scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genomes of Arctic and Antarctic mosses: *Aulacomnium turgidum* and *Polytrichastrum alpinum*

Yuying Zeng[1,2,3,6], Yuqing Cai[1,2,3,6], Xuping Zhou[4], Sibo Wang[2,3], Linzhou Li[2,3], Yifeng Yao[5], Jin Yu[1,3], Xin Liu[1], Huanming Yang[1], Tong Wei[2,3], Shanshan Dong[4] & Yang Liu[2,3,4] ✉

Bryophytes play a crucial role in the ecosystems of polar regions. These simple plants are among the predominant vegetation types in both Arctic and Antarctic landscapes, where they contribute significantly to biodiversity and ecological stability. Here, we report the chromosome-level genomes of two polar moss species, the Arctic *Aulacomnium turgidum* and Antarctic *Polytrichastrum alpinum*. Utilizing a combination of Illumina short reads, Nanopore long reads, and Hi-C data, we assembled genomes of 277.84 Mb for *A. turgidum* and 498.33 Mb for *P. alpinum*, respectively. These assemblies were anchored to 11 chromosomes for *A. turgidum* and 8 chromosomes for *P. alpinum*. Both species exhibited a sex chromosome with distinct genomic characteristics. Gene annotations revealed 25,999 protein-coding genes in *A. turgidum* and 28,070 in *P. alpinum*. The high completeness of the gene space was validated via BUSCO, achieving impressive scores of 98.2% and 98.0%. These high-quality genomes provide critical resources for studying the adaptive evolution and stress tolerance mechanisms of mosses in extreme polar environments.

Background & Summary

Bryophytes, including hornworts, liverworts, and mosses, form the sister group to vascular plants, and exhibit a haploid-dominated lifecycle with unbranched sporophytes remain attaching to and reliant on gametophytes[1]. Bryophytes excel in enduring some of the most severe environmental conditions, ranging from low light intensity and extreme temperatures to nutrient scarcity and desiccation. As a result, they often act as pioneering species within various ecosystems. Notably, bryophytes are the predominant flora in Antarctica's terrestrial ecosystems[2] and can also thrive in arid deserts[3]. These adaptabilities may stem from an evolutionarily refined genetic toolkit geared towards stress tolerance[4].

Aulacomnium turgidum (Aulacomniaceae, Rhizogoniales), a moss species widespread across the Arctic Circle, including Svalbard, Greenland, and Alaska, is notable for its ability to regenerate after being entombed in ice for 400 years, demonstrating exceptional freeze stress tolerance[5]. The *A. turgidum* (Fig. 1a) was collected from Svalbard, a Norwegian archipelago, epitomizes a typical Arctic environment. The annual mean surface air temperatures vary from 2.5 to 5.8 °C in summer and to −11.4 to −9.4 °C in winter[6]. Despite these extreme conditions, resilient plants, including bryophytes, flourish due to their high tolerance for abiotic stresses[7,8].

Polytrichastrum alpinum (Polytrichaceae, Polytrichales) is a common Antarctic moss species, and thrives in moist, rocky environments near glacier moraine peaks, as well as in dry areas[2]. The *P. alpinum* sample (Fig. 1c) was collected from King George Island, located at the northern tip of the Antarctic Peninsula, an area characterized by its semi-desert landscape[9]. The average annual temperature of this area is −2.3 °C, with summer months

[1]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 101408, China. [2]BGI Research, Wuhan, 430074, China. [3]State Key Laboratory of Agricultural Genomics, BGI Research, Shenzhen, 518083, China. [4]Key Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen, 518004, China. [5]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, No. 20 Nanxincun, Xiangshan, Beijing, 100093, China. [6]These authors contributed equally: Yuying Zeng, Yuqing Cai. ✉e-mail: yang.liu0508@gmail.com

