Metagenome-assembled genome of the glacier alga Ancylonema yields insights into the evolution of streptophyte life on ice and land

Alexander M. C. Bowles\textsuperscript{1,2}, Tom A. Williams\textsuperscript{2}, Philip C. J. Donoghue\textsuperscript{2}, Douglas A. Campbell\textsuperscript{3} and Christopher J. Williamson\textsuperscript{1}

\textsuperscript{1}School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK; \textsuperscript{2}Bristol Palaeobiology Group, School of Biological Sciences and School of Earth Sciences, Life Sciences Building, University of Bristol, Bristol, BS8 1TQ, UK; \textsuperscript{3}Department of Biology, Mount Allison University, Sackville, NB, E4L 1H3, Canada

Summary

- Contemporary glaciers are inhabited by streptophyte algae that balance photosynthesis and growth with tolerance of low temperature, desiccation and UV radiation. These same environmental challenges have been hypothesised as the driving force behind the evolution of land plants from streptophyte algal ancestors in the Cryogenian (720–635 million years ago).
- We sequenced, assembled and analysed the metagenome-assembled genome of the glacier alga Ancylonema nordenskiöldii to investigate its adaptations to life in ice, and whether this represents a vestige of Cryogenian exaptations.
- Phylogenetic analysis confirms the placement of glacier algae within the sister lineage to land plants, Zygnematophyceae. The metagenome-assembled genome is characterised by an expansion of genes involved in tolerance of high irradiance and UV light, while lineage-specific diversification is linked to the novel screening pigmentation of glacier algae. We found no support for the hypothesis of a common genomic basis for adaptations to ice and to land in streptophytes. Comparative genomics revealed that the reductive morphological evolution in the ancestor of Zygnematophyceae was accompanied by reductive genome evolution.
- This first genome-scale data for glacier algae suggests an Ancylonema-specific adaptation to the cryosphere, and sheds light on the genome evolution of land plants and Zygnematophyceae.

Introduction

The evolution of terrestrial flora (Embryophyta) (de Vries & Archibald, 2018) transformed Earth’s continents, atmosphere and climate (Lenton et al., 2016; Bengtson et al., 2017; Leebens-Mack et al., 2019; Delaux & Schornack, 2021; Bowles et al., 2023), and promoted the diversification of multiple lineages spanning the tree of life (Lutzoni et al., 2018; Cortona et al., 2020). The first land plants evolved from a streptophyte algal ancestor (Wickett et al., 2014; de Vries & Archibald, 2018), the last common ancestor of Anydrophyta, with timescale analyses inferring a likely Neoproterozoic divergence between Embryophyta and its closest Zygnematophycean algal relatives, with the youngest estimates falling within the Ediacaran period (Morris et al., 2018), and others dating this divergence as far back as the Mesoproterozoic (Su et al., 2021). To make the fundamental transition from an aquatic to terrestrial environment required adaptations to tolerate extremes in temperature, desiccation and UV radiation (de Vries et al., 2016, 2018a,b, 2020; de Vries & Archibald, 2018; Bowles et al., 2020, 2021). It is likely that ancestral streptophytes possessed a suite of exaptations (preadaptations) available to be co-opted to these ends. Studying closely related extant streptophytes could provide significant insight into the genomic capacity of ancestral lineages and is therefore critical to unravelling the processes involved in land plant terrestrialisation (de Vries & Archibald, 2018; Donoghue et al., 2021).

Streptophyte glacier algae live on the surfaces of contemporary glaciers and ice sheets, and proliferate in widespread algal blooms during summer melt seasons when sunlight and liquid water are available to power photosynthesis (Yallop et al., 2012; Williamson et al., 2018, 2020; Holland et al., 2019; Cook et al., 2020). To inhabit icy environments, glacier algae must balance their requirements for photosynthesis and growth with tolerance of extremes in temperature, desiccation and UV radiation (Williamson et al., 2018, 2020), raising interesting parallels with the adaptation of plants to life on land. Glacier algae also belong to the Zygnematophyceae, the sister lineage to all embryophytes (Williamson et al., 2019). One currently untested hypothesis suggests that plants moved onto land in the aftermath of global glaciations in the Cryogenian, whereby ice provided an intermediate habitat.
between water and land (Williamson et al., 2019; Žárský et al., 2022). Considering the niche of modern-day glacier algae, and their phylogenetic proximity to embaphytes, these species thus represent an important model system to explore adaptations to extreme conditions and potential processes of plant terrestrialisation.

Here, we describe and analyse the metagenome-assembled genome (MAG) of the glacier alga Ancylonema nordenskioldii to investigate the genomic basis of its adaptations to life on ice and the process of plant terrestrialisation more broadly. Specifically, we explore whether modern-day adaptations of glacier algae represent expatations derived from an anyrophyte ancestor, which would support the hypothesis that Cryogenian glaciations were a major driver of land plant evolution, or whether these adaptations emerged more recently within the glacier algal lineage. The MAG also enables comparative analyses of genome evolution in Zygnematophyceae, to test the recent hypothesis (Hess et al., 2022; Bowles et al., 2024) that gene loss has dominated their evolutionary history. As sequence data for glacier algae has historically been limited and with successful culturing protocols only very recently established (Jensen et al., 2023; Remias & Procházková, 2023), our metagenome-assembled genome represents a significant step forward in our genomic understanding of glacier algae and their evolutionary history.

