA agarose column pre-equilibrated with Tris–NaCl buffer at 4 °C. The column was washed extensively with Tris–NaCl buffer containing 20 mM imidazole, and the 6xHis-tagged protein was eluted with Tris–NaCl buffer containing 250 mM imidazole. The elution product containing pure protein was washed thrice with Tris–NaCl buffer and concentrated using Centricon (Millipore, PM10). The purified His-tagged protein was detected by western blot using an HRP-conjugated monoclonal antibody (HRP-66005); the His-tagged protein is about 20 kDa in size.

The fall warm *S. frugiperda* and cotton bollworm *H. armigera* were obtained from the Agricultural Genomics Institute, Chinese Academy of Agricultural Sciences at Shenzhen, and reared on artificial diet[124]. To avoid cannibalism, larvae were raised separately in a 5.0 cm × 1.5 cm glass tube after the third instar. Two-day-old third instars of *S. frugiperda* and *H. armigera* with an average body weight of 5.15 mg per individual and 6.35 mg per individual were used for the experiments. Toxin was added to the diet of the larvae, setting the initial concentration gradient of the toxin treatment to 0, 1, 2, 4, 8 and 16 μg ml^{-1} diet, respectively. Larvae fed with PBS were used as a control. In total, 14 larvae

were treated for each group, and three replicates were carried out. The mortalities were recorded 48 h after feeding on a toxin-containing diet. The larvae were weighed (M1) and transferred to the toxin-added diet. Larvae were reweighed 48 h later (M2). The relative growth rate (%) was calculated as ((M2−M1) × 100/M1). Toxin concentration of 4 μg ml^{-1} diet (corresponding to 0.47 μg toxin per mg larvae weight) caused 97.62% death in *H. armigera*. However, this FBT toxin did not cause mortality in *S. frugiperda*, and 8 μg ml^{-1} concentration of toxin diet (corresponding to 1.5 μg toxin per mg larvae weight) distinctly inhibited the growth rate of the larvae.

Statistics

To assess the statistical differences in the weight change dynamics of *S. frugiperda* larvae and the mortality rates of *H. armigera* larvae between the PBS control and FBT protein treatment groups, we used two-sided Welch's *t* test. This approach was chosen due to its robustness in handling unequal variances between groups, ensuring reliable comparisons. All analyses were conducted using R version 4.2.2 (R Foundation for Statistical Computing).

Reporting summary

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

Data availability

Genome assemblies and annotations can be found at the website www.bryogenomes.org. The genomes, raw genomic, transcriptomic and Hi-C data, as well as the WGBS data for DNA methylation analyses that were generated in this study have been deposited in the NCBI Sequence Read Archive (BioProject PRJNA1117192), and the CNGB data center (https://db.cngb.org/) under the project CNP0002895. The genome annotation files, orthogroup information, data alignments and phylogenetic trees are available in the Figshare repository (https://doi.org/10.6084/m9.figshare.23528667.v6 (ref. 125)).

Code availability

No custom code or mathematical algorithms were developed in this study. All analyses were performed using publicly available software tools, with either default settings or parameters as specified in the Methods section. Details regarding software versions and analysis pipelines are provided in the Methods section.

References

63. Söderström, L. et al. World checklist of hornworts and liverworts. *PhytoKeys* **27**, 1–828 (2016).
64. Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J. & Ecker, J. R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483 (2015).
65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
67. Liu, B. et al. Estimation of genomic characteristics by analyzing *k*-mer frequency in de novo genome project. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1308.2012 (2013).
68. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
69. Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 727–736 (2017).
70. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
71. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
72. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
73. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
74. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
75. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
76. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
77. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**, i152–i158 (2005).
78. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
79. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
80. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
81. Jurka, J. et al. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
82. McCarthy, E. M. & McDonald, J. F. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**, 362–367 (2003).
83. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
84. Kamisugi, Y. et al. A sequence-anchored genetic linkage map for the moss, *Physcomitrella patens*. *Plant J.* **56**, 855–866 (2008).
85. 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. Bioinform.* **3**, lqaa108 (2021).
86. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
87. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
88. 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).
89. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
90. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
91. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
92. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
93. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **13**, 555–556 (1997).

94. Sensalari, C., Maere, S. & Lohaus, R. ksrates: positioning whole-genome duplications relative to speciation events in *KS* distributions. *Bioinformatics* **38**, 530–532 (2021).
95. Proost, S. et al. i-ADHoRe 3.0-fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* **40**, e11 (2012).
96. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
97. Laetsch, D. & Blaxter, M. KinFin: software for taxon-aware analysis of clustered protein sequences. *G3 (Bethesda)* **7**, 3349–3357 (2017).
98. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
99. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
100. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
101. Mirarab, S. et al. Astral: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).
102. Smith, S., Moore, M., Brown, J. W. & Yang, Y. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol. Biol.* **15**, 150 (2015).
103. Smith, S. A., Brown, J. W. & Walker, J. F. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS ONE* **13**, e0197433 (2018).
104. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T. & Drummond, A. J. Beast 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
105. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
106. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
107. Miklós, C. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**, 1910–1912 (2010).
108. Yi, Z. et al. Itak: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**, 1667–1670 (2016).
109. Kong, L. et al. Origins and evolution of cuticle biosynthetic machinery in land plants. *Plant Physiol.* **184**, 1998–2010 (2020).
110. Whitewoods, C. D. et al. *CLAVATA* was a genetic novelty for the morphological innovation of 3D growth in land plants. *Curr. Biol.* **28**, 2365–2376 (2018).
111. Radhakrishnan, G. V. et al. An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. *Nat. Plants* **6**, 280–289 (2020).
112. Wang, H., Kong, F. & Zhou, C. From genes to networks: the genetic control of leaf development. *J. Integr. Plant Biol.* **63**, 1181–1196 (2021).
113. Xu, Z. et al. Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. *Mol. Plant* **11**, 983–994 (2018).
114. Popper, Z. A. et al. Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant. Biol.* **62**, 567–590 (2011).
115. Fangel, J. U. et al. Cell wall evolution and diversity. *Front. Plant Sci.* **3**, 152 (2012).
116. Cheng, S. et al. Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* **179**, 1057–1067 (2019).
117. Wang, K. et al. The regulation of sporopollenin biosynthesis genes for rapid pollen wall formation. *Plant Physiol.* **178**, 283–294 (2018).
118. Clayton, W. A. et al. *UVR8*-mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort *Marchantia polymorpha* and flowering plants. *Plant J.* **96**, 503–517 (2018).
119. Jia, Q. et al. Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *Proc. Natl Acad. Sci. USA* **113**, 12328–12333 (2016).
120. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121 (2013).
121. 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).
122. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
123. Quang, M. B. et al. IQTREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
124. Jin, M. et al. Two ABC transporters are differentially involved in the toxicity of two *Bacillus thuringiensis* Cry1 toxins to the invasive crop-pest *Spodoptera frugiperda* (J. E. Smith). *Pest Manag. Sci.* **77**, 1492–1501 (2021).
125. Dong, S. Genome annotation files, orthogroup information, phylogenetic trees and alignments used for bryophyte genome project. *figshare* https://doi.org/10.6084/m9.figshare.23528667.v6 (2023).