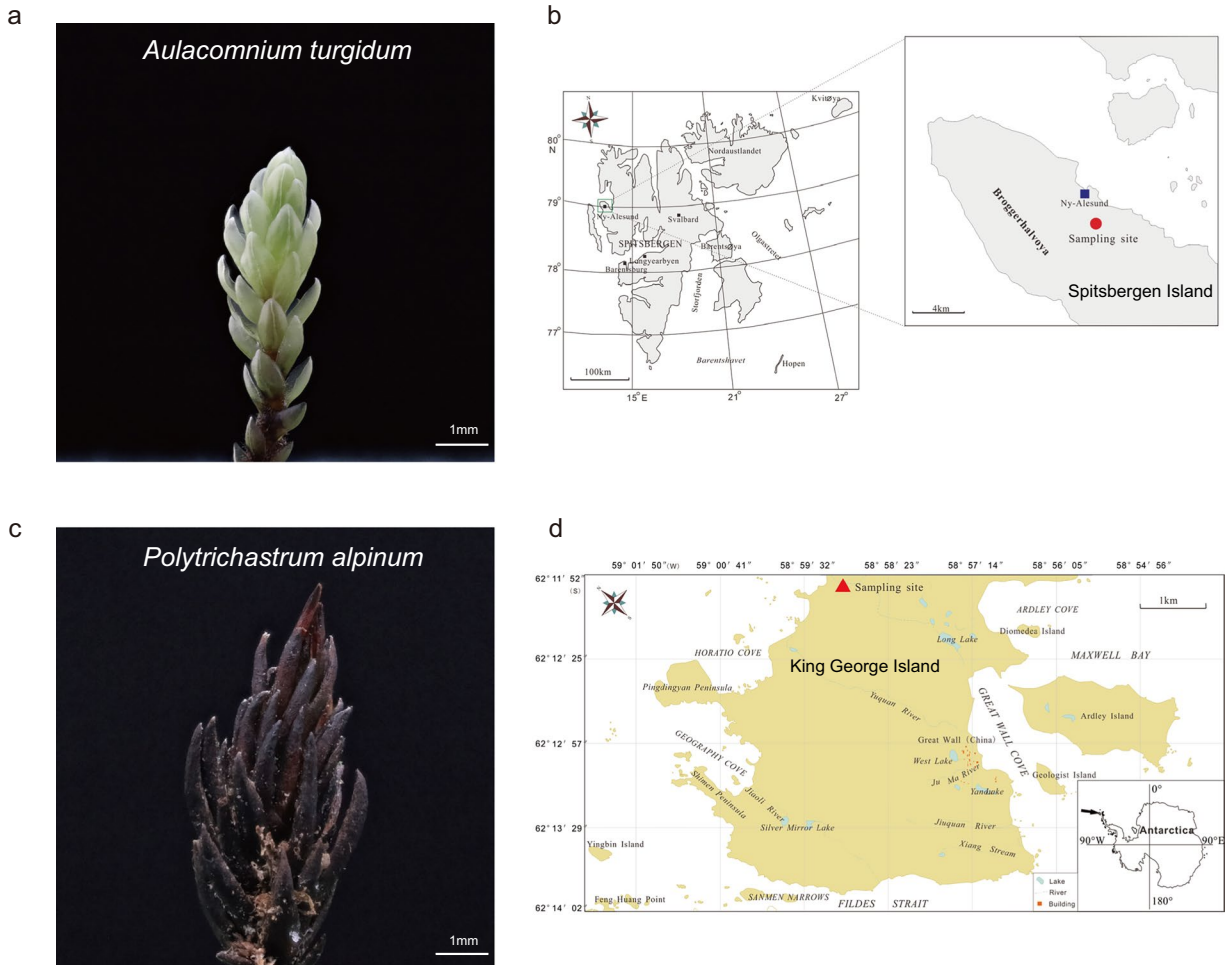

Fig. 1 (**a**) *Aulacomnium turgidum* and (**b**) geographical map showing its sampling site in the Arctic. (**c**) *Polytrichastrum alpinum* and (**d**) geographical map showing its sampling site in the Antarctic.

experiencing temperatures slightly above freezing. The region is subject to frequent strong winds[10]. The island is home to 64 documented species of moss, which are primarily found in humid, sheltered areas with relatively stable and partially organic soils[11,12].

Over the recent decades, research has utilized mosses including *A. turgidum* and *P. alpinum* from the Arctic and Antarctic regions to study the impacts of global warming on Earth[13,14], the interactions between polar microbial communities and plants[15,16], and the molecular mechanisms behind unique survival strategies in extreme conditions[17,18]. Despite the ecological significance of polar mosses, our understanding of their genomes remains limited[19]. These mosses, which thrive under the harsh conditions of the Arctic and Antarctic, are critical for understanding resilience and adaptability to extreme environments. However, the scarcity of complete genomic data hinders our ability to fully comprehend the molecular mechanisms that underpin their unique survival strategies. This gap in knowledge underscores the need for enhanced genomic research to better exploit the potential of polar mosses in studying environmental adaptation and climate change resilience.

In this study, we assembled high-quality chromosome-level genomes of *A. turgidum* and *P. alpinum* using a combination of Illumina short reads, Nanopore long reads, and high-throughput chromosome conformation capture (Hi-C) data. The genome of *A. turgidum* was assembled as 277.84 Mb with a contig N50 of 11.92 Mb, while the genome of *P. alpinum* was 498.33 Mb with a contig N50 of 4.24 Mb (Table 1). A total of 275.60 Mb and 488.51 Mb of the assemblies were anchored to 11 and 8 chromosomes for *A. turgidum* and *P. alpinum*, respectively (Table 2). Both species possess a sex chromosome with lower gene density and higher repetitive sequence density than the autosomes. *A. turgidum* and *P. alpinum* respectively encodes 25,999 and 28,070 protein-coding genes (Table 3). These two high-quality genomes offer valuable new genomic resources for future research into the genetic foundations and adaptive evolution of plants in Arctic and Antarctic environments.

