ISME

www.nature.com/ismej

ARTICLE OPEN

Evolution of snow algae, from cosmopolitans to endemics, revealed by DNA analysis of ancient ice

Takahiro Segawa[1✉], Takahiro Yonezawa[2✉], Ryo Matsuzaki[3,4], Hiroshi Mori[5], Ayumi Akiyoshi[6], Francisco Navarro[7], Koji Fujita[8], Vladimir B. Aizen[9], Zhongqin Li[10], Shuhei Mano[11] and Nozomu Takeuchi[12]

Recent studies of microbial biogeography have revealed the global distribution of cosmopolitans and dispersal of regional endemics, but little is known about how these processes are affected by microbial evolution. Here, we compared DNA sequences from snow/glacier algae found in an 8000-year-old ice from a glacier in central Asia with those from modern snow samples collected at 34 snow samples from globally distributed sites at the poles and mid-latitudes, to determine the evolutionary relationship between cosmopolitan and endemic phylotypes of snow algae. We further applied a coalescent theory–based demographic model to the DNA sequences. We found that the genus *Raphidonema* (Trebouxiophyceae) was distributed over both poles and mid-latitude regions and was detected in different ice core layers, corresponding to distinct time periods. Our results indicate that the modern cosmopolitan phylotypes belonging to *Raphidonema* were persistently present long before the last glacial period. Furthermore, endemic phylotypes originated from ancestral cosmopolitan phylotypes, suggesting that modern regional diversity of snow algae in the cryosphere is a product of microevolution. These findings suggest that the cosmopolitans dispersed across the world and then derived new localized endemics, which thus improves our understanding of microbial community formation by microevolution in natural environments.

The ISME Journal (2023) 17:491–501; https://doi.org/10.1038/s41396-023-01359-3

INTRODUCTION

Global microbial colonization, which gives rise to cosmopolitan species, has been a topic of great interest since 1934, when Baas Becking hypothesized that "Everything is everywhere, but the environment selects" [1]. However, few details are known about the spatial genetic structures within microbial cosmopolitan species as well as the processes by which they have formed. Advances in genome sequencing technology have revealed that most of the observed spatial distribution of microorganisms has been endemic, being found only in localized regions, based on DNA sequencing at single-nucleotide (nt) resolution [2–5]. Moreover, a recent study of photosynthetic algae that grow on the snow surface in polar regions demonstrated that these snow algae consist of limited numbers of cosmopolitan phylotypes and very diverse endemic phylotypes [6]. Current global diversification and regional endemism of microorganisms have been formed by genetic isolation and adaptation [7–10]. Therefore, to understand how microbes migrate globally and adapt locally, it is necessary to study the evolutionary history of cosmopolitan and endemic phylotypes.

Our knowledge on the time scale of development of microbial population structures, and how they have been preserved is very limited. Because ice-core samples from glaciers provide a time series of past genome information [11], analysis of the DNA information in ice core samples offers new insights into the microevolutionary process that shapes the global-scale biogeographic structure. In addition, because the snow microbiota is genetically tractable, and the polar and mid-latitude/high-mountain cryosphere regions are physically isolated, this microbiota is most suitable for understanding the microevolution of communities driven by migration-isolation and subsequent local adaptation [12–15]. Therefore, microorganisms in snow and ice environments are the most appropriate organisms for elucidating the history of biogeographical distribution and microevolution. Although microbial analyses of ice cores have been conducted using conserved gene regions such as 16S or 18S rRNA genes [16–22], no analysis has focused on the microevolution between cosmopolitans and endemic distribution.

To overcome this shortcoming, we here analyze ancient DNA information derived from an ice-core that dates continuously back to 8000 years before the present, and analyze a comprehensive global modern dataset of snow algae collected from both poles and mid-latitudes (Tables S1–S2). The results clarify the time scale on which cosmopolitan species were established and maintained.

[1]Center for Life Science Research, University of Yamanashi, Yamanashi, Japan. [2]Department of Animal Science, Faculty of Agriculture, Tokyo University of Agriculture, Kanagawa, Japan. [3]Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan. [4]Biodiversity Division, National Institute for Environmental Studies, Ibaraki, Japan. [5]Advanced Genomics Center, National Institute of Genetics, Shizuoka, Japan. [6]National Institute of Polar Research, Tokyo, Japan. [7]Departamento de Matemática Aplicada a las Tecnologías de la Información y las Comunicaciones, ETSI de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain. [8]Graduate School of Environmental Studies, Nagoya University, Aichi, Japan. [9]Department of Earth and Space Science, University of Idaho, Moscow, Idaho, USA. [10]Laboratory of Cryospheric Sciences, Northwest Institute of Eco-Environment and National Resources/Tianshan Glaciological Station, Chinese Academy of Sciences, Gansu, China. [11]The Institute of Statistical Mathematics, Tokyo, Japan. [12]Department of Earth Sciences, Graduate School of Science, Chiba University, Chiba, Japan. ✉email: tsegawa@yamanashi.ac.jp; ty205685@nodai.ac.jp

Received: 10 June 2022 Revised: 24 December 2022 Accepted: 6 January 2023
Published online: 17 January 2023

MATERIALS AND METHODS

Ice-core samples

The ice core was drilled at the top of the Grigoriev Ice Cap in the Tien Shan Mountains in Kyrgyz Republic in August, 2007 (Fig. 1). The ice cap summit is a flat snowfield at elevation 4563 m a.s.l. [23]. The 86.87 m length ice core, made up of 19 pieces, spanned the distance from the surface to the bedrock of the ice cap. The ice core was transported in a frozen state to the ice-core laboratory at the Research Institute for Humanity and Nature, Japan. The bottom age was dated as 12,500 calendar years before present (1950 AD) based on radiocarbon dating of particulate organic matter [24].

The 19 ice-core samples were melted using a device developed by our group that enabled us to obtain water from only the inner portion of the core; the 6-mm-thick outer cylindrical layer remained intact [25]. Complete separation of the inner and outer portions of the ice core is required to avoid contamination by modern microorganisms that can adhere to a core during drilling and storage. The resultant water samples (10–20 ml) containing visible dust particles were filtered through a sterilized 0.22 µm membrane filter (Nalgene Analytical Filter Unit, Thermo Scientific, USA) to collect microorganisms.

To determine whether any contamination had occurred during handling or due to ice-core cracks, a solution of ~2 ng/µl bacterial plasmid vector (pCR4-TOPO vector, Life Technologies, Carlsbad, CA, USA) was applied to the ice-core surface for 20 h before sampling of the inner portion. The vector contaminant was not amplified from the inner part of ice-core samples, but amplified only for the outer layer of ice-core samples as assessed by 45 cycles of PCR with vector-specific primers (M13-Forward and M13-Reverse primers).

Modern snow samples

Snow samples were collected into sterile 50 ml plastic conical tubes during the melt season from 10 sites on 5 glaciers or snowpacks in No. 31 Glacier in Suntar-Khayata Mountains in Russia; Glacier No. 1 in the Eastern Tien Shan in China; the Grigoriev Ice cap at the Inner Tien Shan in Kyrgyz Republic; Fedchenko Glacier in Central Pamir in Republic of Tajikistan; and Tateyama Mountains in Japan (Table S2). Samples were kept frozen during transport to the National Institute of Polar Research (Tokyo, Japan) and then stored at −80 °C before use. Although phototrophs of glaciers include both snow and glacier algae, which dominate on snow-covered and bare ice surfaces of the glacier, respectively, in this study we simply refer to them as snow algae since most of the samples analyzed were collected in snow-covered surfaces.