Acknowledgements

This study was supported by the Biological Breeding–National Science and Technology Major Project (2023ZD04073 to H.L.), the Scientific Foundation of the Urban Management Bureau of Shenzhen (202005 and 202403 to Y.L. and 202106 and 202302 to S.D.), the Shenzhen Science and Technology Program (KJZD20230921114607016 and KQTD20230301092839007 to H.L.), the European Research Council under the European Union's Horizon 2020 research and innovation program (833522 to Y.V.d.P.) and Ghent University Methusalem funding (BOF.MET.2021.0005.01 to Y.V.d.P.), and Australian Research Council (CE200100015 to J.L.B.). The authors acknowledge D. Gleny (Landcare Research), D. Liu and W. Ma (Kunming Institute of Botany, CAS), T. Peng (Guizhou Normal University), Y. Qiu (Michigan University), W. Sheng and Q. Zuo (Fairy Lake Botanical Garden), Q. Wang (Institute of Botany, CAS), E. T. Y. Wu (University of British Columbia) for providing or assisting in collecting bryophyte materials, and B. Aguero (Duke Herbarium), D. Callaghan (Bryophyte Surveys), D. Gleny (Landcare Research), J. Duckett (Natural History Museum) and Š. Koval (Agentura ochrany přírody a krajiny ČR) for providing bryophyte photos. The authors then thank C. Schlichting (University of Connecticut) and X. Zhang (Shenzhen Agricultural Genome Research Institute, CAAS) for commenting on an earlier draft of the manuscript. The authors also thank R. Ramnath (University of Connecticut) for his assistance in constructing the website www.bryogenomes.org, and F.-W. Li (Cornell University) for kindly sharing the bryophyte genomic data at an early stage of our analysis. This work is part of the 10KP project (https://db.cngb.org/10kp/) and was also supported by Guangdong Provincial Key Laboratory of core collection of crop

genetic resources research and application, BGI-Shenzhen (NO. 2011A091000047).

Author contributions

Y.L., H.L., Y.V.d.P., J.L.B. and S.D. led and managed the project. Y.L., H.L., Y.V.d.P., J.L.B., B.G. and S.D. conceived the study. S.D., Y.L., B.G., M.I., J.L., Y.Y., R.R., R.M., L.Z., Y.H. and S.Z. prepared materials. S.D., S.W., L. Li, J.Y., Y. Zeng, Y.C., X.Z., T.Z., J.R., H.W., R.H. and J.W. performed bioinformatics analysis. J.H., J.M., S.D., S.W., L.L., H.C. and Y.V.d.P. performed other data analyses. J.W., W.H., M.J., L. Liu and B.X. carried out lab experiments. S.D., Y.L., S.W., L. Li, J.Y., Y. Zhang and J.Y.X. drafted the manuscript. Y.V.d.P., J.L.B., B.G., S.A.R., M.M., J.M., K.S.R., D.W.S., J.C.V.A., H.C., L.D., M.L., T.W., Y.G. and X.X. contributed substantially to revisions. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

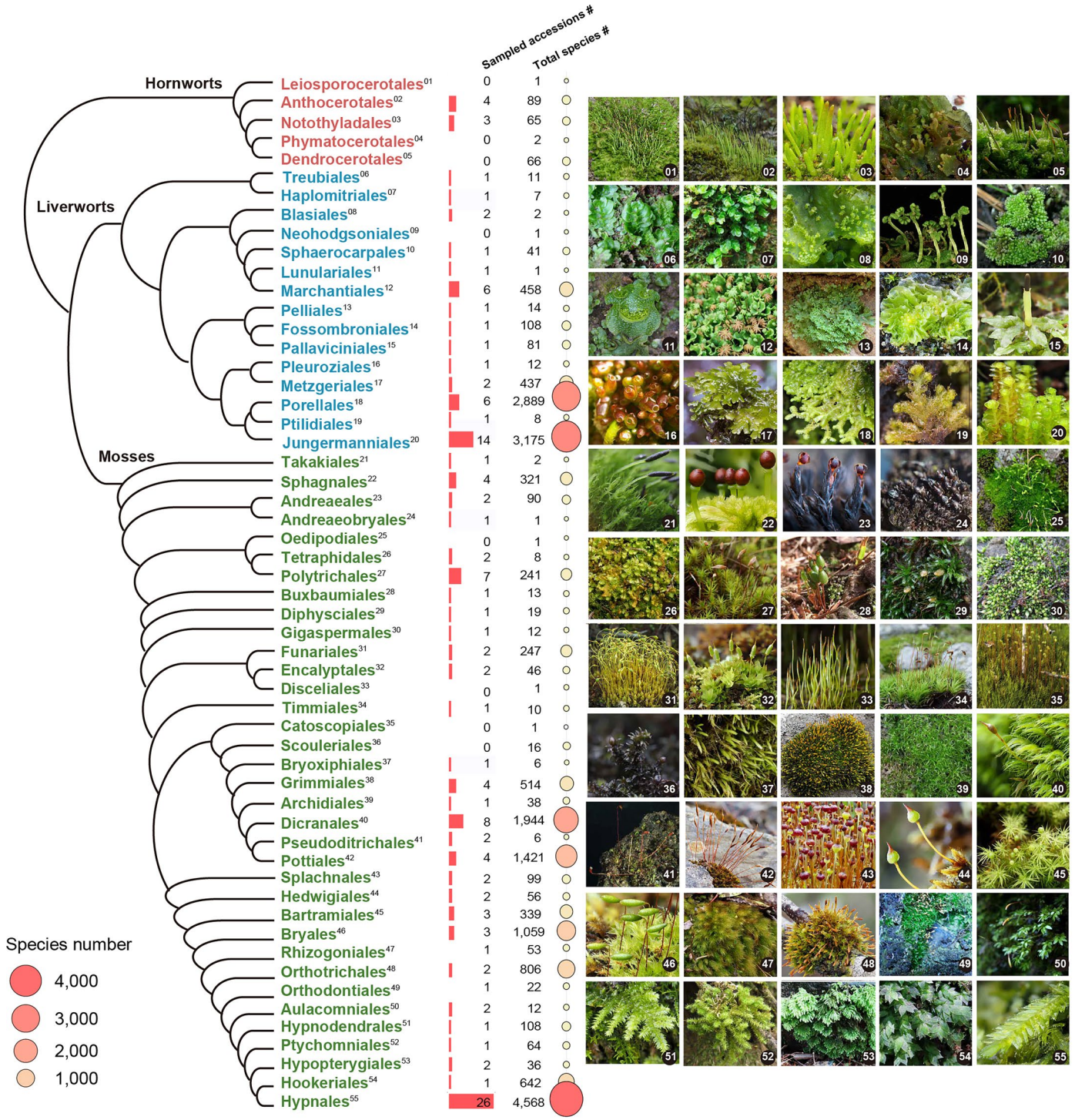
Extended data is available for this paper at https://doi.org/10.1038/s41588-025-02325-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02325-9.