Methods

Plant materials and sequencing. Wild gametophytes of *A. turgidum* were collected from Spitsbergen Island, Svalbard (78°54′41″ N, 11°58′35″ E) (Fig. 1b) on September 18th, 2018, and *P. alpinum* from King George Island, Antarctica (62°12.041” S, 58°59.698” W) (Fig. 1d) on December 23rd, 2018. Voucher specimens were deposited in the Herbarium of Shenzhen Fairy Lake Botanical Garden in Shenzhen, China (SZG). The entire

Assembly	*Aulacomnium turgidum*	*Polytrichastrum alpinum*
Estimated genome size by k-mer (Mb)	272.00	449.50
Total scaffolded assembly size (Mb)	277.84	498.33
Contig N50 (bp)	11,929,215	4,244,457
Longest contig (bp)	33,957,313	17,475,254
Scaffold N50 (bp)	23,694,134	54,366,491
Largest scaffold (bp)	40,940,268	97,437,522
GC content (%)	38.79	40.96
L50	5	3
Chromosome number	11	8
Completeness BUSCOs (%)	98.2	98
Quality values (QV)	30.14	29.59
Assembly completeness (%)	99.42	93.06

Table 1. Characteristics of genome assemblies and genome size estimates for *Aulacomnium turgidum* and *Polytrichastrum alpinum*.

	Chr ID	Contig number	Chr length (bp)
Aulacomnium turgidum	Chr01	5	40,940,268
	Chr02	5	36,916,288
	Chr03	2	34,644,130
	Chr04	2	23,694,134
	Chr05	2	21,801,986
	Chr06	2	19,455,661
	Chr07	2	18,398,267
	Chr08	1	18,375,383
	Chr09	4	18,369,948
	Chr10	2	17,065,196
	Chr11	12	25,965,726
	Chr ID	**Contig number**	**Chr length (bp)**
Polytrichastrum alpinum	Chr1	37	97,437,522
	Chr2	31	89,389,019
	Chr3	32	67,256,542
	Chr4	21	60,794,430
	Chr5	21	54,366,491
	Chr6	20	52,110,541
	Chr7	19	51,807,611
	Chr8	30	15,351,097

Table 2. Summary of the assembled chromosomes of *A. turgidum* and *P. alpinum*.

moss plant, including its leaves, stems, and rhizoids, was used for sequencing. Approximately 5 grams of material were gathered from these samples for genomic and transcriptomic sequencing, respectively. Three distinct types of genome sequencing techniques were performed: the Nanopore long-read sequencing, the Hi-C sequencing, and the Illumina short-read sequencing. Additionally, plant tissues were utilized for transcriptomic sequencing. Transcriptome libraries were constructed using a TruSeq RNA Library Prep Kit v2 (Illumina, CA, USA), with an insert size ranging from 200 to 400 bp, following polyA selection. Sequencing was performed on an Illumina NovaSeq 6000 platform (Illumina, CA, USA), generating 150-bp paired-end reads.

Genome assembly. *Illumina short-read sequencing and genome survey.* The Illumina short reads were processed using Trimmomatic (v0.39)[20] to filter out duplicates, low-quality reads, and adapters. Filtered reads were used for *k-mer* analyses to estimate genome sizes. Jellyfish (v2.0.0)[21] was employed to count the *k-mer* frequencies with the '-m 21' parameter. GenomeScope (v2.0)[22] was used to estimate genome sizes with the settings "-k 21 -p 1" along with other default parameters. These analyses predicted that the genome sizes for *A. turgidum* and *P. alpinum* are 272.00 Mb and 449.50 Mb, respectively, with the proportion of repetitive sequences at 29.52% and 40.00%, respectively (Fig. 2).

ONT long-read sequencing and genome assembly. Nanopore libraries were prepared using SQK-LSK108 and sequenced on a Nanopore PromethION sequencer. The long reads obtained were assembled using NextDenovo

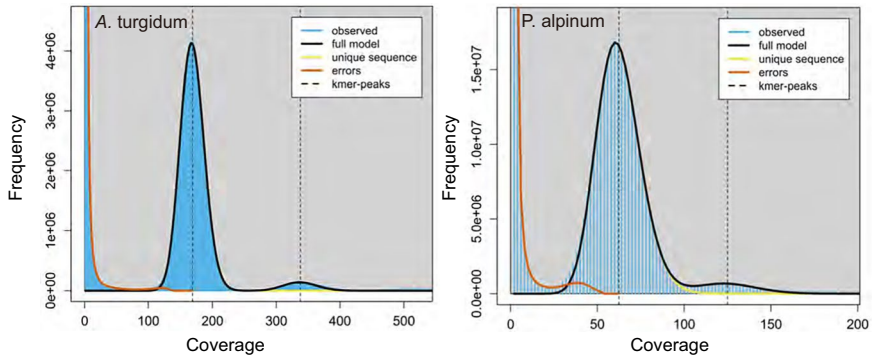

Fig. 2 Genome survey based on k-mer (k = 21) analysis of *A. turgidum* and *P. alpinum*. The X-axis represents the k-mer depth, while the Y-axis indicates the k-mer frequency for a given depth.