Materials and Methods

Study site and sample collection

Communities of ice-inhabiting glacier algae (Ancylonema nordenskioldii Berggren) were sampled from the surface ice of Morteratsch glacier, Switzerland, in August 2020. Surface ice was collected using a presterilised sampling ice-saw directly into sterile Whirl-Pak bags and maintained frozen at −20°C during transport to the University of Bristol, UK, whereby samples were held at −80°C before genomic DNA extraction as below.

Library construction and sequencing

For short-read sequencing, genomic DNA was extracted using the DNeasy PowerSoil Pro kit (Qiagen, Hilden, Germany) and assessed with Qubit (Thermo Fisher Scientific, Waltham, MA, USA), BioAnalyzer (Agilent, Santa Clara, CA, USA) and Femto Pulse (Agilent). Samples were purified through bead-based clean up and polymerase chain reaction-free libraries were prepared by the Earlham Institute. Short-read libraries were sequenced with Illumina NovaSeq 6000 system in paired-end mode. In total, 369.4 Gb raw data were generated, with 333.7 Gb remaining after filtering by Trimmomatic (Bolger et al., 2014).

For long-read sequencing, samples were ground in liquid nitrogen for 15 min and genomic DNA was extracted following a previously described protocol (Auber & Wisecaver, 2023) and assessed with Qubit (Thermo Fisher Scientific), BioAnalyzer (Agilent) and Femto Pulse (Agilent). The NERC Environmental Omics Facility sequenced long-read libraries with PacBio, producing 22.2 Gb raw data.

Genome assembly

Short- (Illumina, San Diego, California, USA) and long-read (PacBio, Menlo Park, CA, USA) datasets were assembled with hybrid assembler OPERA-MS (Bertrand et al., 2019). A minimum read length of 3000 bp was specified. As the sequenced samples represented communities of organisms, a metagenomic approach was used to computationally extract streptophyte algal data, rigorously minimising the likelihood of contamination. EuKrept (West et al., 2018) was used to classify and filter eukaryote and prokaryote contigs. Within the eukaryotic reads, Kraken2 (Wood et al., 2019) was used to identify and filter 9979 streptophyte contigs, discarding data from other taxonomic groups (e.g. fungi). Multiple methods were then used for genome binning, including Maxbin (Wu et al., 2014), MetaBAT2 (Kang et al., 2019), BinSanity (Graham et al., 2017) and Concoct (Alneberg et al., 2014). These were evaluated with the BUSCO Eukarya dataset (Waterhouse et al., 2018) to identify MetaBAT2 as the best binning tool (Supporting Information Fig. S1), as well as to assess completeness.

RNA sequencing and assembly

RNA for Illumina RNA-Seq was extracted with the RNEasy PowerSoil Total RNA kit (Qiagen) following the manufacturer’s protocol. The quality of RNA was assessed using an Agilent BioAnalyzer (Agilent). The Bristol Genomic Centre sequenced RNA-Seq libraries with Illumina HiSeq 2500 system in paired-end mode. RNA sequencing data were mapped against our streptophyte algal MAG with HISAT2 (Kim et al., 2019), using default parameters. Mapped reads were then assembled with Trinity (Grabherr et al., 2011), with default parameters.

Gene prediction

The MAKER-P (Campbell et al., 2014) pipeline was used for gene annotation in two rounds, incorporating multiple annotation sources. First, assembled RNA reads were mapped to the genome using TopHat2 (Kim et al., 2013). Homology-based gene prediction was completed with the protein-coding genes of closely related species including Mesotanium endlicherianum (Cheng et al., 2019), Penium marinaecium (Jiao et al., 2020) and Spiroloea musciola (Cheng et al., 2019). Gene models obtained from RNA-aided and homology-based pipelines were used to train the de novo prediction pipeline, SNAP (Korf, 2004). A final MAKER-P run, combining all these sources, was used to annotate genes for the Ancylonema nordenskioldii MAG (Campbell et al., 2014). Genome completeness was assessed with BUSCO Eukarya (Waterhouse et al., 2018). Functional annotation of protein-coding genes was completed with InterproScan (Jones et al., 2014).

Distinguishing species of microalgae

Morphological analysis has identified two common species of glacier algae, unicellular Ancylonema alaskanum and filamentous A.
nordenskioldii. Marker based analysis of rbcL and 18S genes demonstrated that these species of Ancylonema are very closely related (Procházková et al., 2021); the only published genetic data for glacier algae to date. BLAST comparison, using default parameters, of previous rbcL and 18S gene sequences from Ancylonema alaskanum and filamentous A. nordenskioldii against the MAG finds best hits against the same contig (Tables S1–S4). The best hits identified our MAG as A. nordenskioldii (Tables S1–S4). Additional analysis was used to assess that the MAG derived from a single species. GC content analysis, using SeqKit (Fig. S2; Shen et al., 2016), and kmer analysis, using Jellyfish (Fig. S3; Marcais & Kingsford, 2011), were conducted. The function of duplicated BUSCO Viridiplantae genes were also assessed, using functional categories in OrthoDB (Table S5; Zdobnov et al., 2021).

Phylogenetic analysis
We utilised the latest transcriptome and genome data, as well as the predicted protein-coding genes from the MAG, to infer the evolutionary history of streptophytes. Specifically, these included Zygnematophyceae transcriptomes from the one thousand plant transcriptomes project (1KP) and genomes across the plant tree of life (Figs S4–S6). A benchmark of 30% BUSCO genes missing was used to filter high quality data (Waterhouse et al., 2018). Two datasets were produced, the first constructed based on genome only data from across the green plants and the second constructed from Zygnematophyceae transcriptomes from the 1KP project and complementary Ancynophyta genome data.