Correspondence and requests for materials should be addressed to John L. Bowman, Yves Van de Peer, Huan Liu or Yang Liu.

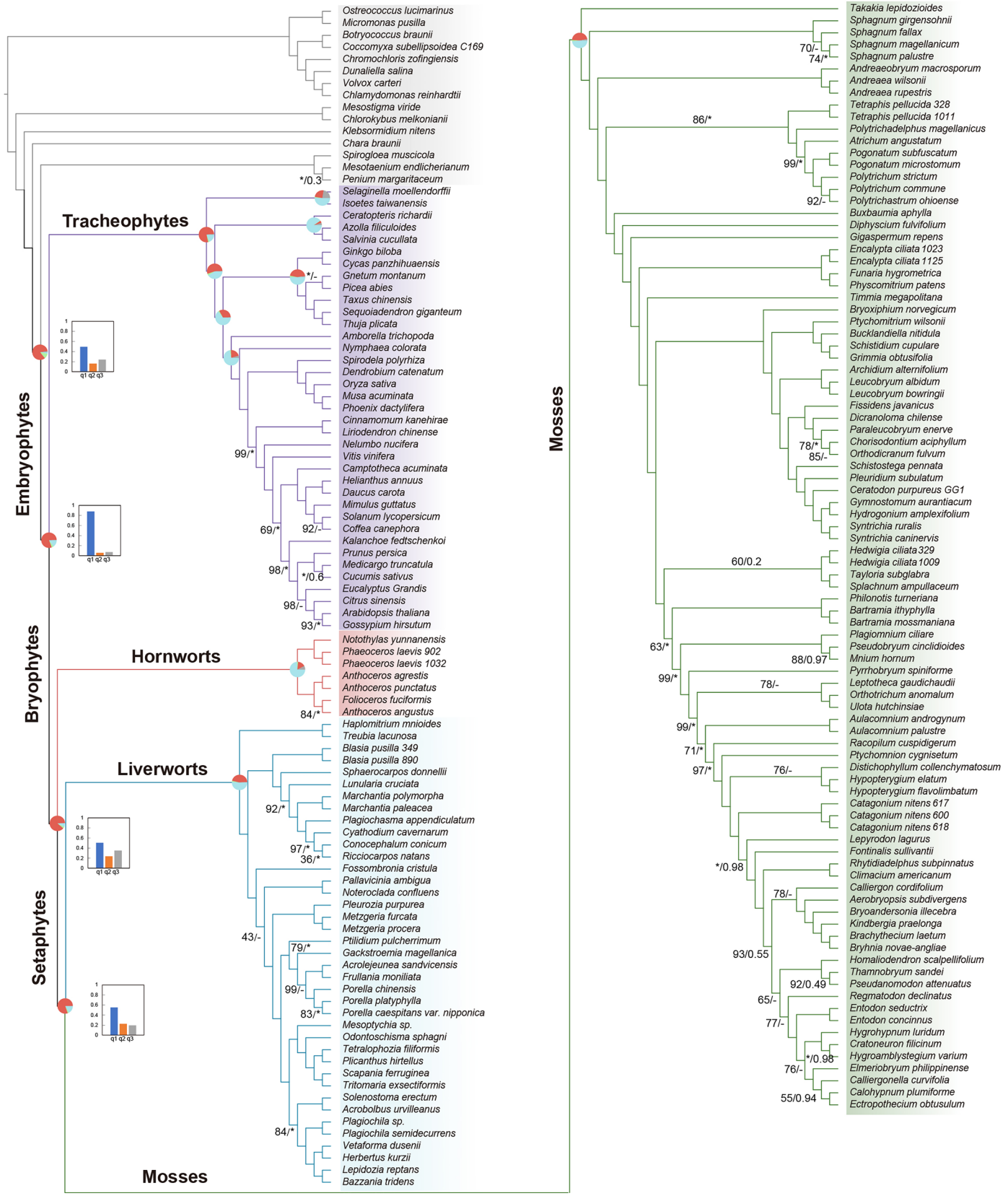
Peer review information *Nature Genetics* thanks Bojian Zhong and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41588-025-02325-9