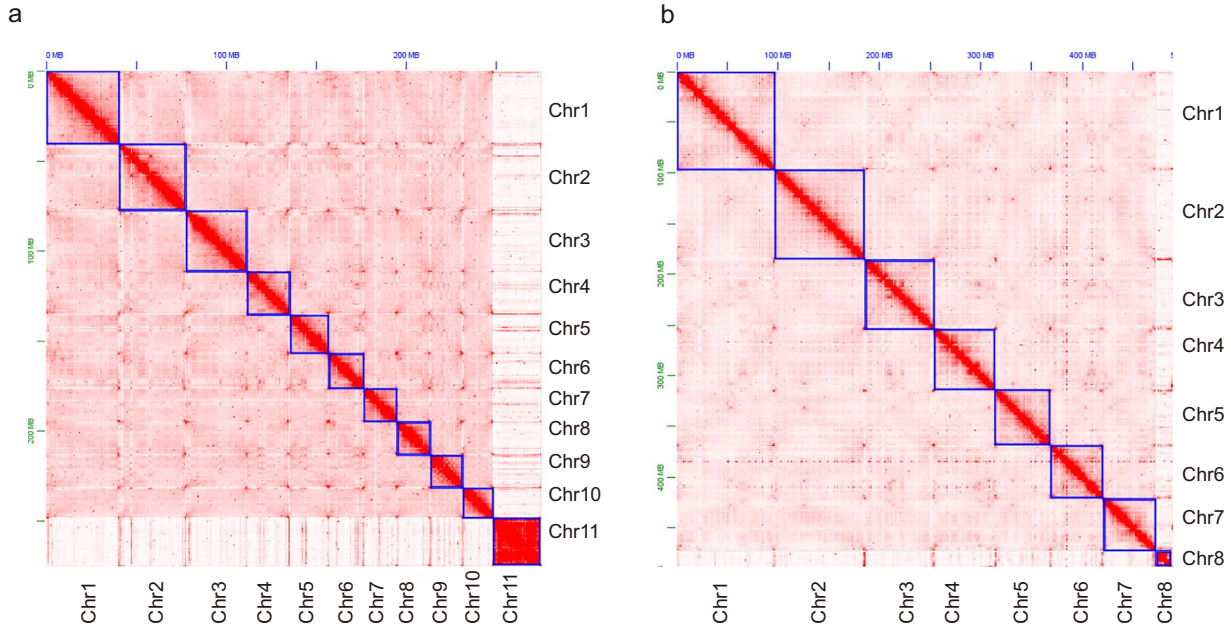

Fig. 3 Hi-C interaction heatmap for *A. turgidum* and *P. alpinum*. The heatmap illustrates the density of Hi-C interactions among distinct chromosomes, showing scaffolded and independently assembled chromosomes at high resolution. The intensity of red color reflects to the number of Hi-C interactions.

(v2.5.0)[23]. The initial assembly was then polished for three iteratives with NextPolish (v1.3.1)[24], using both Nanopore long reads and filtered Illumina reads.

Hi-C scaffolding and chromosome-level genome assembly. The Hi-C library construction involved several steps: cross-linking, restriction enzyme digestion (using MboI), end repair, DNA cyclization, and purification[25]. Paired-end 150-bp reads were generated on an Illumina NovaSeq 6000 platform (Illumina, CA, USA). The raw Hi-C reads were initially processed with Trimmomatic (v0.39)[20] using default parameters. Subsequently, Juicer (v1.6)[26] was employed to extract valid data. Misassembled contigs were corrected, anchored, ordered, and oriented using the 3D-DNA pipeline (v 180922)[27]. Juicebox (v1.11.08)[28] was then utilized for manual adjustments of the anchored results. For scaffolds not anchored to chromosomes, we performed the procedure outlined by Zhou *et al.*[29] to remove those containing contamination or organelle fragments, and to link consecutive contigs to generate a high-quality genome assembly (Fig. 3). This process ultimately enabled the identification of 11 chromosomes for *A. turgidum* and 8 chromosomes for *P. alpinum*, resulting in final chromosome-scale genome assemblies of 277.84 Mb for *A. turgidum* and 498.33 Mb for *P. alpinum* (Table 1).

Repeat annotation. For repeat annotation, a customized *de novo* repeat library was created using a homology-based approach. Programs Piler (v0.4.1)[30], LTR_FINDER (v1.0.5)[31], RepeatScout (v1.0.5)[32], and RepeatModeler (v2.0)[33] were used to generate the libraries. The resulting repetitive sequence libraries were combined and used as input for RepeatMasker (v.4.1.1)[34]. Additionally, Repbase (v21.01)[35] served as the database for

Annotation	*A. turgidum*	*P. alpinum*
Number of predicted protein-coding genes	25,999	28,070
Average gene length (bp)	3,017.35	3,074.81
Average exon length (bp)	315.22	249.23
Average exon number per gene	6.46	4.58
Average intron length (bp)	179.36	519.72
Averrage intron number per gene	5.46	3.58
Average CDS length (bp)	223.00	249.25
Average CDS number per gene	6.15	4.63
Percentage of repeat sequence (%)	42.65	53.37
LTR (%)	15.45	25.16
LINE (%)	4.66	8.54
SINE (%)	0.18	0.44
DNA transposons (%)	20.91	10.72

Table 3. Genome annotation of *A. turgidum* and *P. alpinum*.