For both datasets, Orthofinder (v.2.3.7) was used to cluster protein-coding genes into orthogroups (Emms & Kelly, 2019), based on sequence divergence, using default settings (orthofinder -f data_folder). Single copy orthologs were identified using a previously described python script (Harris et al., 2020), which removes paralogous genes from orthogroups. The script enables the user to specify a minimum taxonomic occupancy of each orthogroup, set at 70%.

Single copy orthologs were aligned using Mafft (Katoh et al., 2002) using –auto parameter and trimmed with Trimal (Capella-Gutiérrez et al., 2009) using the –automated1 parameter. Two complementary approaches were used to reconstruct phylogenies. In a first concatenation approach, multiple sequence alignments were concatenated using Phyutility to create a supermatrix (Smith & Dunn, 2008). A bootstrapped maximum likelihood phylogeny was inferred using IQ-TREE (Nguyen et al., 2015) using the Bayesian Information Criterion (BIC) to select best fitting substitution model and empirical profile mixture models (C10–C60). 1000 ultrafast bootstrap replicates were used.

For the analysis of ice-binding proteins, reciprocal BLASTs were conducted between the protein-coding genes of the A. nordenskioldii MAG against Uniprot ice-binding proteins, using default parameters and specifying a single best hit. The reciprocal BLAST analyses were compared and matching hits were recorded (Table S20).

Divergence time estimation
For molecular clock analysis, we utilised the aligned and trimmed 472 gene dataset from the green plant genome phylogeny as input into MCMCtree (Yang, 2007). Node distributions using minimum and maximum constraints were specified, with full phylogenetic and age justifications listed in Notes S1. These calibrations derive from previous critical reviews of the fossil record (Harris et al., 2022; Bowles et al., 2024). To specify the prior distributions on node ages, all calibrated nodes were given a hard minimum age and a soft maximum age.

Initially, molecular clock analyses were run without sequence data to obtain effective time priors, to ensure that the calibration densities and time priors were appropriate. The single copy...
orthogroups were divided into 4 partitions according to their evolutionary rate, based on total tree length in IQ-Tree (Nguyen et al., 2015) and grouped using k-means clustering in R (R Core Team, 2014). A relaxed clock model was used (Uncorrelated; Independent Gamma Rates). Given the protein dataset, branch lengths were first estimated using codeml (Yang, 2007). The tree topology was fixed based on the focal maximum likelihood analysis above and was analysed using the normal approximation method in MCMCtree (Yang, 2007). After a burn-in of 10 000 generations, parameter values were saved every 20th generation until 20 000 cycles were saved (400 000 generations total). Trees were plotted using MCMCtreeR (Puttick, 2019).

Results

The metagenome of the glacier alga Ancylonema nordenskiöldii

A single glacier algal MAG was assembled using a combination of PacBio High-Fidelity (HiFi) long reads and Illumina short reads; in total, these were assembled into 9979 contigs (25 KB contig N50) generating an estimated genome size of 170 megabases (Mb). While the contig number is high compared to streptophyte algae sequenced from cultures (Hori et al., 2014; Cheng et al., 2019), this is the first genome-scale data for glacier algae and therefore an important dataset to understand their adaptations and evolution. We predicted 19 593 protein-coding genes in the glacier algal MAG while BUSCO (Waterhouse et al., 2018) analysis (Eukaryota_odb10) suggested a genome completeness of 86.7% (Fig. S1). Analysis revealed that 46% of present BUSCO genes were duplicated, which is comparable to other Zygnematophyceae genomes sequenced from cultured samples including Spirolooa musicola (75%) and Penium margaritaceum (22%). This set of 19 593 protein-coding genes was used for downstream analyses.

Current knowledge based on morphology and limited amplion sequencing data (18S and rbcL) recognises two species of glacier algae, the filamentous Ancylonema nordenskiöldii and the unicellular Ancylonema alaskanum (Williamson et al., 2019). Samples for the present study were collected in 2020 from Morteratsch glacier, Switzerland; a site known for the dominance (70% relative abundance) of A. nordenskiöldii (Mauro et al., 2020). Reciprocal BLAST analysis of glacier algal 18S and rbcL marker genes against our contigs identified our MAG as A. nordenskiöldii (Table S1–S4). Additional analysis of GC content and kmers did not identify a bimodal distribution of contigs (Figs S2, S3). Duplicated BUSCO genes were mostly involved in enzymatic processes as opposed to housekeeping genes, suggesting proliferation of metabolism in a single organism, rather than multiple housekeeping genes, which are less prone to duplication, deriving from multiple related organisms (Table S5). Therefore, we concluded that the MAG contained material from a single species, A. nordenskiöldii, and proceeded conservatively with downstream analyses on that basis. Phylogenetic analysis placed our A. nordenskiöldii MAG within the Zygnematophyceae, the sister group to land plants, with which they comprise the group Anydrophyta (Fig. 1). Within the Zygnematophyceae, A. nordenskiöldii is placed within the Zygnematales (Fig. 1). Further analysis placed Anydrophyta within the Streptophyta, with Chlorophyta sister to the latter group (Fig. S4).