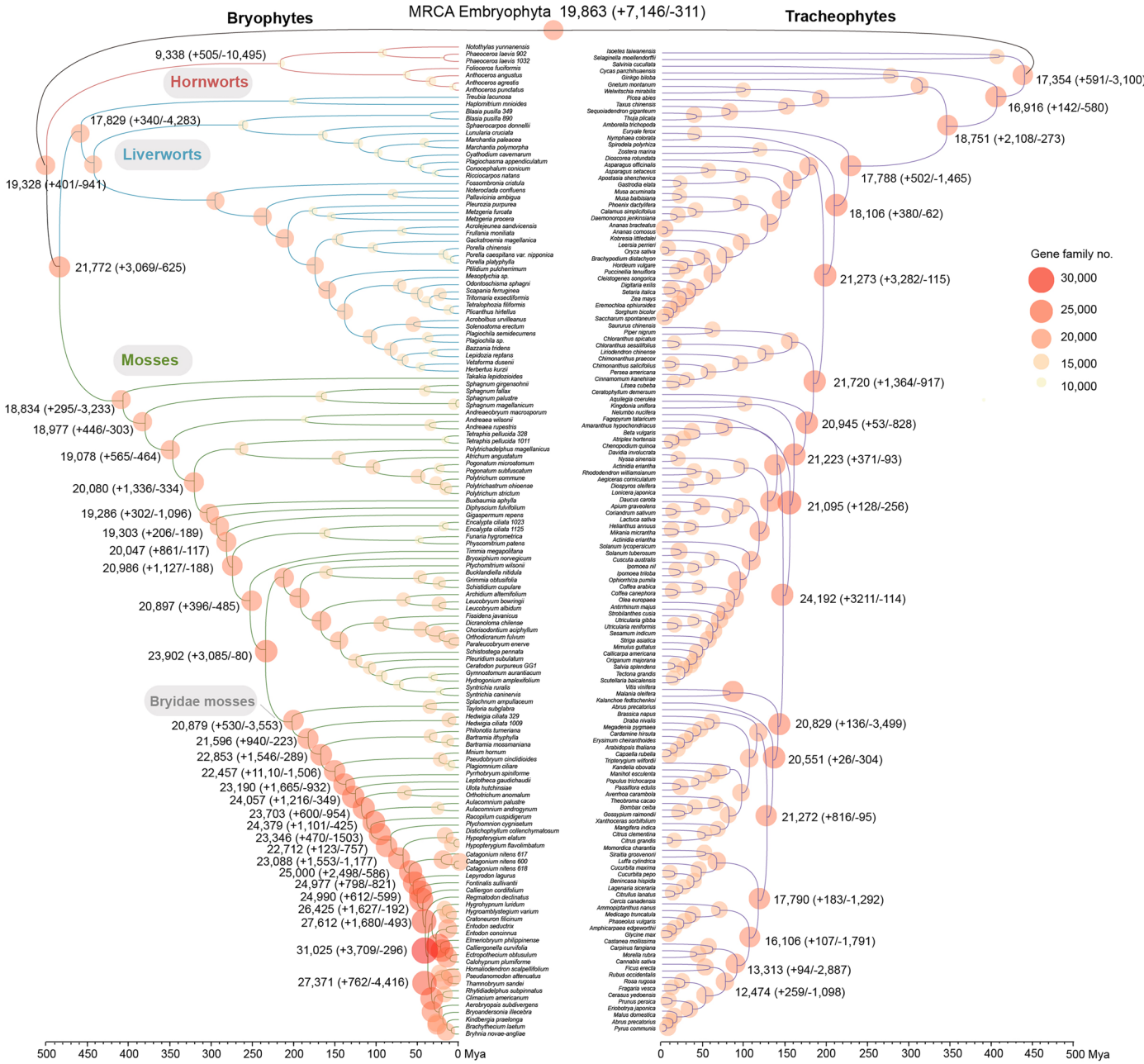
Extended Data Fig. 1 | Ordinal phylogeny of bryophytes with the number of genomes sequenced (123 new + 15 published) and number of species recognized in each order. Current sampling covers 47 of the 55 recognized bryophyte orders. Representative photographs (not to scale) for each order are as follows: (1) *Leiosporoceros dusii*, (2) *Anthoceros angustus*, (3) *Phaeoceros minutus*, (4) *Phymatoceros bubliculosus*, (5) *Dendroceros cavernosus*, (6) *Treubia lacunosa*, (7) *Haplomitrium mnioides*, (8) *Blasia pusilla*, (9) *Neohodgsonia mirabilis*, (10) *Sphaerocarpos texanus*, (11) *Lunularia cruciata*, (12) *Marchantia polymorpha*, (13) *Apopellia endiviifolia*, (14) *Petalophyllum rafsii*, (15) *Symphyogyna brasiliensis*, (16) *Pleurozia acinosa*, (17) *Riccardia longispica*, (18) *Radula ankefinensis*, (19) *Ptilidium pulcherrimum*, (20) *Gottschelia schizopleura*, (21) *Takakia lepidozioides*, (22) *Sphagnum fallax*, (23) *Andreaea frigida*, (24) *Andreaeobryum macrosporum*, (25) *Oedipodium griffithianum*, (26) *Tetraphis pellucida*, (27) *Polytrichum commune*, (28) *Buxbaumia aphylla*, (29) *Diphyscium longifolium*, (30) *Gigaspermum repens*, (31) *Funaria hygrometrica*, (32) *Encalypta ciliata*, (33) *Discelium nudum*, (34) *Timmia* sp., (35) *Catoscopium nigritum*, (36) *Scouleria* sp., (37) *Bryoxiphium norvegicum* ssp. *japonicum*, (38) *Grimmia donniana*, (39) *Archidium ohioense*, (40) *Leucoloma sprengelianum*, (41) *Pseudoditrichum* sp., (42) *Gertrudiella xanthocarpa*, (43) *Splachnum vasculosum*, (44) *Rhacocarpus purpurascens*, (45) *Bartramia halleriana*, (46) *Ptychostomum inclinatum*, (47) *Pyrrhobryum dozyanum*, (48) *Ulota fulva*, (49) *Orthodontium lignicola*, (50) *Aulacomnium heterostichum*, (51) *Racopilum capense*, (52) *Ptychomnion cygnisetum*, (53) *Hypopterygium flavolimbatum*, (54) *Hookeria acutifolia* and (55) *Catagonium nitens*. Photographs provided by J.C. Villarreal (1), L. Zhang (2, 4, 5, 16, 26, 27, 29, 34, 37, 38, 39, 45, 47 and 50), D. Callaghan (3, 08, 14, 15, 17, 18, 20, 22, 32, 40, 42, 43, 44, 46, 48, 51 and 55), J.G. Duckett (6), D. Gleny (9), B. Aguero (10), Š. Koval (11, 13, 19 and 23), S.S. Dong (7 and 12), Y. He (21), M. Ignatov (24, 25, 33, 35, 36, 41 and 49), Y. Liu (28, 53 and 54), B. Goffinet (30, 31 and 52).