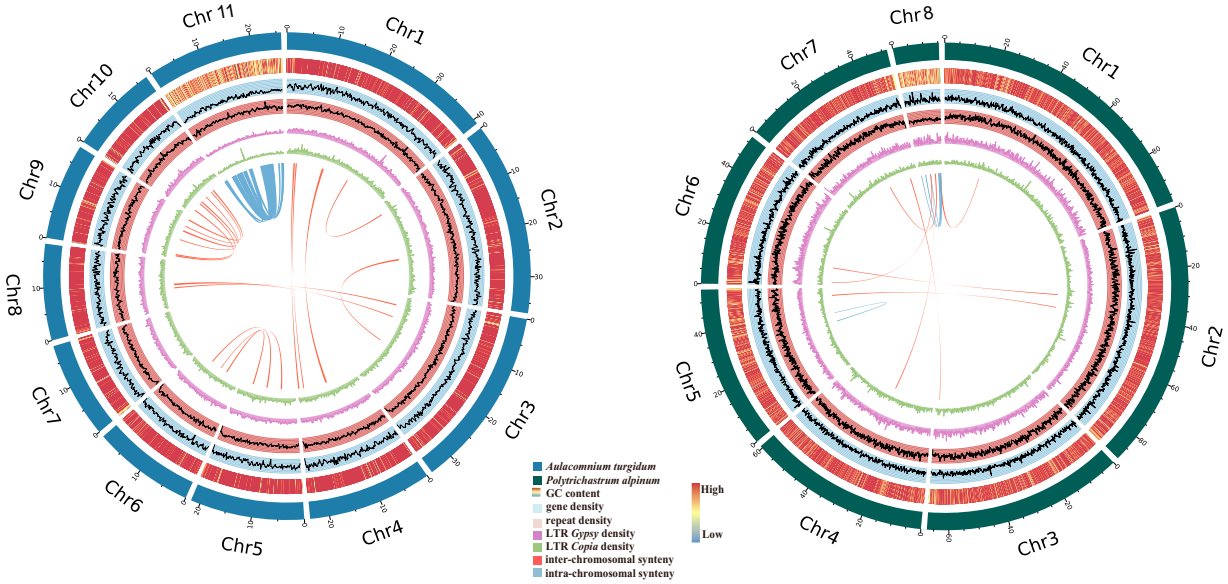

Fig. 4 Circos plots illustrating the chromosomal-level genome assemblies for *A. turgidum* (left) and *P. alpinum* (right). The tracks, from outermost to innermost, are: chromosomes, GC content, gene density, TE density, LTR/Gypsy density, LTR/Copia density and color ribbons representing genome-wide syntenic blocks of inter-chromosome and intra-chromosome. The color bar indicates the proportion of GC content.

known repetitive elements, searched using RepeatMasker (v.4.1.1)[34]. Tandem Repeats Finder (TRF v4.07)[36] was employed to predict tandem repeat sequences across the genome. The results indicated that the genomes of *A. turgidum* and *P. alpinum* contained 42.65% and 53.37% repetitive elements, respectively. The percentages of long terminal repeat (LTR) retrotransposons were 15.45% for *A. turgidum* and 25.16% for *P. alpinum* (Table 3 and Fig. 4).

Gene and functional annotation. To predict protein-coding genes (PCGs) for *A. turgidum* and *P. alpinum*, we integrated two types of evidence. We sourced proteome sequences from *Physcomitrium patens*[37], *Marchantia polymorpha* (https://marchantia.info/), *Arabidopsis thaliana* (https://www.arabidopsis.org/), and the bryophyte proteins from Swiss-Prot (https://www.uniprot.org/) to provide homology-based protein evidence. Transcriptome evidence involved mapping clean illumina reads to the assembled genomes using HISAT (v2.2.0)[38] to produce BAM files. Gene models were then predicted using BRAKER3 pipeline (v3.0.6)[39], in conjunction with AUGUSTUS[40] and GeneMark[41] *de novo* predictions, based on the soft-masked genome, protein evidence, and transcriptome data. Additionally, transcripts were generated using StringTie (v2.2.1)[42] and TransDecoder (v5.5.0) (https://github.com/TransDecoder/TransDecoder). These pieces of evidences were integrated using EVidenceModeler (v2.1.0)[43]. Subsequently, Trinity (v2.8.4)[44] was used for *de novo* transcript assembly, updating gene models to include untranslated regions (UTRs) and alternative splicing variants via the PASA pipeline (v2.5.3)[45]. A total of 25,999 and 28,070 protein-coding genes were predicted in *A. turgidum* and *P. alpinum*, respectively (Table 3). The completeness of the gene space was evaluated using the BUSCO v.3.1.0, based

Gene function annotation	*A. turgidum*	*P. alpinum*
InterProScan (%)	55.98	50.72
Gene Ontology (%)	51.25	44.81
Kyoto Encyclopedia of Genes and Genomes (%)	60.81	56.84
Swiss-prot (%)	52.3	44.8
TrEMBL (%)	71.52	67.69
Transcription factors (%)	2.79	2.49
With Physcomitrium patens blast Hit (%)	67.6	56.38
With Arabidopsis thaliana blast Hit (%)	56.71	49.01
Percentage of overall annotated genes (%)	74.56	72.11

Table 4. Overview of gene functional annotations in *Aulacomnium turgidum* and *Polytrichastrum alpinum*.