Glacier algal adaptations to life in ice

Comparative genomic analysis combining the A. nordenskiöldii MAG with previously published data indicated that glacier algae adapted to the cryosphere through lineage-specific diversification of existing genetic pathways (Fig. 2a–c; Tables S6–S15). The 19 593 protein-coding genes from our A. nordenskiöldii MAG were clustered into 6242 orthogroups (OGs) using OrthoFinder (Emms & Kelly, 2019). Gene ontology (GO) analysis of 2195 expanded gene families, identified with CAFE (Bie et al., 2006), revealed functions associated with water transport (e.g. lipid localisation), protein repair (e.g. autophagy, PSII associated light-harvesting complex), response to abiotic stimulus (e.g. response to high light intensity, UV, radiation), wax metabolic process (e.g. lipid biosynthesis and modification) and plant-type cell wall modification (e.g. chloroplast organisation; Figs 2b, S7; Tables S6, S8, S12). Due to the identification of the wax metabolic process term, the evolution of the cuticle biosynthetic machinery (Kong et al., 2020) was investigated, which identified the DGAT family in streptophytes, including A. nordenskiöldii (Table S24).

Our analysis also identified extensive gene loss and gene family contraction during the evolution of glacier algae (Fig. 2). Gene losses and gene family contractions outweighed gains and expansions (Fig. 2), suggesting extensive gene turnover with the evolution of glacier algae. Gene ontology analysis of 1583 contracted gene families indicated a reduction in functions associated with phenylpropanoid metabolism, clathrin coat assembly and cellular catabolic process (Figs 2c, S9; Tables S10, S14). GO terms associated with the 2184 lost gene families, identified with Count (Csíkos, 2010), mirrored the contracted gene families of A. nordenskiöldii. These lost genes were associated with phenol-containing compound metabolism, clathrin coat assembly and cellular catabolic assembly (e.g. cell wall organisation) as well as regulation of development (e.g. cell–cell signalling; Figs 2a, S10; Tables S11, S15). Together, these suggest a loss of intracellular trafficking, via clathrin coat assembly, and a reduction in chemical degradation pathways.

Despite the conspicuous phenolic pigmentation of A. nordenskiöldii, a known physiological adaptation to life in surface ice (Remias et al., 2012a,b; Williamson et al., 2020), genes involved in the biosynthesis of many phenylpropanoids and phenolic compounds were lost or became reduced in copy number in this MAG (Figs S9, S10), although several were retained (de Vries et al., 2021; Table S25). Comparative genomic and phylogenetic analysis instead suggested that lineage-specific gene diversification of a particular pathway underpinned the novel screening pigmentation of A. nordenskiöldii (Fig. 3). In glacier algae, purpurogallin pigments absorb ultraviolet and visible light, providing photo-protection against high levels of UV irradiance associated with life in supraglacial surface ice (Remias et al., 2012a,b; Williamson...
Purpurogallin accumulation has also been provisionally associated with protection against low temperatures, protection of the photosynthetic machinery and tolerance to desiccation (Williamson et al., 2019). While genes involved in the purpurogallin biosynthetic pathway were found across the green plants, one component, dehydroquinate dehydratase/shikimate dehydrogenase (DHQD/SD), was present at a high copy number (9) in *Ancylonema nordenskioldii*; greater than in any other green plant species (Fig. 3; Table S16). DHQD/SD catalyses the dehydration of dehydroquinate (DHQ) to dehydroshikimate (DHS) and the reduction of dehydroshikimate (DHS) to shikimate (Bontpart et al., 2016; Lynch, 2022), leading to the spontaneous synthesis of gallic acid. Phylogenetic analysis demonstrated that the expansion of this gene family was specific to *Ancylonema nordenskioldii* (Fig. 3c).

Aside from purpurogallin, comparative genomic analysis did not find any *Ancylonema nordenskioldii* specific gene radiations in relation to well-characterised light screening pathways, including UV-A or UV-B stress (Table S17). This is further supported by genetic analysis of photosystem I and II here, showing no significant patterns of expansion or contraction in known genes in the photosynthetic machinery (Table S18).

Analysis of ice-binding proteins suggested that cold adaptations of *Ancylonema nordenskioldii* represent a unique adaptation of...
glacier algae, rather than an ancient exaptation derived from Anydrophyta (Tables S19, S20). Reciprocal BLAST of Uniprot ice-binding proteins from across the tree of life (e.g. fungi, bacteria) revealed 1194 OGs with hits against *A. nordenskiöldii* (Table S19). Based on the taxonomic occupancy of OGs, many were distributed across the green plant phylogeny (Table S20). These included a protein kinase superfamily, ATP-binding cassette protein family and heat shock protein
family. Only 35 of these 1194 OGs were novel to Anhydrophyta, indicating a relatively small evolutionary response to cold stress. Additionally, *A. nordenskiöldii* had the highest copy number amongst all green plants for some gene families, including a lipase, ATP-dependent Clp protease and Cytochrome C Oxidase.

Further analysis of land plant cold tolerance pathways did not find any gene family expansion in *A. nordenskiöldii* (Fig. S17; Table S21), instead indicating that a core element of land plant cold tolerance emerged after the split with Zygnematophyceae.

Cold stress signalling in land plants is coordinated by the CBF (C-REPEAT BINDING FACTOR)-COR (COLD REGULATED) signalling pathway (Ding *et al*., 2019). Comparative genomics revealed that the majority of this pathway predated the origin of Anhydrophyta (Table S21). However, ICE (INDUCER OF CBF EXPRESSION), a key regulator of CBFS, emerged in the ancestor of land plants. ICE, also known as SCRM (SCREAM), is involved in stomatal development (Chater *et al*., 2017), and may have evolved in land plants as an adaptation to both cold and drought stresses.