https://doi.org/10.1038/s41588-025-02325-9



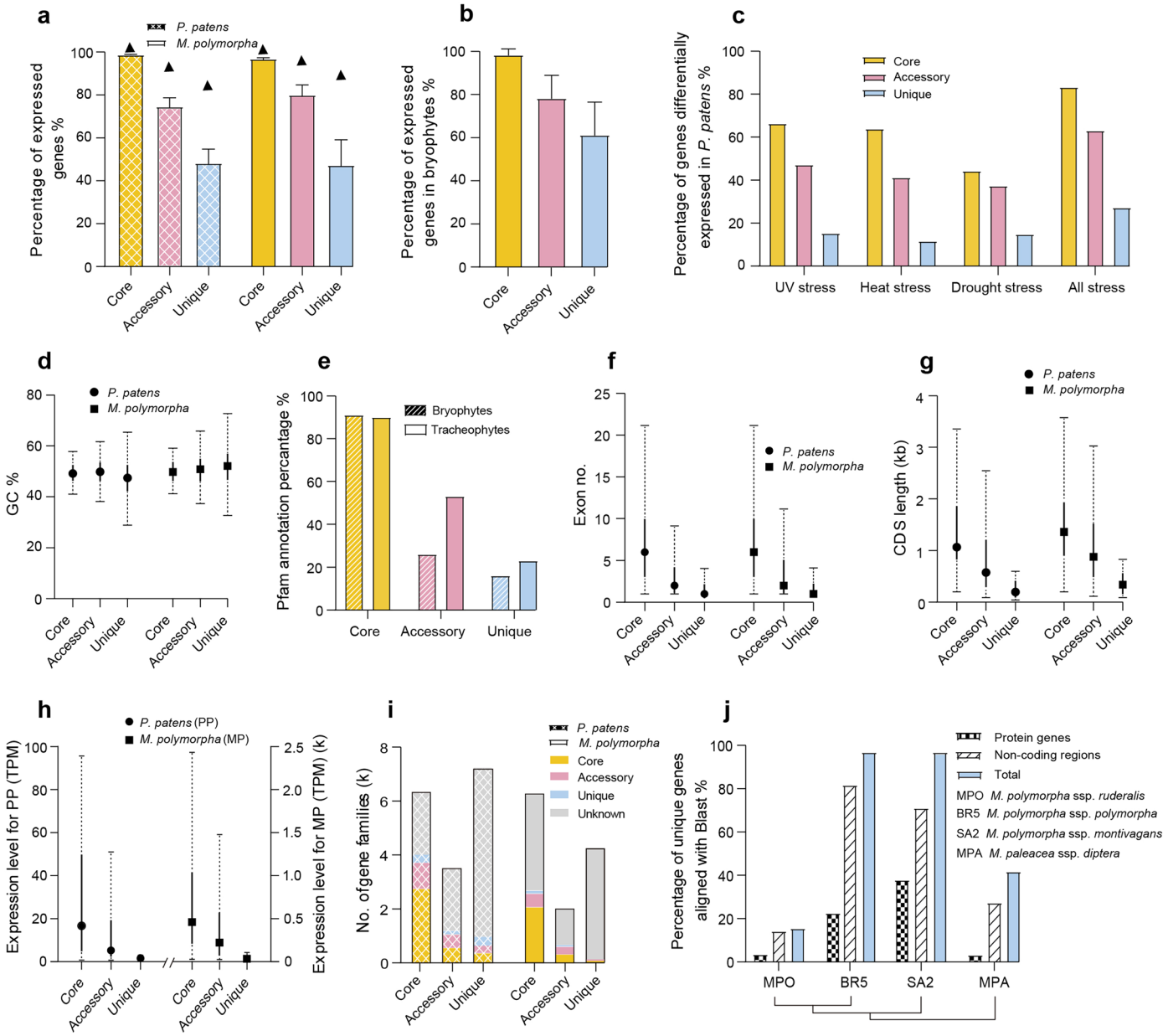
Extended Data Fig. 2 | Cladogram of maximum likelihood (ML) tree of 190 green plants based on 587 low copy nuclear protein-coding genes. All branches are maximally supported by both the ML bootstrap supports and ASTRAL posterior probabilities unless otherwise indicated. Maximum supports by either method were indicated as '*', nonsupported branches were indicated as '–'. Gene tree conflicts are shown on focal nodes as pie charts indicating concordance (light blue), top conflict (lime green), other conflict (red) and no signal (gray). Quartet supports are shown beside the focal nodes as bar charts, where blue, orange and gray represent quartet supports for the main topology (q1), the first alternative topology (q2) and the second alternative topology (q3), respectively.