DNA extraction

Prior to genome library amplification, all work on ancient DNA in ice cores was conducted in a dedicated clean room for ancient DNA at the National Institute of Polar Research, Japan. Membrane filters containing microbial DNA were transferred into 2 ml Matrix-E tubes (Qbiogene, USA) with 600 µl of extraction buffer as described in Willerslev et al. [16]. Two sequential homogenizations were carried out with a multi-beads shocker (Yasui Kikai, Japan) at 1000 rpm for 30 s to remove microorganisms and then at 2500 rpm for 30 s to disrupt microorganisms. The homogenized samples were added with proteinase K (Roche) and incubated at 55 °C for at least 4 h under agitation. Each solution was adjusted with NaCl to 1.15 M, treated with ½ vol. of chloroform/octanol (24:1), and slowly agitated overnight at room temperature. The mixture was centrifuged at 12,000 × *g* for 2 min, and the aqueous phase was collected and transferred into a separate microtube for incubation at 2 °C for at least 1 h. Further purification steps were the same as described by Orlando et al. [26]. The supernatant was incubated with 100 µl silica beads and 40 ml of binding buffer (Qiagen PB buffer; 25 mM NaCl, 87 mM Na acetate) for 3 h at room temperature. The supernatant was discarded, and the pellet was washed twice with 1 ml of 80% ethanol before eluting the DNA with 60 µl elution buffer (Qiagen, Germany).

The modern snow samples were melted at 4 °C, and 5–10 ml of each sample were centrifuged at 5000 × *g* for 10 min to obtain a pellet. The pellets from five replicate samples collected at each site were pooled and used for DNA extraction. Genomic DNA was extracted from each pellet using a FastDNA spin kit for soil (Qbiogene) with a Multi-beads shocker (Yasui Kikai) at 2500 rpm for 30 s. All DNA extractions were conducted on a class 100 clean bench (Sanyo, Japan), and subsequent procedures were carried out on another class 100 clean bench.

ITS2 region amplicon sequencing

The ITS2 region was amplified by PCR using primers c and b (5′-GCATC GATGAAGAACGCAGC-3′ and 5′-GGGATCCATATGCTTAAGTTCAGCGGGT-3′, respectively; [27]) with Illumina overhang adaptor sequences at their 5′ ends. Each PCR mixture (25 µl) contained 1 × KAPA HiFi HS ReadyMix (Kapa Biosystems, USA), 0.2 µM of each primer, and 2–4 µl template DNA. The PCR cycling conditions were as follows: initial annealing at 95 °C for 3 min, followed by 25 cycles (for modern samples) or 45 cycles (ice-core samples) at 95 °C for 30 s, 50 °C for 30 s, and 72 °C for 60 s, with a final extension at 72 °C for 5 min. The amplicons with Illumina overhang adapter sequences were generated in triplicate, and each amplicon was pooled before index PCR. The PCR products were labeled with two sample-specific indices each containing a sample-unique index [28] and Illumina adapter sequences at their 5′ end (Nextera XT index kit v2, Illumina). The PCR mixture (10 µl) contained 1×KAPA HiFi HS ReadyMix, 2 µl each of forward and reverse primers, and 1 µl of the recovered PCR products. PCR was performed under the following cycling conditions: 95 °C for 3 min, followed by 8 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 60 s, with final extension at 72 °C for 5 min. After agarose gel electrophoresis, PCR products were excised from the gel and purified using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel, Germany). Tagged amplicons were mixed with PhiX control DNA at a ratio of 80:20 and used as a template for MiSeq paired-end sequencing (2 × 300 bp) using Reagent kit v3 (Illumina). To avoid index switching with MiSeq [28], two separate MiSeq runs were carried out for the ice-core and modern samples.

Sequence quality filtering and taxonomic assignments

Adapter sequences in the reads were removed, and quality filtering was conducted using fastp version 0.20 with default parameters [29]. The forward and reverse MiSeq reads were merged for the paired-end libraries of ITS2 using USEARCH (version 10.0.240) [30] with the fastq_mergepairs parameter. Only paired reads carrying the exact index combinations were assigned to each sample's reads and used for subsequent analyses. Reads were discarded if they (i) contained ambiguous nts, and/or (ii) mapped to the PhiX genome sequence as determined by Bowtie 2 (version 2.2.4) with default parameters [31]. The forward and reverse PCR primer sequences were removed with cutadapt (version 2.6) with default parameters [32].

The unique ITS2 sequence clusters, i.e., the non-redundant sequence clusters, were constructed by clustering high-quality reads of all samples using USEARCH (version 10.0.240) with the fastx_uniques parameter. The unique sequences of the ITS2 region were aligned using MAFFT v7.429 [33]. These alignments were carefully inspected by eye, and incomplete sequences and those that were not full length and the singleton clusters were discarded.

Taxonomic assignments of unique ITS2 sequences were conducted by a BLASTn search [34] with a top-hit *E* value of $<1e^{-8}$, identity of >90%, and alignment length of >200 bp against (i) the UNITE fungal ITS2 sequence database [35], (ii) Viridiplantae ITS2 sequences obtained from NCBI, and (iii) unique ITS2 sequences of snow algae detected from red snow in a previous study [6]. Sequence clustering was then conducted with a 98% nt identity level based on the ITS2 sequences in each sample, using the furthest neighbor algorithm in Mothur 1.44.1 [36].

Using the unique sequences of the ITS2 region detected from each geographical region, we divided the distribution into three types: cosmopolitans, multi-regions, and endemics. Cosmopolitan types are distributed at both poles and middle latitudes (detected in Antarctica, Svalbard, Greenland, Alaska, and mid-latitude regions); multi-region types are distributed in two regions (Antarctica and a mid-latitude region, the Arctic and a mid-latitude region, or at both poles only). The endemic types are distributed in one region (either the Arctic, Antarctica, or a mid-latitude region) (Table S8). In addition, ice core samples were investigated for cosmopolitan/endemic phylotypes in the past.

We compared the number of unique sequences obtained by usearch with the amplicon sequence variants (ASVs) obtained by DADA2. All sequences were clustered into ASVs using the R package DADA2 [37] by following the procedure on the DADA2 analysis tutorial web page.