Fig. 3  Comparative genomics of purpurogallin biosynthesis. (a) Current understanding of the purpurogallin biosynthetic pathway. Orange boxes indicate secondary metabolites, with arrows indicate enzymatic reaction. Genes in pathway have been abbreviated as follows: ADT, arogenate dehydratase; CM, chorismate mutase; CS, chorismate synthase; DHQD/SD, dehydratase/shikimate dehydrogenase; DHQS, 3-dehydroquinic acid synthase; DHS1, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase; ESPS, 5-enolpyruvylshikimate 3-phosphate synthase; PAT, prephenate aminotransferase; SK1, Shikimate kinase 1. Enzymes in the pathway have abbreviated as follows: DAHP, D-arabino-petulosonic acid 7-phosphate synthase; DHQ, 3-dehydroquinic acid; DHS, 3-dehydroshikimic acid; EPSP, 5-enolpyruvylshikimate 3-phosphate; PEP, phosphoenolpyruvate; S3P, shikimate 3-phosphate; UGGT, U-glutamyl transpeptidase. The yellow box highlights DHQD/SD. (b) Count of purpurogallin biosynthesis genes, with the size of circle corresponding to gene copy number. The tree is based on phylogenetic analysis in Supporting Information Fig. S4. Orange box highlights gene count for Ancylonema nordenskiöldii. Yellow circle highlights gene expansion of dehydroquinate dehydratase/shikimate dehydrogenase (DHQD/SD) in *A. nordenskiöldii*. (c) Phylogenetic analysis of dehydroquinic acid synthase/shikimate dehydrogenase gene family. Numbers on branches highlight bootstrap support values. The green background denotes land plants while blue background denotes Zygnematophyceae. (d) Illustration of pigmentation in glacier algae. The lower panel image shows heavily pigmented glacier algae assemblages. The upper panel image demonstrates surface ice environments of Morteratsch glacier with dark zone dominated by glacier algal blooms.
Due to their importance in abiotic stress responses, the evolution of the biosynthesis and signalling of phytohormones was also investigated at a broad phylogenetic level (e.g. presence in Anydrophyta and land plants). Similar to previous analyses (Bowman et al., 2017; Nishiyama et al., 2018; Bowles et al., 2020; Jiao et al., 2020), we demonstrated these phytohormone genes are mostly estimated to originate before the transition of plants onto land (Table S22; Fig. S18). CBF/DREBs (DEHYDRATION RESPONSIVE ELEMENT BINDING) are transcription factors that bind to DRE/CRT cis-acting elements in response to abiotic stress (e.g. drought, low temperatures). As such, the evolution of DREBs and all other transcription factors was investigated (Table S23; Fig. S19), with all major families being found in the A. nordenskiöldii MAG, consistent with previous studies (Catarino et al., 2016; Lai et al., 2020; Bowles et al., 2022).

Timescale of glacier algal evolution

Timescale analysis based on 472 genes from the genomes of 24 green plant species suggested that Ancylonema nordenskiöldii split from its closest living algal relative 520–455 million years ago (Ma) during the Cambrian-Ordovician (early Phanerozoic). Consistent with previous estimates (Morris et al., 2018; Nie et al., 2020; Harris et al., 2022), our analyses also suggested that Anydrophyta emerged 703–623 Ma likely during the Cryogenian-earliest Ediacaran, a period of dynamic environmental change characterised by two major global glaciations (Fig. 4) (Stern et al., 2006; Hoffman et al., 2017). The above comparative genomic and phylogenetic analyses have suggested that adaptations to life in ice are lineage-specific to Ancylonema. As such, further comparative genomics (Tables S6, S7, S26–S49) was used to investigate patterns and functions of genome evolution in Anydrophyta, land plants and Zygnematales, and potential exaptive Cryogenian evolution.

Genome and gene family evolution

Analysis of gene family evolution highlighted increasing genome complexity of Anydrophyta and land plants, contrasting with large-scale genome reduction in the Zygnematales (Fig. 2; Table S23).
The origin of Anydrophyta was accompanied by a peak of gene novelty (1113 gained & 1848 expanded OGS) and low gene loss (210 lost & 132 contracted OGS). Within the first-splitting lineages of Anydrophyta, the land plant ancestor exhibited another peak of gene novelty (954 gained & 898 expanded OGS). By contrast, within the ancestor of Zygnumatophyceae gene loss and contraction were both high (927 lost & 482 contracted OGS; Tables S36, S37). Further reductive evolution was also prevalent within the Zygnumatophyceae, in the ancestor of Zygnematales (500 lost & 683 contracted OGS) and the common ancestor of Ancylonema nordenskiöldii and Zygnum species (419 lost & 274 contracted OGS; Fig. 2a; Tables S6, S7).

Analysis of gene ontology (GO) terms and Pfam domains documented how ancestral anydrophytes and land plants evolved increasingly complex anatomies to establish in terrestrial environments (Fig. 2d,e; Tables S38, S42). Gained OGS in Anydrophyta were associated with responses to gravity, polysaccharide metabolism (e.g. xyloglucan metabolism) and anatomical structure development (e.g. cell differentiation; Figs 2d, S12). In land plants, gained OGS were associated with hydrotopism, cutin biosynthesis (e.g. meristem development), plant-type secondary cell wall biogenesis and system process (e.g. vascular and phloem transport; Figs 2e, S14).

Gene family expansions resulted in an increased number of stress response genes in the first Anydrophyta and in land plants (Fig. 2a; Table S39, S43). Expanded OGS in Anydrophyta were linked to water transport, response to abiotic stimulus (e.g. responses to salt, osmotic stress, heat, cold) and plastid organisation (Fig. S11). Land plant expanded OGS were associated with regulation of development, intercellular transport and response to abiotic stimulus (responses to salt, temperature, osmotic stress; Fig. S13).