https://doi.org/10.1038/s41588-025-02325-9



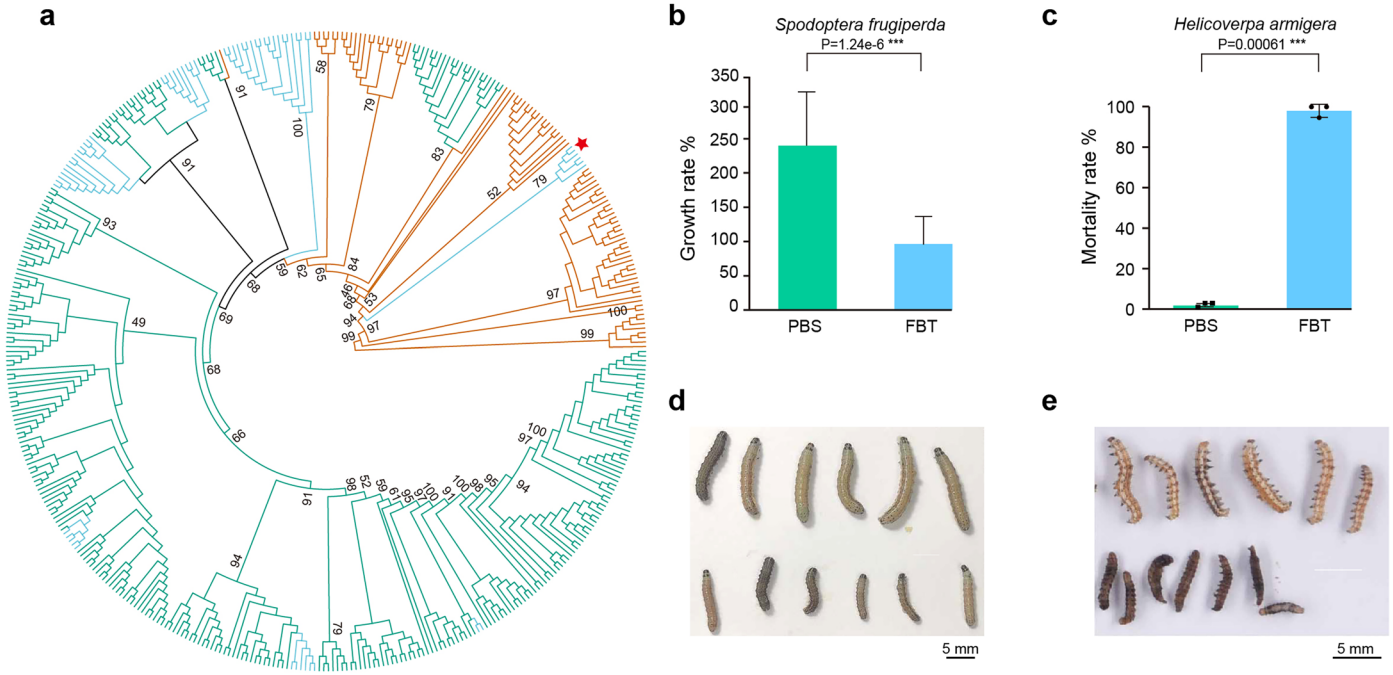
Extended Data Fig. 3 | Reconstruction of the number of gene families (orthogroups) in Embryophyta. The numbers of gene families were shown on ancestral nodes along the chronogram of bryophytes and tracheophytes. The bubble on each node denotes the total number of gene families reconstructed by COUNT with the Dollo parsimony model. The numbers besides the stem nodes indicate the total, gained and lost gene families. The chronogram of 285 embryophytes was calculated by BEAST v2.4.6.

https://doi.org/10.1038/s41588-025-02325-9



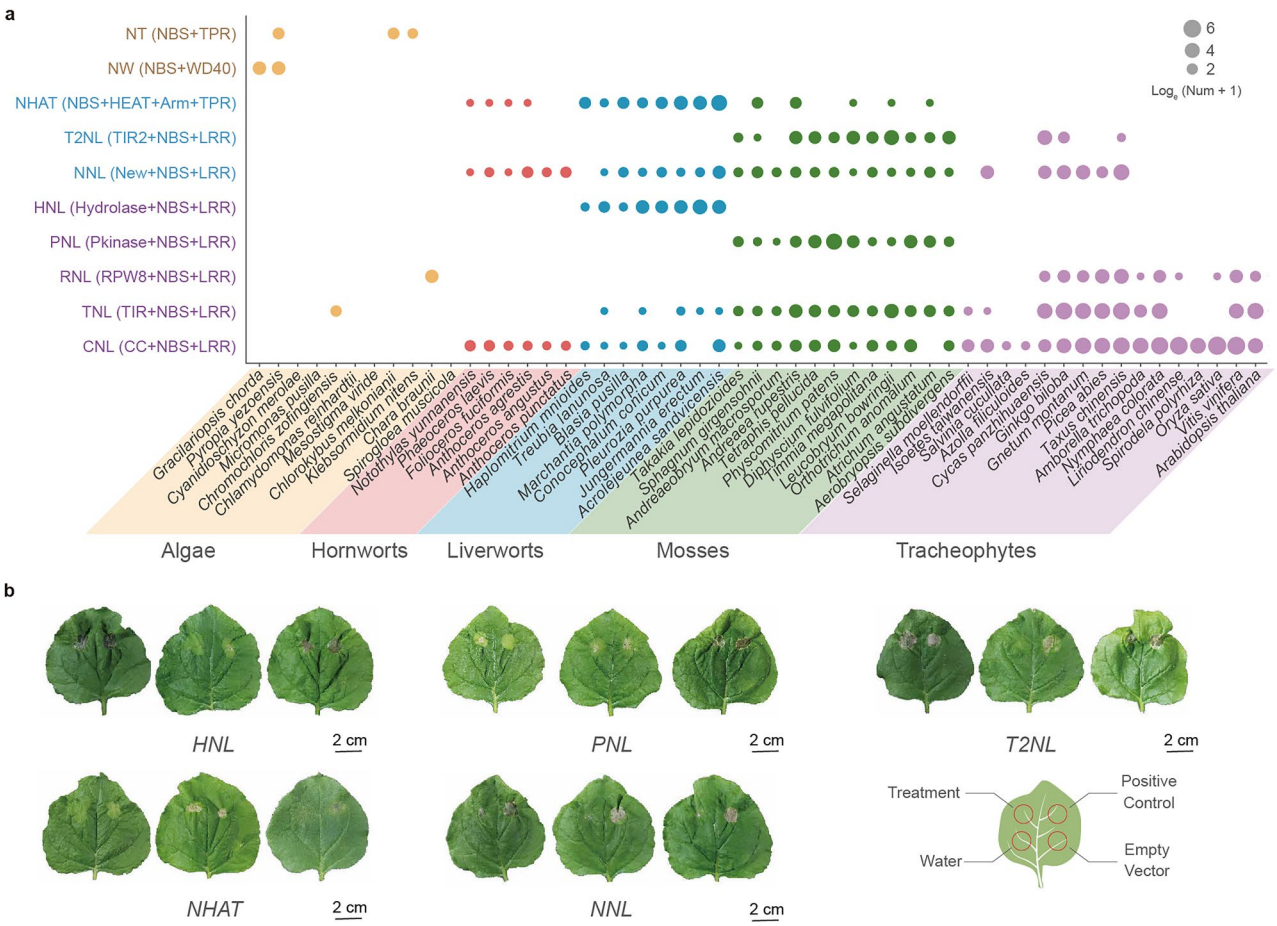
Extended Data Fig. 4 | Characteristics of bryophyte core, accessory and unique gene families. a, The percentage of expressed genes of the core, accessory and unique genes in *P. patens* and *M. polymorpha*. The black triangles above the bar charts indicate the percentage of overall expressed genes of *P. patens* ($n = 58$) and *M. polymorpha* ($n = 80$) transcriptomes from various studies (see details in Methods). **b**, Percentages of expressed genes among the core, accessory and unique gene families across bryophytes ($n = 61$) with transcriptome data. **c**, Percentage of differentially expressed genes (fold change ≥ 2) responding to the three types of stresses in the core, accessory and unique genes in *P. patens* (for results, see Supplementary Data 6). **d**, Percentage of GC content among the core, accessory and unique genes in *P. patens* and *M. polymorpha*. **e**, Proportion of genes with identified InterPro domains in core, accessory and unique gene families in bryophytes and tracheophytes. **f**,**g**, Comparison of gene structure (exon number (**f**) and length of coding sequences (**g**)) among the core, accessory and unique genes in *P. patens* and *M. polymorpha*. **h**, Gene expression level of the core ($n = 19{,}183/11{,}227$), accessory ($4{,}257/2{,}753$) and unique ($5{,}811/3{,}812$) expressed genes in *P. patens* and *M. polymorpha*. **i**, The number of core, accessory and unique gene families getting hits when comparing (BLAST e-value ≤ 1×10^{-5}) to all three categories in *P. patens* and *M. polymorpha*, respectively. **j**, Statistics of BLAST search results of the unique gene families of *M. polymorpha* subsp. *ruderalis* (MPO) against the protein genes and the noncoding regions of genome assemblies of three closely related *Marchantia* species/subspecies with an e-value threshold of 1×10^{-6} and 1×10^{-10} for BLASTP and BLASTN searches, respectively. In **a** and **b**, the error bar denotes the s.d. among different samples. In **a**, **d**–**g**, the number of genes analyzed for *P. patens* is $n = 20{,}506$ (core), $n = 4{,}257$ (accessory) and $n = 5{,}811$ (unique). For *M. polymorpha*, the corresponding numbers are $n = 11{,}723$ (core), $n = 3{,}181$ (accessory) and $n = 4{,}383$ (unique). In **d**, **f**, **g** and **h**, each dot/square denotes the median value, thick solid line indicates the interquartile range (IQR) and the dashed lines represent the whiskers.