Groups based on the secondary structure of the ITS2 region

For estimating the diversity of snow algae within the snow samples, the ITS2 region sequences were classified at the species level according to the genetic species concept based on structural differences in the ITS2 region [38]. Hereafter, these unique sequences are defined as "phylotypes" and ITS2 sequences with ≥98% nt sequence identity as operational taxonomic units (OTUs). The 5.8S-28S rRNA interaction region within the sequence of

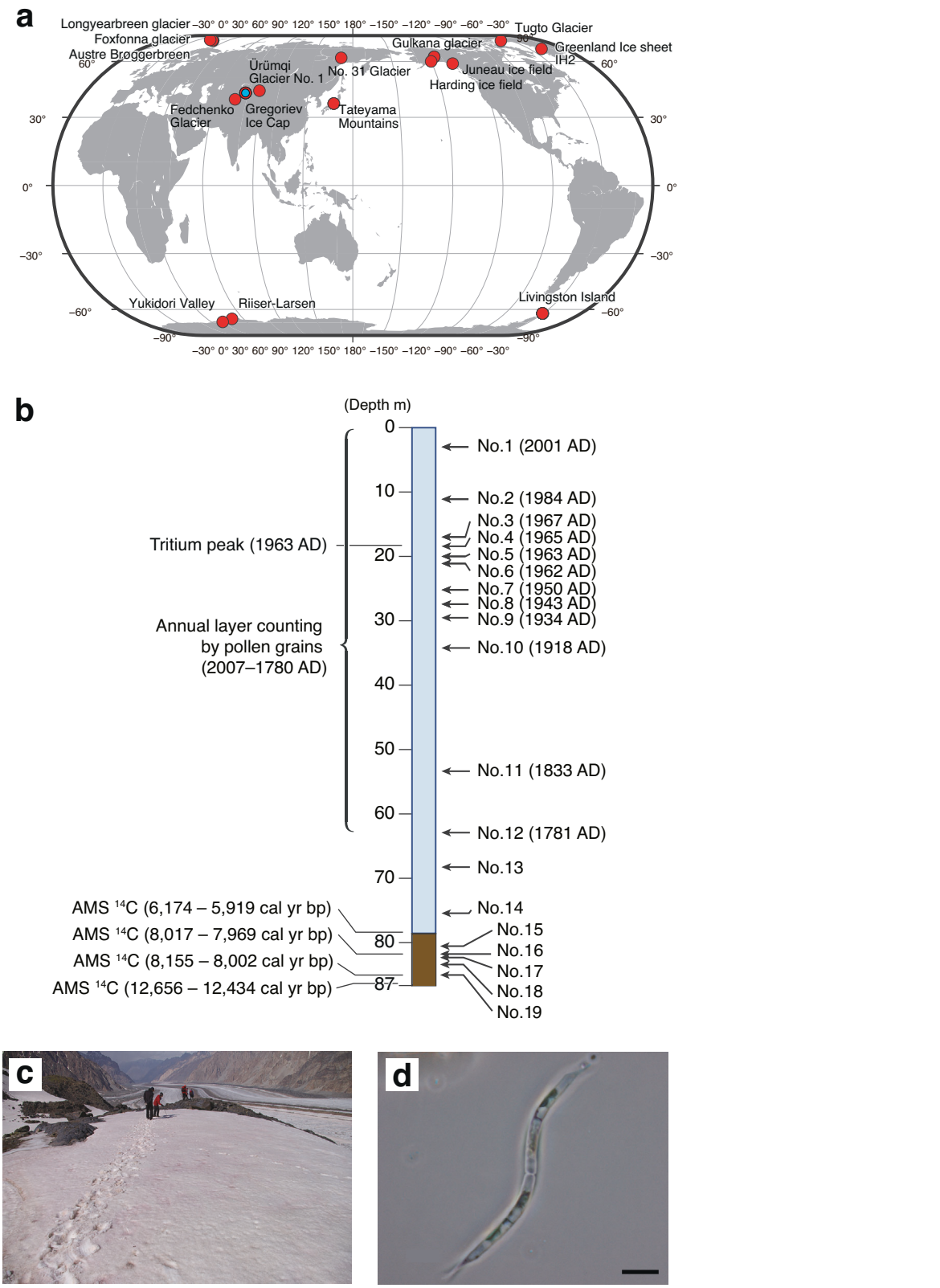

Fig. 1 Collection sites of red snow and ice core samples, ice core information, and photographs of a study site and of a snow alga. **a** Locations of the Grigoriev Ice Cap in the Tien Shan Mountains in the Kyrgyz Republic, where the ice core was drilled (blue circle), and of snow sampling sites (red circles, Tables S1–S2). **b** Schematic drawing of the ice core, with indication of depth from the surface, the dating age of the core, and the positions of the samples analyzed. The depth (in meters) and dating of the dust layer sections are also shown. **c** Red snow on a study site at mid-laditudes (Fedchenko Glacier, Pamir). **d** Microscopic view of *Raphidonema* sp. (collected on Gulkana Glacier, Alaska). Scale bar, 10 μm.

each OTU was annotated using the web interface for Hidden Markov Models–based annotation [39] with the ITS2 database [40]. Secondary structures for ITS2 were predicted using Centroidfold [41] and RNAfold WebServer [42] and were manually refined. We confirmed that the ITS2 secondary structures of the OTUs examined in this study contained four helices as well as a U-U mismatch in helix II and a YGGY motif on the 5′ side near the apex of helix III, which are common structural hallmarks of eukaryotic ITS2 sequences [43, 44]. The ITS2 sequences of the OTUs were then compared with published sequences using BLASTn in the NCBI non-redundant nt database. Based on the BLASTn results, the OTUs were classified into five chlorophycean and six trebouxiophycean groups (Fig. S1 and Tables S3–S4). Within each group, species boundaries among the OTUs were estimated based on the compensatory base change near the apex of helix III encompassing the YGGY motif in the ITS2 secondary structure (the most conserved region of the ITS2 secondary structure of eukaryotes) [45]; the compensatory base change correlates with the separation of biological species [38].

Chimeric clusters were primarily removed by BLASTn to the NCBI nt database. After grouping the sequences based on the secondary structures, the possible chimeric sequences were checked manually and deleted. The unique sequences and OTUs of the algal ITS2 sequence clusters were identified from the taxonomic assignment results, and the remaining unique sequences and OTUs of ITS2 that were not assigned to algae were discarded.

Phylotype network of ITS2 sequences in the *Raphidonema* group

The unique ITS2 sequences of the *Raphidonema* group in the mid-latitudes and the ice core from this study, together with the previous ITS2 sequences of snow algae in Antarctica and the Arctic [6] and those from the NCBI nt database (BLASTn searches were performed to identify the closest sequences in the NCBI nt database, i.e., those with >98.0% similarity) were aligned using the program MAFFT v7.429 [33]. These alignments were carefully checked by eye, and all ambiguous sites and sequences were manually deleted. Sequence clustering was then conducted with ≥98% nt sequence identity using the furthest neighbor algorithm in Mothur 1.44.1 [36]. The maximum-likelihood (ML) tree was inferred using IQ-tree ver. 1.6.12 [46] with the K3P + G4 model, and 1000 replications were carried out for standard bootstrap analysis.

For analyzing the evolutionary approaches of the *Raphidonema* group, the OTUs that were defined with ≥98% nt sequence identity among sequences were subdivided into five subgroups (Groups A–E) based on a phylogenetic analysis. Using the unique sequences of each subgroup, we performed the following molecular evolutionary analyses. Phylotype networks of unique sequences were then constructed within the *Raphidonema* group with the median-joining (MJ) method [47] using NETWORK 4.6.11 (http://www.fluxus-engineering.com/sharenet.htm). Because the number of unique sequences was too large for computational and visual network analysis, 200 unique sequences were used for network analysis, including the top 100 unique sequences with the largest number of reads plus 100 randomly picked unique sequences from the 101st and later unique sequences. We additionally generated multiple sets of random 101st–200th sequences by random selection and confirmed that all results were consistent.

Pairwise differences among the four regions (Antarctica, Arctic, mid-latitudes, and the ice core), and ice cores between the newer (1800–2001 AD) and the older (6000–8000 years before present) layers with respect to unique sequence cluster compositions were statistically analyzed with permutational multivariate analysis of variance (PERMANOVA) with R (version 3.6.1) and the VEGAN package.