Our GO analysis suggested that extensive loss and contraction of gene families in the first Zygnumatophyceae may have contributed to the reductive evolution of their simple morphology from a more complex ancestor (Fig. 2; Tables S48, S49). The OGS lost from Zygnumatophyceae were associated with regulation of molecular function (e.g. regulation of cell communication), lignan biosynthesis and plasma membrane fusion (e.g. extracellular matrix and structure organisation; Figs 2f, S15). Contracted OGS in Zygnumatophyceae were linked with regulation of molecular function (e.g. auxin polar transport), responses to endogenous stimuli (e.g. salt stress, hypoxia, biotic stimulus), regeneration (e.g. multicellular organism development) and extracellular matrix organisation (Figs 2g, S16).

### Discussion

**Genomic survey of the glacier alga Ancylonema nordenskiöldii**

The first genome-scale data produced here for streptophyte glacier algae allowed us to test hypotheses on the genomic basis of glacier algal adaptations to life within ice, to examine their phylogenetic and evolutionary history, and to explore processes of plant terrestrialisation over geological timescales. Our metagenome approach, utilised to circumvent the longstanding absence of viable glacier algal cultures (e.g. Jensen et al., 2023; Remias & Procházková, 2023), produced a metagenome-assembled genome (MAG) consistent with previous Zygnumatophycean genomes in terms of both genome size and the number of protein-coding genes predicted. For example, our Ancylonema nordenskiöldii MAG was estimated at 170 megabases (Mb), within the range of previous Zygnumatophycean genomes including Spioroega muscula (174 Mb, Mesotaenium endlicherianum: 163 Mb (Cheng et al., 2019), Zygnum circumcarinatum SAG 698-1b: 71 Mb (Feng et al., 2024), Zygnum circumcarinatum UTEX1559: 71 Mb (Feng et al., 2024), Zygnum circumcarinatum UTEX1560: 67 Mb (Feng et al., 2024), Zygnum cf. ciliatrum: 360 Mb (Feng et al., 2024), and Penium margaritaceum: 4700 Mb (Jiao et al., 2020). This suggests that although the large size of the Penium margaritaceum genome is an outlier amongst the Zygnumatophyceae, likely deriving from an expansion of transposons (Jiao et al., 2020). The number of protein-coding genes (19 593) of our A. nordenskiöldii MAG were highly consistent with its closest sequenced relative, Mesotaenium endlicherianum, with 19 233 protein-coding genes (Dadras et al., 2023). Results from our phylogenetic analysis were also consistent with previous work on Zygnumatophyceae (Hess et al., 2022), Streptophyta (Wickett et al., 2014) and green plant evolution (Leebens-Mack et al., 2019). Morphological analysis has classified glacier algae as the filamentous A. nordenskiöldii and the unicellular A. alaskanum (previously Mesotaenium berggreni) (Williamson et al., 2019; Procházková et al., 2021). Our analysis of genome-scale data indeed placed glacier algae separate from Mesotaenium endlicherianum (Fig. 1), highlighting the need for re-classification of Mesotaenium species as well as species in the genus Cylindrocystis.

**Glacier algal adaptation to ice**

Several lines of evidence suggest here that glacier algal adaptation to life in ice are lineage-specific and do not derive from ancient exaptations of an ancestral Cryogenian anydrophyte. These include the lack of gained or expanded OGS in Anydrophyta linked to cold tolerance, the emergence of a core element of land plant cold tolerance after the split with Zygnumatophyceae, and the absence of gene family expansion of land plant cold tolerance genes in A. nordenskiöldii (Fig. S17; Table S21). Our work thus argues against the previous hypothesis that ice may have provided an intermediate habitat between water and land during processes of plant terrestrialisation (Williamson et al., 2019; Žárauský et al., 2022).

By contrast, analysis of our A. nordenskiöldii MAG highlighted low gene gain (276 genes gained), suggesting expactive evolution of a zygnumatophycean alga to glacial environments (Figs 2a, S8; Tables S7, S9, S13). Expanded gene families suggested that glacier algae gained adaptations to key abiotic stressors, principally high-light and UV stress, consistent with their high-light surface ice environment (Williamson et al., 2020), potentially achieved
through duplication of genes involved in light sensing and photodamage repair. These expansions were not seen in known light screening pathways (e.g. UV-A, UV-B, PSI, PSII) suggesting genes in *A. nordenskiöldii* for high light stress derive from outside these well-characterised pathways. Importantly, gene duplication followed by neo-functionalisation may have enabled glacier algae to synthesise their novel purpuror gallin pigment that underpins their dominance of surface ice environments (Williamson et al., 2018, 2019, 2020). The high production of purpuror gallin within glacier algal cells, that is to 11-times the cellular content of Chl *a* (Williamson et al., 2020), may have made other phenolic and phenylpropanoid-based compounds functionally redundant, explaining the loss and contraction of genes important for secondary metabolite biosynthesis highlighted here (Fig. 2c). The lack of evolutionary signal in relation to alternate light screening mechanisms supports previous assertions that purpuror gallin provides the bulk of photoprotection within glacier algal cells, while chloroplasts remain typically light-adapted for green algae (Williamson et al., 2020). Indeed, *Ancylonema nordenskiöldii* has been shown to be very well protected against high light, with oxygen production apparently unaffected up to 2000 μmol photons m⁻² s⁻¹ (Remias et al., 2012a,b). While purpuror gallin is produced in high quantities in glacier algae, other members of the Zygnematophyceae synthesise sunscreens with similar chemical composition to gallic acid (Busch & Hess, 2022). For example, *Zygogonium erictorum* contains an unusual phenolic compound, a glycosylated ferric Fe₃ + (gallate)₂ complex, giving this alga a distinct purple colour (Aigner et al., 2013). Thus, more work is required to understand the evolution and genetics of phenolic biosynthesis in the Zygnematophyceae.