	Class	Type	Number	Average length(bp)	Total length(bp)	% of genome
A. turgidum	miRNA		14	116.64	1,633	0.000588
	tRNA		284	74.59	21,184	0.007625
	rRNA	rRNA	132	251.67	33,220	0.011957
		18S	38	623.53	23,694	0.008528
		28S	40	124.35	4,974	0.00179
		5.8S	11	104.27	1,147	0.000413
		5S	43	79.19	3,405	0.001226
	snRNA	snRNA	235	143.83	33,799	0.012165
		CD-box	60	110.20	6,612	0.00238
		HACA-box	4	117.00	468	0.000168
		splicing	171	156.25	26,719	0.009617
	Class	**Type**	**Number**	**Average length(bp)**	**Total length(bp)**	**% of genome**
P. alpinum	miRNA		7	109.71	768	0.000154
	tRNA		568	75.31	42,775	0.008584
	rRNA	rRNA	716	104.76	75,011	0.015052
		18S	624	108.19	67,511	0.013547
		28S	33	96.85	3,196	0.000641
		5.8S	8	96.88	775	0.000156
		5S	51	69.20	3,529	0.000708
	snRNA	snRNA	608	141.46	86,005	0.017259
		CD-box	135	116.08	15,671	0.003145
		HACA-box	3	119.33	358	0.000072
		splicing	470	148.89	69,976	0.014042

Table 5. Non-coding RNA classification and genome proportion in *Aulacomnium turgidum* and *Polytrichastrum alpinum*.

on the Viridiplantae_odb10 set[46]. For functional annotation, gene models were aligned against the UniProt (Swiss-Prot and TrEMBL), KEGG, and TAIR databases. Protein domains and gene ontologies were annotated using InterProScan (v5.51–85.0)[47]. Additionally, the iTAK online tool (v1.6)[48] was used to identify transcription factors, predicting 727 transcription factors in *A. turgidum* and 698 in *P. alpinum*, (Table 4, Tables S1, S2).

Non-coding RNA annotation. Transfer RNAs (tRNAs) were identified using tRNAscan-SE (v1.3.1)[49] with default parameters. Given the high conservation of ribosomal RNAs (rRNAs), rRNA sequences from closely related species were downloaded from the Ensembl database and used as references in BLAST (v2.2.26)[50] search with an e-value threshold of 1e-5. Other non-coding RNAs (ncRNAs), including micro RNAs (miRNAs) and small nuclear RNAs (snRNAs), were identified by searching against the Rfam database (v12.0)[51] using Infernal (v1.1.1)[52] with default parameters. In total, 1,032 and 3,223 ncRNAs were identified in *A. turgidum* and *P. alpinum*, respectively (Table 5).

Genome synteny analysis and Circos diagram construction. The python library jcvi (v1.1.8)[53] was employed to identify intra-genomic syntenic blocks and detect inter- and intra-chromosomal synteny (Table S3). The syntenic regions were first identified using the jcvi.compara.catalog ortholog function with a default C-score threshold of 0.7 to filter out low-quality hits. Syntenic depths were then calculated using the jcvi.compara.synteny depth function. To further refine the syntenic blocks, we applied the jcvi.compara.synteny screen function with the parameters–minspan = 30–simple. Visualization of syntenic relationships and genomic feature distributions

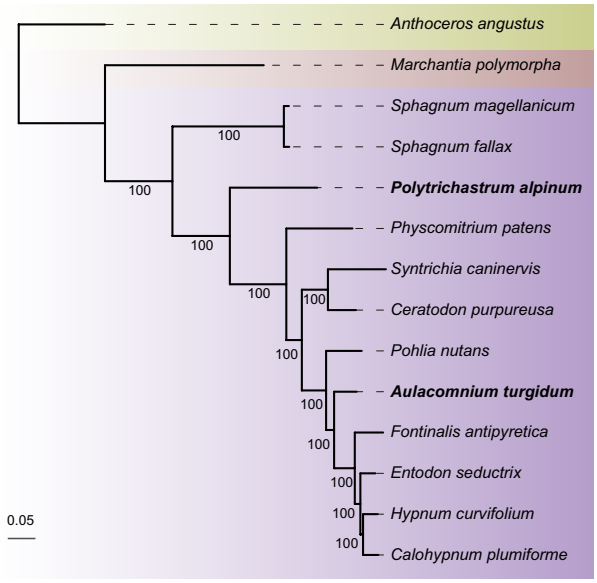

Fig. 5 A maximum likelihood phylogenetic tree of bryophytes based on 61 single-copy ortholog protein sequences from 14 bryophyte species. The newly sequenced two mosses are highlighted in bold.

was carried out using Circos[54]. The Circos plot (Fig. 4) includes synteny blocks, GC content, gene density, transposable element (TE) density, LTR/*Gypsy* density, and LTR/*Copia* density, with all features calculated in 1 Mb windows.