https://doi.org/10.1038/s41588-025-02325-9



Extended Data Fig. 5 | Phylogenetic tree of FBT (fungal body lectin domain PF07367 containing) proteins and insect toxicology experiments. a, Maximum likelihood (ML) tree of FBT proteins in bryophytes inferred with IQtree2 with ML method based on the aligned amino acid sequences. The best-fit model was estimated by ModelFinder and branch support values were assessed with UFBoot2 tests. Bootstraps were only shown on stem branches. Subclades of mosses, liverworts and fungi were highlighted in green, blue and orange, respectively. Star labeled *Acrolejeunea sandvicensis* from which the gene *888_jg2710.t1* was used for insect toxicology experiments. **b**, Growth rate (in weight changes) of *Spodoptera frugiperda* ($n = 14 \times 3$ samples per group). **c**, Mortalities of *Helicoverpa armigera* ($n = 3$ independent experiments, each with 14 samples) after treatment with PBS and FBT protein. **d**, Morphologies of *Spodoptera frugiperda* (scale bar = 5 mm). **e**, *Helicoverpa armigera* (scale bar = 5 mm) before (above) and after (below) receiving PBS, and FBT protein treatments. The statistical differences in weight change dynamics (**b**) and mortality rates (**c**) were analyzed with the two-sided Welch's *t* test. The error bars in **b** and **c** denote s.d. of the data. Solid squares (PBS group) and circles (FBT group) in **c** represent the mortality rates from three independent experiments, respectively.

https://doi.org/10.1038/s41588-025-02325-9



Extended Data Fig. 6 | The distribution of resistance (R) genes in green plants, and hypersensitive response (HR) experiments. a, Comparison of the numbers of different types of R genes in green plants. NBS-encoding proteins are divided into 10 major subclasses based on domain structure and phylogenetic relationships (Supplementary Fig. 18). The names of the three types of newly identified resistance gene families are highlighted in light blue, two types specific to algae are in light brown and the five types that were previously reported are in purple. **b**, Macroscopic cell death symptoms induced by overexpression of five selected types of bryophyte R genes (*NNL*: *Anthoceros angustus*, gene ID: *AANG_009324*. *T$_2$NL*: *Funaria hygrometrica*, gene ID: *jg20677.t1*. *NHAT*: *Marchantia polymorpha*, gene ID: *Mapoly0095s0062.1*. *PNL*: *Physcomitrium patens*, gene ID: *Pp3c22_2140V3.3*. *HNL*: *Marchantia polymorpha*, gene ID: *Mapoly0200s0002.1*) in *Nicotiana benthamiana* leaves; each experiment consists of three biological replicates. *XopQ* was used as the positive control; empty vector and water were used as negative controls. Photos were taken seven days after agroinfiltrations. Scale bars are 2 cm for all the leaves.

nature portfolio

Corresponding author(s): Dr. Yang Liu

Last updated by author(s): 2025/08/1

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	The long read data were generated by the Nanopore PromethION sequencer, and the short read data were sequenced by the MGI-SEQ or Illumina HiSeq X platform, with 150-bp read length and 300-500 bp insert size.
Data analysis	The softwares used in this study are listed as follows: 3D-DNA pipeline v180922 (https://github.com/aidenlab/3d-dna) ALEOBSERVE v1.0 (https://github.com/ssolo/ALE) ANNOVAR v2019Oct24 (http://annovar.openbioinformatics.org) ASTRAL v5.7.3 (https://github.com/smirarab/ASTRAL) BCFTOOLS v0.2.0-rc6- 37 -g17d49cf (https://samtools.github.io/bcftools) BEAST v2.4.6 (https://www.beast2.org) BRAKER v2.1.5 (https://github.com/Gaius-Augustus/BRAKER/releases) BISMARK v0.23.1 (https://github.com/FelixKrueger/Bismark) BUSCO v5.3.2 (https://busco.ezlab.org/) COUNT v10.04 (https://sourceforge.net/projects/count/) DIAMOND v0.9.25 (https://github.com/bbuchfink/diamond) EDGER v3.42.2 (https://git.bioconductor.org/packages/edgeR) EGGNOG-MAPPER v2 (http://eggnog-mapper.embl.de/) EMBOSS v6.5.7.0 (http://emboss.sourceforge.net/download/) FEATURECOUNTS v1.34.7 (http://subread.sourceforge.net) GBLOCKS v0.91b (http://molevol.cmima.csic.es/castresana/Gblocks.html) GENOMESCOPE v2.0 (https://github.com/molikd/genomescope2.0)