Inference of demography based on the coalescent model

Let us consider a simple model of demography: a population expanded t_0 generations ago, and the sizes before and after the expansion were N_0 and N_1, respectively. The time $s > 0$ in terms of generations is measured backward from the present. The population size is N_0 for $s \geq t_0$ and N_1 for $s < t_0$. Rogers and Harpending [48] derived a partial differential equation that is satisfied by an approximation of the probability that the number of mismatched sites between two randomly chosen sequences is k in generation s for sufficiently large k. However, their approximation is not justified for small k. Moreover, their estimation fits the solution curve to the histogram, for which statistical assessment of the estimation is difficult.

Let K be the number of mismatched sites between two randomly chosen sequences. The probability density of the coalescence time for two sequences is: $f(s) = e^{-t_0/(2N_1)}\frac{1}{2N_0}e^{-(s-t_0)/(2N_0)}$ for $s \geq t_0$ and $\frac{1}{2N_1}e^{-s/(2N_1)}$ for $s < t_0$. Given s, we assume that the number of mutations along the lineages to the most recent common ancestor of the two sequences follows a Poisson distribution of intensity $2us$, where u is the mutation rate of the sequence per generation. Then, the probability mass function of K is given as

$$\begin{aligned} P(K=k) &= \int_0^\infty f(s)\frac{e^{-2us}}{k!}(2us)^k ds \\ &= \left(\frac{\theta_1}{1+\theta_1}\right)^k \frac{1}{1+\theta_1}\frac{1}{k!}\gamma\left(k+1,\tau\left(1+\frac{1}{\theta_1}\right)\right) \\ &\quad + \left(\frac{\theta_0}{1+\theta_0}\right)^k \frac{1}{1+\theta_0}\frac{1}{k!}\Gamma\left(k+1,\tau\left(1+\frac{1}{\theta_0}\right)\right)e^{-\tau\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)}, \end{aligned}$$

where γ and Γ are incomplete gamma functions defined as $\Gamma(c,x) = \int_0^x e^{-z}z^{c-1}dz = 1-\gamma(c,x)$, $\tau = 2ut_0$, $\theta_0 = 2N_0u$, and $\theta_1 = 2N_1u$. The expectation $E(K) = \int_0^\infty f(s)2usds = \theta_1 + (\theta_0 - \theta_1)e^{-\tau/\theta_1}$ reproduces Tajima's result [49].

The probability mass function of K may be used as the likelihood; let it be denoted as $L(k;\theta_0,\theta_1,\tau)$. We conducted maximum likelihood estimation based on a composite log-likelihood $\sum_{i=1}^{\lfloor n/2 \rfloor}\log L(k_{2i-1,2i};\theta_0,\theta_1,\tau)$, where $k_{2i-1,2i}$ is the number of mismatched sites between the $(2i-1)$-th and $2i$-th phylotypes in randomly ordered phylotypes. The reason why we did not exhaust all pairs is twofold: to avoid computational burden and to avoid dependence between pairs. The sampling variance-covariance matrix was estimated as the inverse of the observed Fisher information matrix. Note that the composite log-likelihood is still not the true likelihood because any pair of sequences taken from a population may share edges in a genealogy. The correlation among genealogies may underestimate the variances. See Larribe and Fearnhead [50] for further discussion.

The maximum likelihood estimates of τ, θ_0 and θ_1 are shown in Table S13 with standard deviation values. For the cosmopolitans, the standard deviations of θ_0 and θ_1 were not available because the likelihood was constant for $\theta_0 \in [0.108, 0.010]$. If the standard deviation was estimated to be a tiny value, it was omitted. The maximum likelihood estimate provides estimates of $\tau = 2ut_0, \theta_0 = 2N_0u$, and $\theta_1 = 2N_1u$, but we were interested in estimates of t_0, N_0, and N_1. To derive estimates of N_0, N_1, and t_0 in years, the mutation rate and the generation time are needed. Ness et al. [51] estimated the single-base mutation rate of the unicellular green alga *Chlamydomonas reinhardtii* to be 2.08×10^{-10} (95% CI: 1.09×10^{-10} to 3.74×10^{-10}) per site per generation based on a mutation accumulation experiment followed by whole-genome resequencing of two replicate lines. Because the ITS sequence has 381 base pairs, we assumed the mutation rate to be $u = 2.08 \times 10^{-10} \times 381 = 7.9 \times 10^{-8}$ per generation. The generation time of an alga corresponds to the doubling time of algal population size. On the Greenland Ice Sheet, doubling times have been estimated as 3.75–5.5 days [52], while Onuma et al. [53] estimated the growth rate per day to be 0.42, which leads to a doubling time of (log 2)/0.42 = 1.65 days. Algae grow only during the period when snow melts and the duration depends on both altitude and latitude, but a typical duration of algal growth is ~2 months. Taking the reciprocal as the effective doubling time yields 11–36 days. Therefore, we assumed the generation time (annual mean) to be 24 days.

RESULTS

Classification of snow algae in the ice core based on ITS2 sequences

We used high-throughput sequencing to obtain DNA sequences of algae from 19 layers of an ice core drilled on a glacier in central Asia, dated from present time to 8000 years ago (Fig. 1 and Table S1). In total, 17,016 unique sequences (phylotypes) for the fast-evolving algal nuclear rDNA internal transcribed spacer 2 (ITS2) region were determined in the ice core, from which 290 OTUs were defined with ≥98% nt sequence identity among all OTUs.

The ITS2 sequences were classified at the species level according to the genetic species concept based on secondary structural differences in the ITS2 region, which correlate with the boundaries of most biological species [38]. The ITS2 sequences from ice core samples were classified into 24 subgroups consisting of 17 chlorophycean, 5 trebouxiophycean, and 2 ulvophycean groups based on their secondary structures and BLASTn results

(Fig. S1 and Tables S3–S4). The 17 subgroups of Chlorophyceae were subdivided into 10 subgroups of the *Chloromonadinia* clade, 1 subgroup of the *Monadinia* clade (recently treated as the genus *Microglena* [54]), 3 subgroups of the *Reinhardtinia* clade, 2 subgroups of the *Stephanosphaerinia* clade, and 1 subgroup corresponding to an unnamed group (which is related to *Ploeotila* sp. CCCryo 086-99) (for the clade names, see [55]). Although the *Chloromonadinia* clade contains several snow species belonging to *Chloromonas* or *Chlainomonas*, the 10 subgroups of the *Chloromonadinia* clade were considered to be *Chloromonas*. The 5 trebouxiophycean subgroups were composed of 2 subgroups of the *Chlorella* group, 1 subgroup of the *Raphidonema* group, 1 subgroup of the *Trebouxia* group, and 1 subgroup of the *Neocystis* group. The 2 subgroups of Ulvophyceae were closely related to the genus *Chamaetrichon* and *Planophila*, respectively. It is noted that *Sanguina* ('*Chlamydomonas*'-snow group B [6]), *Ancylonema*, and *Mesotaenium*, which are snow algal genera found throughout the world [56, 57], were not detected in the ice core samples (Tables S3–S4).