### Timescale of glacier algal evolution

Our divergence time estimates of 520–455 Ma for the split of *Ancylonema nordenskiöldii* from its closest relatives emphasises the sparseness of genome-scale sampling in this important region of the streptophyte tree (Fig. 4). This estimate suggests either that no other close relatives have survived to the present day, or that close relatives have not yet been identified. While no long-term glaciations occurred during this period, this estimate is pre-dated and postdated by a number of glacial episodes (e.g. the mid Ediacaran Gaskiers, Late Ordovician and the late Carboniferous glaciations) (Pohl et al., 2016), potential drivers of glacier algal evolution. Additionally, regardless of their exact origination point, this represents a long evolutionary time in which to assemble glacial adaptations. While several studies produce an older estimate for crown Anhydrophyta (Su et al., 2021; Yang et al., 2023), they require that key fossils (e.g. *Proterocladus antiquus*, *Bangiomorpha pubescens*) are assigned to derived phylogenetic positions that, in our view, are not justifiable based on the phenotypic evidence preserved (Notes S1).

### Genome and gene family evolution

The divergent genomic trajectory of land plants and Zygnematophyceae identified here, mirror their morphological trajectories (Hess et al., 2022; Bowles et al., 2024) and suggest that genome streamlining underpins the reductive morphological evolution of Zygnematophyceae. Contrasting patterns of genome evolution between closely related groups have been seen elsewhere in the tree of life (Paps & Holland, 2018; Guijarro-Clarke et al., 2020; Harris et al., 2022; Ocaña-Pallarés et al., 2022), highlighting the role of gene gains as well as losses in driving phenotypic evolution (Clark, 2023). Our gene ontology analysis indicated that gene gains in Anhydrophyta and land plants were associated with anatomical structural development. A greater repertoire of drought stress response pathways emerged in Anhydrophyta and land plants, which potentially enabled them to tackle temperature and osmotic stress of the Cryogenian through to the Cambrian. These gained genes could derive from *de novo* formation from noncoding sequences, shuffling from a combination of existing domains or through horizontal gene transfer as recently identified in land plants (Bowles et al., 2020; Ma et al., 2022; Xue et al., 2023). In Zygnematophyceae, gene losses were linked to multicellular development. As extant Zygnematophyceae are unicellular or filamentous (Hess et al., 2022), the loss and contraction of gene families involved in extracellular matrix organisation suggest a clear molecular signature of the evolution of their simple morphology.

### Conclusions

Production of the first genome-scale dataset for ice-inhabiting glacier algae demonstrated their unique adaptation to life in ice and served to reject recent hypotheses that processes of terrestrialisation dating to the Cryogenian involved the adaptation of ancestral streptophytes to ice as an intermediate habitat between water and land. Instead we highlight the more recent (Orдович – Cambrian) divergence of glacier algae from their closest sequenced relatives and their exaptive evolution to life in ice surrounding novelties of high-light tolerance and specialised pigment production. Corresponding to their divergent morphology, we identify the divergent genomic trajectories of Zygnematophyceae and land plants, which functionally link to the loss of multicellularity and gain of abiotic stress tolerance, respectively. Our analysis, therefore, adds to the growing body of work demonstrating the stress tolerance capabilities of the common ancestor of land plants and Zygnematophyceae (de Vries et al., 2016, 2018a, 2020; Zhao et al., 2019; Becker et al., 2020), eventually leading to the establishment of plants on land. Indeed, this bears out the expectation that the freshwater algal relatives of embryophytes were confronted by much the same challenges as their land plant relatives, and the adaptive solutions that they established primed stem-embryophytes for life on land (Donoghue & Paps, 2020).

### Acknowledgements

We wish to acknowledge funding from the Leverhulme Trust (RPG-2020-199 ‘iDAPT’ project to CW, DC, PD, TW; RF-2022-167 to PD); the Natural Environment Research Council (NE/P013678/1 to PD); part of the Biosphere Evolution,
Transitions and Resilience (BETR) programme cofunded by the Natural Science Foundation of China (NSFC); the John Templeton Foundation (62220 to PD, TW); the Gordon and Betty Moore Foundation (GBMF9741 to PD, TW) and a University Research Fellowship to TW (URF 1R210124).

Competing interests
None declared.

Author contributions
The work was devised by AMCB, TAW, PCJD, DAC and CJW. AMCB conducted the analyses and drafted the manuscript. All authors contributed to writing, reviewing and editing the finished manuscript.

ORCID
Alexander M. C. Bowles https://orcid.org/0000-0002-7487-3811
Douglas A. Campbell https://orcid.org/0000-0001-8996-5463
Philip C. J. Donoghue https://orcid.org/0000-0003-3116-7463
Tom A. Williams https://orcid.org/0000-0003-1072-0223
Christopher J. Williamson https://orcid.org/0000-0002-8239-7368

Data availability
The Ancylonema nordenskiöldii MAG, associated proteins, and genes used to infer species phylogenies were deposited (FigsHare: doi: 10.6084/m9.figshare.25305640), whilst raw Illumina, PacBio and RNA sequencing data were deposited to NCBI (PRJNA1103419). Other data supporting the findings of the paper are available in the Supporting Information.