HISAT2 (https://github.com/DaehwanKimLab/hisat2)
HMMER v3.2.1 (http://hmmer.org/download.html)
I-ADHORE v.3.0.01 (https://github.com/VIB-PSB/i-ADHoRe)
INPARANOID v4.1 (https://inparanoid.sbc.su.se/cgi-bin/index.cgi)
INTERPROSCAN v5.51-85.0 (www.msi.umn.edu/sw/interproscan)
IQTREE2 v2.2.0 (https://github.com/iqtree/iqtree2/releases)
iTAK v1.6 (https://github.com/kentnf/iTAK/releases)
JCVI v.1.1.8 (https://github.com/tanghaibao/jcvi)
JUICEBOX ASSEMBLY TOOLS v1.11.08 (https://github.com/aidenlab/Juicebox)
JUICER v1.6 (https://github.com/aidenlab/juicer)
KEGG (https://www.genome.jp/kegg/)
KINFIN v1.0.3 (https://github.com/DRL/kinfin/releases)
KSRATES v1.1.1 (https://github.com/VIB-PSB/ksrates)
LTR_STRUC (https://mcdonaldlab.biology.gatech.edu/ltr_struc/)
LTR_FINDER v1.0.5 (https://github.com/xzhub/LTR_Finder)
MAFFT v5.0 (https://mafft.cbrc.jp/alignment/software/)
MEGA 7.0 (https://www.megasoftware.net/)
MODELFINDER v1.5.4 (www.iqtree.org/ModelFinder)
MUSCLE v3.8.1551 (https://www.drive5.com/muscle/)
NCBI BLAST+ (https://www.ncbi.nlm.nih.gov/books/NBK279690/)
NEXTDENOVO v2.4.0 (https://github.com/Nextomics/NextDenovo)
NEXTPOLISH v1.4.1 (https://github.com/Nextomics/NextPolish)
OrthoFinder v2.4.0 (https://github.com/davidemms/OrthoFinder)
PHYPARTS v0.0.1 (https://bitbucket.org/blackrim/phyparts/)
PILER v0.4.1 (https://github.com/esamattis/piler)
PILON v1.23 (https://github.com/broadinstitute/pilon)
PUG v2.0 (https://github.com/mrmckain/PUG)
RAxML v8.2.12 (https://pkgs.org/download/raxml)
REPBASE v21.01 (https://www.girinst.org/repbase)
REPEATMASKER v4.1.2 (www.repeatmasker.org/RepeatMasker)
REPEATMODELER v2.0 (www.repeatmasker.org/RepeatModeler)
REPEATSCOUT v1.0.5 (https://github.com/mmcco/RepeatScout)
SAMTOOLS v1.8-20-g4ff8062 (https://sourceforge.net/projects/samtools/)
SMART (http://smart.embl-heidelberg.de/)
SORTADATE v1.0 (https://github.com/FePhyFoFum/SortaDate)
SUPERNOVA v2.1.1 (https://www.getsupernova.com/download)
SWISSPROT (https://www.uniprot.org/)
TopHat2 v2.1.1 (https://anaconda.org/bioinfo/tophat2)
TRIMMOMATIC v0.39 (https://github.com/usadellab/Trimmomatic)
Trinity v2.8.4 (https://github.com/trinityrnaseq/trinityrnaseq/releases)
WGD v1.1.2 (https://github.com/arzwa/wgd)
WHALE v2.0.3 (https://github.com/arzwa/Whale.jl)

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

Genome assemblies and annotations can be found at the website www.bryogenomes.org. The genomes, raw genomic, transcriptomic, and Hi-C data, as well as the whole-genome bisulfite sequencing (WGBS) data for DNA methylation analyses that were generated in this study have been deposited in the NCBI Sequence Read Archive (SRA, BioProject No. PRJNA1117192), and the CNGB data center (https://db.cngb.org/) under the project number CNP0002895. The orthogroup information, data alignments, and phylogenetic trees are available in the Figshare repository, doi: 10.6084/m9.figshare.23528667).

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	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA

Recruitment	NA
Ethics oversight	NA

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

Life sciences study design

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

Sample size	For 123 genome sequencing, a single collection was used for each bryophyte species. This sample size was chosen to cover the current ordinal diversity of bryophytes. For insect toxicology experiment, 14 samples were used for each treatment. This sample size was chosen based on the need to achieve sufficient statistical power to detect significant effects while considering the practical constraints of rearing and testing a large number of insects.
Data exclusions	All bryophyte species sampled were presented, no data were excluded.
Replication	For genome sequencing, the data were generated from a single collection for each species, no replication was applied. For transcriptome sequencing and insect toxicity experiment, three biological replicates were used.
Randomization	For genome sequencing, the data were generated from a single collection for each species, no randomizations were required. For insect toxicology experiment, samples were randomly allocated into three groups for each treatment.
Blinding	For genome sequencing, the data were generated from a single collection for each species, no blinding experiment was performed. For insect toxicology experiment, no blinding experiment was performed. While the treatments were randomly assigned and replicated biologically, the distinct differences in appearance between the toxin and PBS diets made blinding impractical for both the researchers and the insects. However, we ensured experimental rigor through standardized procedures, independent data recording, and robust statistical analysis to minimize potential biases and validate our results.

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	Details for the voucher specimens, such as collection location, specimen number, and herbarium, are provided
Novel plant genotypes	No novel plant genotypes were generated.
Authentication	Not involved.