Global distribution of the *Raphidonema* group

To understand the process by which snow algae form geographically specific population structures and how they migrate globally across the glaciers and snow fields, it is necessary to focus on the microbial species that inhabit the global cryosphere. Previous work elucidated that the *Raphidonema* group and '*Chlamydomonas*'-snow group B (*Sanguina*) are the cosmopolitans at both poles [6], but the latter was not detected in ice core samples examined in this study. Therefore, to elucidate the evolutionary history of the *Raphidonema* group, we further analyzed the ITS2 sequences obtained from the ice core sample as well as the glacier-surface samples from both poles [6] and from the mid-latitudes (samples from 10 sites, obtained in this study) (hereafter, surface samples; Table S2). Members of the *Raphidonema* group were detected in the older (deep core) layers of the ice core and at the glacier surface of central Asia (Fig. S1 and Tables S3–S4), as well as in the red snow samples from both poles [6]. In central Asia, the *Raphidonema* group was found in the Russian, Chinese, and Kyrgyz samples but was not detected in the Japanese and Tajik samples (Fig. S1 and Tables S3–S4). Combining these sequences yielded 893,649 reads and 22,389 unique sequences for subsequent detailed analysis (Tables S5–S6). The taxonomic composition of the *Raphidonema* communities differed among the mid-latitude, ice core, Arctic, and Antarctic samples as determined by PERMANOVA (Table S7). Most of the unique sequences in the *Raphidonema* group were consistent with an endemic distribution (Tables S8–S10). An average of 77% of the unique sequences of the *Raphidonema* group were endemic to a specific region (mid-latitude, 96%; Antarctic, 66%; Arctic, 79%), accounting for 40% of the total sequencing reads (mid-latitude, 77%; Antarctic, 74%; Arctic, 22%) (Fig. 2a, b and Tables S9–S10). This result suggested that most of the unique sequences are endemic, indicating that their dispersal has been limited to their respective regions [58–61].

Next, we analyzed the global distribution of the cosmopolitan phylotypes of the *Raphidonema* group, because a previous study analyzed their distribution only at the poles [6]. Only a limited number of unique sequences were distributed in all regions (mid-latitude, 1.4%; Antarctic, 5.6%; Arctic, 3.1%), accounting for a large proportion of the sequencing reads in polar regions but for only a small proportion in the mid-latitudes (mid-latitude, 2.8%; Antarctic, 20%; Arctic, 55%) (Figs. 2a, b, S2–S3, and Tables S9–S10).

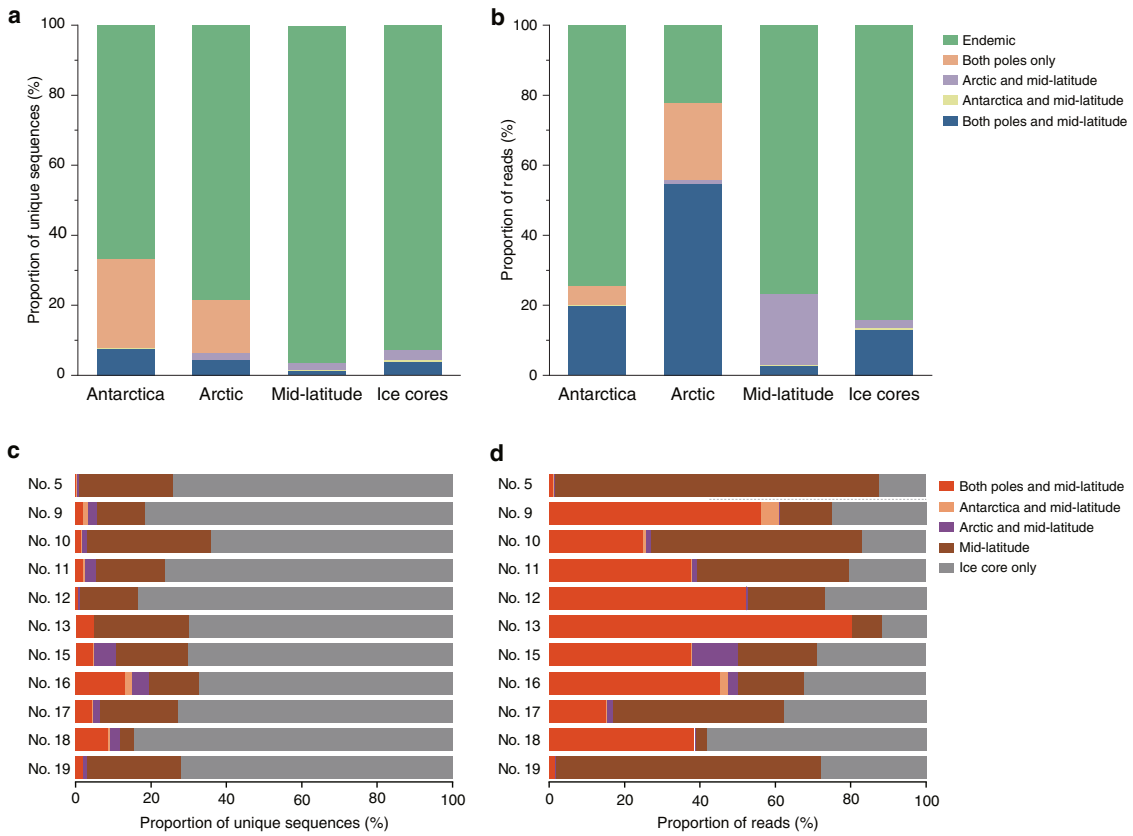

Fig. 2 Distribution types of the Raphidonema group obtained from each region and the ice core based on ITS2 unique sequences. Proportions of unique sequence and number of sequencing reads are shown. **a** Unique sequences from surface snow and ice-core samples. **b** Number of sequencing reads from surface snow and ice core samples. **c** Unique sequences from the indicated locations within the ice core. **d** Number of sequencing reads of the unique sequences from the indicated locations within the ice core.

The distribution types of the *Raphidonema* group obtained from each region and the ice core were similar between the USEARCH and DADA2 analyses (Figs. 2, S4). In addition, we note that in ancient samples, post-mortem nt substitutions, such as cytosine to thymine, accumulate over many years of deposition [62], and these are not included in the DADA2 error model, which leads to the elimination of minor sequences in the DADA2 analysis. Therefore, we based our analysis on the results of the USEARCH unique sequences. These results suggested that only a few snow algae in the *Raphidonema* group were detected in samples from the mid-latitude regions.

Snow algae of the *Raphidonema* group were detected in different ice core layers, corresponding to different time periods. The ice core records revealed that the distribution types of the *Raphidonema* group have not changed significantly for the last 8000 years, with $p = 0.1924$ based on a PERMANOVA between the newer (1800–2001 AD) and the older (6000–8000 years before present) layers (Fig. 2c, d). In ice core samples, 77% of the unique sequences of the *Raphidonema* group were detected only in the ice core samples, accounting for 23% of the total sequencing reads (Fig. S5). Although some of these unique sequences may be artifacts of the post-mortem nt substitution or sequencing errors, because we conducted sequence quality filtering and removed the majority of artifact sequences by removing the singleton clusters, most of the unique sequences in the ice core are not likely to be artifacts, but they could represent endemic phylotypes (Figs. 2a, b, S5).

The cosmopolitan phylotypes were detected over a broad period as represented by ice core samples. They were present in approximately similar ratios in the newer and older layers (Fig. 2c, d). The cosmopolitan phylotypes were relatively abundant in the ice core samples (average, 4.0%; range, 0.2–13%), accounting for 13% (0.9–81% in the samples) of the total sequencing reads (Figs. 2c, d and S5).