References


**Supporting Information**

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

**Fig. S1** Benchmarking Single Copy Orthologs analysis of *Ancylonema nordenskiöldii* metagenome-assembled genome, as well as previously published Zygnematophyceae genomes.

**Fig. S2** Analysis of GC content for the contigs of *Ancylonema nordenskiöldii* metagenome-assembled genome.

**Fig. S3** Analysis of kmers for the contigs of *Ancylonema nordenskiöldii* metagenome-assembled genome.

**Fig. S4** Phylogenetic analysis of green plant genomes based on 472 genes using a concatenation based approach.

**Fig. S5** Phylogenetic analysis of green plant genomes based on 472 genes using a coalescence based approach (Astral).

**Fig. S6** Phylogenetic analysis of green plants incorporating data from the 1 KP project as well as genome data.

**Fig. S7** Reduce and Visualise Gene Ontology plot for expanded gene families in *Ancylonema nordenskiöldii*.

**Fig. S8** Reduce and Visualise Gene Ontology plot for gained gene families in *Ancylonema nordenskiöldii*. 

---

© 2024 The Authors

*New Phytologist* © 2024 *New Phytologist* Foundation

[www.newphytologist.com](http://www.newphytologist.com)
Fig. S9 Reduce and Visualise Gene Ontology plot for contracted gene families in Ancylonema nordenskiöldii.

Fig. S10 Reduce and Visualise Gene Ontology plot for lost gene families in Ancylonema nordenskiöldii.

Fig. S11 Reduce and Visualise Gene Ontology plot for expanded gene families in Anydrophyta.

Fig. S12 Reduce and Visualise Gene Ontology plot for gained gene families in Anydrophyta.

Fig. S13 Reduce and Visualise Gene Ontology plot for expanded gene families in land plants.

Fig. S14 Reduce and Visualise Gene Ontology plot for gained gene families in land plants.

Fig. S15 Reduce and Visualise Gene Ontology plot for contracted gene families in Zygnematophyceae.

Fig. S16 Reduce and Visualise Gene Ontology plot for lost gene families in Zygnematophyceae.

Fig. S17 The evolution of land plant cold stress signalling pathways.

Fig. S18 The evolution of phytohormone signalling.

Fig. S19 The evolution of transcription factors.

Notes S1 Fossil calibrations for Molecular Clock Analysis.

Table S1 BLAST hits against Ancylonema nordenskiöldii rbcL.

Table S2 BLAST hits against Ancylonema alaskanum rbcL.

Table S3 BLAST hits against Ancylonema nordenskiöldii 18S.

Table S4 BLAST hits against Ancylonema alaskanum 18S.

Table S5 Duplicated Benchmarking Single Copy Orthologs genes and their functions.

Table S6 Gene family expansion and contraction analysis with CAFÉ.

Table S7 Gene gain and loss analysis with COUNT.

Table S8 Expanded genes in Ancylonema nordenskiöldii.

Table S9 Gained genes in Ancylonema nordenskiöldii.

Table S10 Contracted genes in Ancylonema nordenskiöldii.

Table S11 Lost genes in Ancylonema nordenskiöldii.

Table S12 Gene Ontology Analysis for expanded genes in Ancylonema nordenskiöldii.

Table S13 Gene Ontology Analysis for gained genes in Ancylonema nordenskiöldii.

Table S14 Gene Ontology Analysis for contracted genes in Ancylonema nordenskiöldii.

Table S15 Gene Ontology Analysis for lost genes in Ancylonema nordenskiöldii.

Table S16 Occupancy of purpurogallin biosynthesis genes.

Table S17 Occupancy of light adaptation genes.

Table S18 Occupancy of Photosystem I & II genes.

Table S19 Taxonomic occupancy of potential ice-binding proteins.

Table S20 BLAST hits against ice-binding proteins.

Table S21 Occupancy of cold stress genes.

Table S22 Occupancy of phytohormone biosynthesis and signalling genes.

Table S23 Occupancy of transcription factors.

Table S24 Occupancy of wax metabolic genes.

Table S25 Occupancy of phenylpropanoid biosynthesis genes.

Table S26 Expanded genes in Anydrophyta.

Table S27 Gained genes in Anydrophyta.

Table S28 Contracted genes in Anydrophyta.

Table S29 Lost genes in Anydrophyta.

Table S30 Expanded genes in land plants.

Table S31 Gained genes in land plants.

Table S32 Contracted genes in land plants.

Table S33 Lost genes in land plants.

Table S34 Expanded genes in Zygnematophyceae.

Table S35 Gained genes in Zygnematophyceae.

Table S36 Contracted genes in Zygnematophyceae.

Table S37 Lost genes in Zygnematophyceae.
Table S38 Gene Ontology Analysis for expanded genes in Anydrophyta.

Table S39 Gene Ontology Analysis for gained genes in Anydrophyta.

Table S40 Gene Ontology Analysis for contracted genes in Anydrophyta.

Table S41 Gene Ontology Analysis for lost genes in Anydrophyta.

Table S42 Gene Ontology Analysis for expanded genes in land plants.

Table S43 Gene Ontology Analysis for gained genes in land plants.

Table S44 Gene Ontology Analysis for contracted genes in land plants.

Table S45 Gene Ontology Analysis for lost genes in land plants.

Table S46 Gene Ontology Analysis for expanded genes in Zygnematophyceae.

Table S47 Gene Ontology Analysis for gained genes in Zygnematophyceae.

Table S48 Gene Ontology Analysis for contracted genes in Zygnematophyceae.

Table S49 Gene Ontology Analysis for lost genes in Zygnematophyceae.

Table S50 Phylopic links.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.