Phylogenetic reconstruction. A total of 14 bryophyte species, including the newly sequenced *A. turgidum* and *P. alpinum*, were sampled to identify orthologs using OrthoFinder (v2.3.11)[55]. Alignment for protein sequences of 61 single-copy orthologs was performed using MAFFT (v7.453)[56]. The resulting alignments were then concatenated and used as the input data for IQ-TREE2 (v2.0.6)[57] to construct a maximum likelihood phylogenetic tree, with the JTT model and 1,000 ultrafast bootstrap replicates. The resulting phylogenetic tree was rooted with *Anthoceros angustus*, and visualized using the interactive Tree of Life (iTOL)[58] (Fig. 5).

Data Records

The raw data from Nanopore, Hi-C, Illumina short-read sequencing used for genome assembly and annotation have been deposited in the Genome Sequence Archive (GSA) of the National Genomics Data Center (NGDC) with the accession number CRA017596[59] under the BioProject accession number PRJCA027760[60]. All the genomic sequencing raw data were also deposited in the China National GeneBank Database (CNGB) Nucleotide Sequence Archive (CNSA) under accession numbers CNP0002895[61]. The final contigs and chromosome assemblies are submitted to NCBI under the accession number GCA_048933245.1[62], GCA_048933195.1[63] of *A. turgidum* and *P. alpinum*, respectively. The contigs and chromosome-scale genome assemblies have also been made available in the GSA at the NGDC. The specific accession numbers for sequences of *A. turgidum* and *P. alpinum* are GWHEUUP00000000.1[64] and GWHEUUQ00000000.1[65], respectively. The annotation files are available in figshare[66].

Technical Validation

The completeness and accuracy of the genome assemblies for *A. turgidum* and *P. alpinum* were assessed using multiple approaches. First, Benchmarking Universal Single-Copy Orthologs (BUSCO)[46] analyses with the Viridiplantae_odb10 indicated completeness scores of 98.20% for *A. turgidum* and 98.00% for *P. alpinum*. Additionally, QUAST[67] was employed to evaluate assembly contiguity metrics, revealing final assembly size of 277.84 Mb for *A. turgidum* with a L50 value of 5, and 498.33 Mb for *P. alpinum* with a L50 value of 3. To further assess assembly accuracy, Merqury[68] was used, yielding quality values (QV) of 30.14 for *A. turgidum* and 29.59 for *P. alpinum*, with assembly completeness rates of 99.42% and 93.06%, respectively. The interaction contact pattern, centered along the principal diagonal in the Hi-C heatmap (Fig. 3), further supports the accuracy of the chromosome -level assemblies.

Code availability

All codes and pipelines used for data processing was executed following the manual and protocols of the respective bioinformatic tools. Detailed information on the software versions is provided in the Methods section. No custom code was developed for this study.

Received: 20 August 2024; Accepted: 7 April 2025;
Published online: 29 April 2025