Microevolution of cosmopolitan and endemic phylotypes

We analyzed the evolutionary relationship between cosmopolitan and endemic phylotypes of the *Raphidonema* group among all snow surface and ice core samples. In total, 22,389 unique sequences of the *Raphidonema* group were clustered into 170 OTUs that were defined with ≥98% nt sequence identity among sequences within OTUs. The OTU sequences were subdivided into five subgroups (Groups A–E) based on phylogenetic analysis (Figs. S6–S11 and Tables S11–S12). Based on a previous study [63], Groups A–C and Group E were assigned to *R. sempervirens* and *R. nivale*, respectively, but Group D was not consistent with any species examined in that study (Fig. S6).

The phylotypes were categorized into three subsets: the cosmopolitan phylotypes found at both poles and the mid-latitude regions; the multi-region phylotypes found in any two of the Antarctic, Arctic, and mid-latitude regions; and the endemic phylotypes found in only one of the three regions. Cosmopolitan phylotypes were found in Groups A, B, and C and accounted for 64.6% of the unique sequences. We then analyzed the dispersal of the three groups in detail.

MJ networks [47] for the ITS2 sequences in each subgroup revealed that the cosmopolitan phylotypes were located at the center of the networks in Groups A and C that contained any types (endemics, multi-regions, and cosmopolitans) of the phylotypes, whereas the endemic phylotypes were considered to be derived from the cosmopolitan phylotypes (Figs. 3 and S12–S13). Moreover, the outgroup phylotypes were directly connected to the cosmopolitan phylotypes. These findings clearly showed that the cosmopolitan phylotypes were ancestral, whereas the endemic

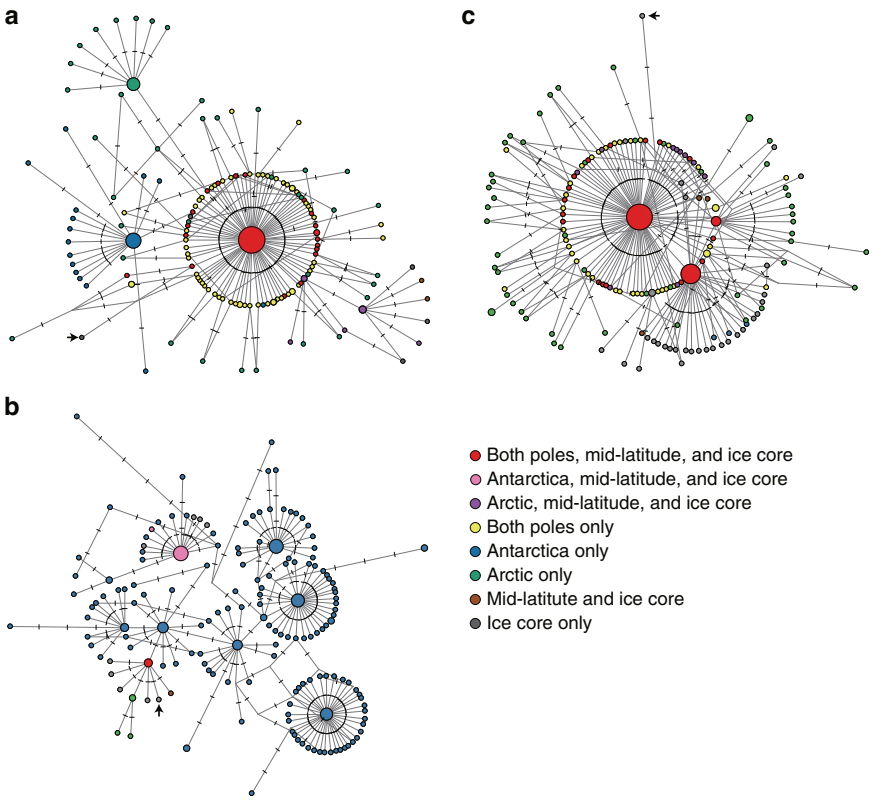

Fig. 3 Phylogenetic relationships among phylotypes of the *Raphidonema* groups. Phylotype networks for ITS2 sequences within Groups A (**a**), B (**b**), and C (**c**) of the Raphidonema group that include the cosmopolitan phylotypes in this study. The median-joining method was used. Circles indicate phylotypes; the size of each circle is proportional to the number of unique sequences. Each notch on the edges represents a mutation. Phylotypes are colored according to geographic region. The arrow represents the phylotype in the outgroup (see Fig. S6).

phylotypes were derived. In contrast, there were remarkable differences in the shape of the networks between Group B and the others (Groups A and C). In Group B, the Antarctic endemic phylotypes formed a distinct clade, and multi-region phylotypes seemed to be recently derived from this clade. In addition, the Arctic endemic phylotypes formed another distinct clade. These two Group-B clades split directly from a cosmopolitan phylotype (5.3% of the total sequencing reads). For Groups A and C, however, major portions of the total sequencing reads belonged to cosmopolitan phylotypes in Groups A (48.2%) and C (62.4%), and the endemic and multi-region phylotypes were directly connected to these major cosmopolitan phylotypes in a radial manner—the so-called "star-like" pattern [64]. These contrasting network shapes seem to have been formed as a consequence of the unique evolution of each of these groups. We also found that sequences from ice cores did not represent a basal position (Figs. 3 and S12–S13). This is because the haplotypes found in the modern samples have existed from times earlier than the ice core ages, due to the very small mutation numbers expected to have occurred since the ice core ages. Therefore, detected ice core ages were not included in the molecular evolution calculations of our demographic model. However, the phylogenetic networks themselves do not provide information on the evolutionary time scale. Hence, the ice core samples provide further direct evidence that *Raphidonema*, especially cosmopolitans belonging to this genus, persistently grew on snow and ice at least during the Holocene, and their ITS2 sequences have not changed over the last 8000 years.

Referring to "ancestral" phylotypes as those having a longer history than other, more recently derived phylotypes, it is possible that individuals not closely related can share the same ancestral phylotype. In such cases, if genetically far-related individuals from various geographical regions share the same ancestral phylotype, they appear to be "cosmopolitan" (Fig. S14a). In order to distinguish between these "apparent cosmopolitans", and "true cosmopolitans" that migrate globally, it is necessary to show that the cosmopolitan and endemic phylotypes have distinct demographic histories rather than being part of a continuous population sharing certain demographic dynamics (Fig. S14). Because phylotype networks are not useful for quantifying the rate(s) of microevolution, we used the coalescent model to quantify phylotype demographics [65]. As numerous phylotypes must be analyzed with this approach, we concentrated on statistical inference based on pairwise comparisons of phylotypes, for which the likelihood can be determined in a practical manner (see Materials and Methods). Histograms for the number of mismatched sites between two phylotypes chosen from a set of phylotypes, which will be called the pairwise mismatch distribution, are shown in Figs. 4 and S15. For Groups A and C, the distribution among cosmopolitans, multi-regions, and endemics was unimodal, in which the modes align from left to right with the order cosmopolitans, multi-regions, and endemics. Rogers and Harpending [48] noted that this "wave" propagation results from the expansion in size of a population, which leads to large mismatches, and the mode shifts to the right (see Fig. 2 of [48]). As time passes, the mode shifts to the left and eventually returns to the origin, i.e., representing a population that has not undergone an expansion event. Rogers and Harpending obtained an approximate solution for the wave and fitted the solution to human mitochondrial sequence data. We improved upon their method based on the coalescent model (see Materials and Methods) and applied it to the ITS2 sequence data for snow algae.

For Group A, when we fit the single demographic model to all phylotypes, the log-likelihood was –414,487. In contrast, when we fit the demographic model to each subset, that is, cosmopolitans, multi-regions, and endemics, separately, the log-likelihood was –341,964. Because the latter is larger than the former, we fit the model to each subset of phylotypes separately. For Group C, when we fit the demographic model to the cosmopolitans, multi-regions, and endemics separately, the log-likelihood was –142,106, which is larger than the log-likelihood, –218,080, when we fit the single demographic model to all phylotypes. In contrast to Groups A and C then, we fit the single demographic model to all phylotypes of Group B because the log-likelihood, –196,070, was

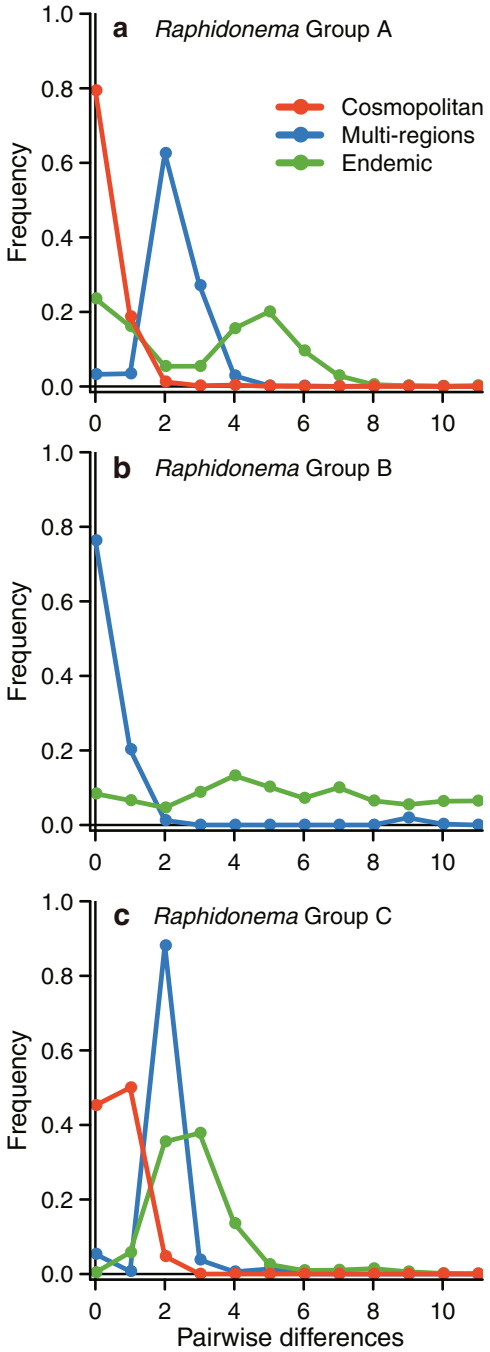

Fig. 4 Mismatch distribution based on the number of pairwise differences in each distribution type in *Raphidonema* groups. The lines represent the observed number of pairwise differences in each distribution type (cosmopolitan, multi-region, endemic) within the *Raphidonema* Groups A (**a**), B (**b**) and C (**c**). Calculations were performed for all distribution types of *Raphidonema* Groups A and C, for which various cosmopolitan phylotypes were detected. On the other hand, calculations for only multi-region and endemic phylotypes were performed for *Raphidonema* group B, because no variation was found in cosmopolitan phylotypes.

larger than the log-likelihood, −220,145, when we fit the demographic model to the cosmopolitans, multi-regions, and endemics separately. These results suggested that cosmopolitans, multi-regions, and endemics experienced different demographic histories in Groups A and C, whereas they had the same demographic history in Group B (Table S13). These results indicate the cosmopolitans in Group A and C are true cosmopolitans, whereas the those in Group B can be regarded as an apparent cosmopolitan.

The ML estimates of $\tau = 2ut_0$, $\theta_0 = 2N_0u$, and $\theta_1 = 2N_1u$ are shown in Table S13 with standard deviation values. The population expanded t years ago, with the size before and after the expansion being represented by N_0 and N_1, respectively. The mutation rate (u) was assumed to be 7.9×10^{-8}/ sequence/generation, and the generation interval was assumed to be 24 days (Materials and Methods). In Group A, for the cosmopolitans, the estimates of t, N_0, and N_1 were $33.8/(2 \times 7.9) \times 10^8 \times \frac{24}{365} = 1.4 \times 10^7$ years, $(0.108 - 0.010)/(2 \times 7.9) \times 10^8 = (6.8 - 0.63) \times 10^5$, and $(0.217)/(2 \times 7.9) \times 10^8 = 1.4 \times 10^6$, respectively. In the same way, we computed estimates of t, N_0, and N_1 of other phylotypes and other groups (Table S14). For the endemics, the respective values were 9.2×10^6 years, 80, and 2.1×10^7, and the values were 4.6×10^6 years, 139, and 1.5×10^7 for the multi-regions. Taking into account the minimum and maximum ranges of the mutation rates per generation as well as the generation intervals, t for cosmopolitans was 3.6×10^6–4.0×10^7 years ago, and t for endemics was 2.3×10^6–2.6×10^7 years ago (Table S14). These results suggested that the cosmopolitans existed at least 1.4×10^7 years ago, and the endemics were derived from the cosmopolitans 9.2×10^6 years ago. The size of the endemics expanded 2.6×10^5-fold, which may have resulted from extensive dispersal. The multi-regions tended to mimic the endemics. Note that our demographic model was simplified to avoid overparameterization. In reality, considering the branching patterns of the MJ network, it is plausible that the endemic phylotypes have been repetitively and continuously derived from the cosmopolitans in multiple lineages—from 9.2×10^6 years ago to the present. In the same way, as for Group C, our results suggested that the cosmopolitan population expanded 3.9-fold ~3.2×10^6 years ago, and the endemics were derived from the cosmopolitans 1.9×10^5 years ago. The size of the endemics expanded 59-fold. In contrast to the phylotypes of Groups A and C, those of Group B experienced no significant expansion (Supplementary Results). In Groups A and C, the derived endemics (and multi-regions) expanded greatly as compared with the ancestral cosmopolitans (Table S14). These extraordinary expansions constitute evidence for local adaptation by the endemic/multi-region populations. In contrast, there was no evidence of local adaptation in Group B. The mismatch distribution of the entire Group B (multi-regions + endemics) showed a multimodal pattern (Fig. 4), which is present in the populations with stable sizes for a long period. When the populations finally reach equilibrium, the mismatch distributions show the exponential distribution [48]. Based on our ML estimates (Table S14), the historical population of Group B has been stable.

DISCUSSION

Our study provides substantive insight into the global microbial dispersal process, which has been controversial for more than a century [1]. The ongoing regional expansion of snow algae is likely to be affected substantially by global climate change because the cryosphere is particularly sensitive to the climate. Our results demonstrate that the current microbial distribution was maintained throughout the Holocene, i.e., at least for the last 8000 years, based on the analysis of ancient DNA from ice core samples. Furthermore, our inference of the demography of snow algae suggests that local adaptation of the endemic phylotype of snow algae occurred millions of years ago rather than after the last glacial period. Microorganisms residing in snow and ice were affected by variations in the extent of the cryosphere associated with climatic changes such as glacial-interglacial cycles. However, the detailed association between environmental fluctuations and the evolution of microorganisms in the cryosphere remains unknown. Therefore, elucidating this association will shed more light on the evolution of microorganisms that populated snow and ice in past glacial cycles.