References

1. Goffinet, B. & Buck, W. R. The evolution of body form in bryophytes. *Annual Plant Reviews*, 51–89 (2013).
2. Ochyra, R., Bednarek-Ochyra, H. & Smith, R. I. L. *The Illustrated Moss Flora Of Antarctica* (Cambridge University Press, 2008).
3. Pan, Z. *et al*. The upside-down water collection system of *Syntrichia caninervis*. *Nature Plants* **2**, 16076 (2016).
4. Kulshrestha, S. *et al*. Stress, senescence, and specialized metabolites in bryophytes. *Journal of Experimental Botany* **73**, 4396–4411 (2022).
5. La Farge, C., Williams, K. H. & England, J. H. Regeneration of little ice age bryophytes emerging from a polar glacier with implications of totipotency in extreme environments. *Proceedings of the National Academy of Sciences* **110**, 9839–9844 (2013).
6. Maturilli, M., Herber, A. & König-Langlo, G. Climatology and time series of surface meteorology in Ny-Alesund, Svalbard. *Earth System Science Data* **5**, 155–163 (2013).
7. Crawford, R. M. M., Chapman, H. M. & Hodge, H. Anoxia tolerance in high arctic vegetation. *Arctic and Alpine Research* **26**, 308 (1994).
8. Liu, S. *et al*. Complementary DNA library construction and expressed sequence tag analysis of an arctic moss, *Aulacomnium turgidum*. *Polar Biology* **33**, 617–626 (2010).
9. Rakusa-Suszczewski, S. *King George Island — South Shetland Islands, Maritime Antarctic* (Springer Berlin Heidelberg, 2002).
10. Oliva, M. *et al*. Recent regional climate cooling on the antarctic peninsula and associated impacts on the cryosphere. *Science of The Total Environment* **580**, 210–223 (2017).
11. Ochyra, R. *The Moss Flora Of King George Island, Antarctica* (Polish Academy of Sciences, 1998).
12. Li, S. *et al*. *Drepanocladus longifolius* (Amblystegiaceae), an addition to the moss flora of king george island, south shetland islands, with a review of antarctic benthic mosses. *Polar Biology* **32**, 1415–1425 (2009).
13. Shortlidge, E. E. *et al*. Passive warming reduces stress and shifts reproductive effort in the antarctic moss, *Polytrichastrum alpinum*. *Annals of Botany* **119**, 27–38 (2017).
14. Koncz, P., Hermanutz, L., Marino, P., Wheeler, J. & Cranston, B. Bryophyte community diversities and expected change under a warming climate in contrasting habitats of the Torngat Mountains, Labrador. *The Bryologist* **121**, 174–182 (2018). 179.
15. Koranda, M., Rinnan, R. & Michelsen, A. Close coupling of plant functional types with soil microbial community composition drives soil carbon and nutrient cycling in tundra heath. *Plant and Soil* **488**, 551–572 (2023).
16. de Carvalho, C. R. *et al*. Cultivable fungi associated with bryosphere of bipolar mosses *Polytrichastrum alpinum* and *Polytrichum juniperinum* in King George Island, South Shetland Islands, Maritime Antarctica. *Polar Biology* **43**, 545–553 (2020).
17. Alavilli, H., Lee, H., Park, M., Yun, D. J. & Lee, B. H. Enhanced multiple stress tolerance in Arabidopsis by overexpression of the polar moss peptidyl prolyl isomerase *FKBP12* gene. *Plant Cell Reports* **37**, 453–465 (2018).
18. Kang, P., Yoo, Y.-H., Kim, D.-I., Yim, J. H. & Lee, H. De novo transcriptome assembly and comparative analysis of differentially expressed genes involved in cold acclimation and freezing tolerance of the arctic moss *Aulacomnium turgidum* (Wahlenb.) Schwaegr. *Plants* **12**, 1250 (2023).
19. Liu, S. *et al*. The antarctic moss *Pohlia nutans* genome provides insights into the evolution of bryophytes and the adaptation to extreme terrestrial habitats. *Frontiers in Plant Science* **13**, 920138 (2022).
20. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
21. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
22. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* **11**, 1432 (2020).
23. Hu, J. *et al*. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biology* **25**, 107 (2024).
24. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2019).
25. Yu, J. *et al*. Chromosome-Level genome assemblies of two Hypnales (mosses) reveal high intergeneric synteny. *Genome Biology and Evolution* **14** (2022).
26. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems* **3**, 95–98 (2016).
27. Dudchenko, O. *et al*. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
28. Durand, N. C. *et al*. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems* **3**, 99–101 (2016).
29. Zhou, X. *et al*. Chromosome-level genome assembly of *Niphotrichum japonicum* provides new insights into heat stress responses in mosses. *Frontiers in Plant Science* **14** (2023).
30. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**, 152–158 (2005).
31. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* **35**, W265–W268 (2007).
32. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**(Suppl 1), i351–358 (2005).
33. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 9451–9457 (2020).
34. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **25**, 4.10.11–14.10.14 (2009).
35. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
36. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573–580 (1999).
37. Bi, G. *et al*. Near telomere-to-telomere genome of the model plant *Physcomitrium patens*. *Nature Plants* **10**, 327–343 (2024).
38. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907–915 (2019).
39. Gabriel, L. *et al*. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research* **34**, 769–777 (2024).
40. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**, W435–439 (2006).
41. Bruna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics* **2** (2020).
42. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**, 290–295 (2015).
43. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology* **9**, R7 (2008).
44. Haas, B. J. *et al*. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494–1512 (2013).

45. Haas, B. J. *et al*. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654–5666 (2003).
46. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral Genomes. *Molecular Biology and Evolution* **38**, 4647–4654 (2021).
47. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
48. Zheng, Y. *et al*. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Molecular Plant* **9**, 1667–1670 (2016).
49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25**, 955–964 (1997).
50. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).
51. Nawrocki, E. P. *et al*. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research* **43**, D130–D137 (2014).
52. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
53. Tang, H. *et al*. JCVI: a versatile toolkit for comparative genomics analysis. *Imeta* **3**, e211 (2024).
54. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome Research* **19**, 1639–1645 (2009).
55. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).
56. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
57. Minh, B. Q. *et al*. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).
58. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2006).
59. *NGDC Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA017596 (2025).
60. *NGDC BioProject* https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA027760 (2025).
61. *CNGB Nucleotide Sequence Archive* https://db.cngb.org/search/project/CNP0002895 (2025).
62. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_048933245.1 (2025).
63. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_048933195.1 (2025).
64. *NGDC Genome Warehouse* https://ngdc.cncb.ac.cn/gwh/Assembly/85929/show (2025).
65. *NGDC Genome Warehouse* https://ngdc.cncb.ac.cn/gwh/Assembly/85930/show (2025).
66. Zeng, Y. Y. The annotation files of *Aulacomnium turgidum* and *Polytrichastrum alpinum*. *figshare*. https://doi.org/10.6084/m9.figshare.28595150 (2025).
67. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
68. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**, 245 (2020).

Acknowledgements

This study was supported by the 10KP project (https://db.cngb.org/10kp/) and China National GeneBank (CNGB; https://www.cngb.org/). Additionally, this study was supported by the Scientific Foundation of the Urban Management Bureau of Shenzhen (202005, 202403 to Y.L., and 202106, 202302 to S.D.).

Author contributions

Y.L., S.D., H.Y. and T.W. conceived the study. Y.Y. collected the materials. Y.Z., Y.C., X.Z., J.Y. and X.L. performed the analyses. Y.Z. drafted the manuscript. All authors read and approved the manuscript. Y.Z. and Y.C. contributed equally to this work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-04960-7.

Correspondence and requests for materials should be addressed to Y.L.