Our results reveal the existence of cosmopolitan phylotypes, which have dispersed globally and regionally to form endemic phylotypes. Although the *Raphidonema* group makes up the majority of present-day cosmopolitan species [6], details pertaining to their phenotypic variations and life cycles remain unknown. Two main explanations have been given for how snow algae disperse globally, namely by aerial transport and by animal transport (e.g., via fecal droppings) [4]. Based on field surveys and laboratory experiments [66, 67], vegetative cells of *Raphidonema* can be transferred by strong winds and dispersed via atmospheric circulation. Chlorophycean snow algae (e.g., *Sanguina* and *Chloromonas*) generally produce zygotes or cysts to become dormant with resistance to desiccation and possibly to UV radiation. Such zygotes/cysts are suitable for long-distance dispersal; however, zygote formation of *Raphidonema* was reported only once in 1953 [68], but has not been observed in subsequent studies using field-collected and cultured materials (e.g., [69, 70]). Therefore, further research is necessary to understand the mechanism by which microorganisms are dispersed via the atmosphere and how they ultimately accumulate in the global cryosphere, as well as any potential effects of UV radiation.

A previous study has suggested that snow species of *Raphidonema* (*R. sempervirens* and *R. nivale*) are actually soil algae that are only occasionally transported to snow [66]. This appears to be similar to some terrestrial microalgae, such as *Chlorella* and *Stichococcus* (Trebouxiophyceae), which have broad temperature ranges for growth and are also found on snow/glacier surfaces. We note, however, that the optimum growth temperatures for both species are below 6 °C, and *R. sempervirens* cannot grow above 13 °C [71]. In addition, the temperature ranges for growth for *R. nivale* and *R. tatrae* are 0–15 °C and 0–10 °C, and the optimum temperatures are 5 and 4 °C, respectively [72, 73]. Therefore, if these species were truly soil algae, their habitats would likely be restricted to cold environments, such as alpine and polar regions. Recent amplicon sequencing analysis for soils from the French Alps reported that *Raphidonema* spp. were rarely detected from soils at high elevations, although the samples contained enough *Sanguina* spp., a well-known genus of snow algae that are associated with red snow, to be detected [74]. Therefore, a large proportion of *Raphidonema* sequences in the surface samples indicated that the snow species of the genus probably grow on snow and ice rather than simply being transported from the periglacial soils.

Although *Raphidonema* is reported to be a minor genus as compared with other snow/glacier algae (e.g., *Ancylonema*, *Mesotaenium*, and *Sanguina*) on modern snow surfaces [75], the latter "dominant" snow/glacier algae were not detected in the ice cores analyzed in this study. According to the altitudinal distribution of phototrophs on a glacier in Alaska [76], *Raphidonema* is categorized as an opportunist, appearing on the surface under certain conditions. The ice core samples analyzed in this study were from central Asia, where the pH of glacial meltwater is higher than in other areas due to abundant carbonates derived from nearby deserts [77]. Under these conditions, cyanobacteria are most dominant, and only a few green algae grow on such glaciers (e.g., [78]). According to Hoham and Remias [75], snow green algae are usually found in snow with an acidic pH. We believe that this higher pH is probably one reason for the absence of green algae other than *Raphidonema* in this area, although future studies are needed to confirm this point.

In conclusion, we examined the genetic structures of snow algae from both poles and mid-latitude high-mountain regions to determine the history of biogeographical distributions and evolutionary relationships between cosmopolitan and endemic phylotypes of snow algae. We found that the genus *Raphidonema* (trebouxiophycean group) was distributed over both poles and mid-latitude regions and was detected in different ice core layers, corresponding to distinct time periods. A phylogenetic network analysis consistently showed that the cosmopolitan phylotypes were located near the root positions, whereas the endemic phylotypes were generally located at the tips, suggesting that the endemic phylotypes originated from the cosmopolitan phylotypes. To examine this hypothesis, we applied a coalescent theory–based demographic model to our data, which confirmed the better fit of an evolutionary scenario with the assumption that the cosmopolitan and endemic phylotypes experienced different population histories, as compared with a scenario with the assumption that these phylotypes shared common population histories. The evolutionary parameters estimated using this model indicated that the modern cosmopolitan phylotypes were present long before the last glacial period (Supplementary Results, Table S14) and that the endemic phylotypes originated from the ancestral cosmopolitan phylotypes. These findings suggest that the cosmopolitans dispersed worldwide, from which were derived new localized endemics.

Although the local adaptations of endemics were inferred from their demographic histories, a precise picture of how newly derived endemics have dispersed and adapted to local environments and of the process by which they have migrated across the global cryosphere is still unclear. Future work will benefit from single-cell genomic technologies to shed light on their morphological and metabolic aspects and will provide new insights into their dispersal and adaptive evolution.

DATA AVAILABILITY

The raw sequence data are available under DDBJ DRA ID DRA012482. The nucleotide alignments of 98% sequence clustering of unique ITS2 sequences are available from http://redsnow2021.paleogenome.jp/.

REFERENCES

1. De Wit R, Bouvier T. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol. 2006;8:755–8.
2. Hahn MW, Jezberová J, Koll U, Saueressig-Beck T, Schmidt J. Complete ecological isolation and cryptic diversity in Polynucleobacter bacteria not resolved by 16S rRNA gene sequences. ISME J. 2016;10:1642–55.
3. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T, et al. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J. 2018;12:742–55.
4. Brown SP, Tucker AE. Distribution and biogeography of Sanguina snow algae: Fine-scale sequence analyses reveal previously unknown population structure. Ecol Evol. 2020;10:11352–61.
5. Okazaki Y, Fujinaga S, Salcher MM, Callieri C, Tanaka A, Kohzu A, et al. Microdiversity and phylogeographic diversification of bacterioplankton in pelagic freshwater systems revealed through long-read amplicon sequencing. Microbiome 2021;9:24.
6. Segawa T, Matsuzaki R, Takeuchi N, Akiyoshi A, Navarro F, Sugiyama S, et al. Bipolar dispersal of red-snow algae. Nat Commun. 2018;9:3094.
7. Chown SL, Clarke A, Fraser CI, Cary SC, Moon KL, McGeoch MA. The changing form of Antarctic biodiversity. Nature 2015;522:431–8.
8. Segawa T, Yonezawa T, Edwards A, Akiyoshi A, Tanaka S, Uetake J, et al. Biogeography of cryoconite forming cyanobacteria on polar and Asian glaciers. J Biogeogr. 2017;44:2849–61.
9. Thompson PL, Fronhofer EA. The conflict between adaptation and dispersal for maintaining biodiversity in changing environments. Proc Natl Acad Sci USA. 2019;116:21061–7.
10. Karcher N, Pasolli E, Asnicar F, Huang KD, Tett A, Manara S, et al. Analysis of 1321 Eubacterium rectale genomes from metagenomes uncovers complex phylogeographic population structure and subspecies functional adaptations. Genome Biol. 2020;21:138.
11. Priscu J, Christner B, Foreman C, Royston-Bishop G. Biological Material in Ice Cores. In: Elias S, editor. Encyclopaedia of Quaternary Sciences, 2nd Edn. London: Elsevier; 2007. p. 1156–66.
12. Gawor J, Grzesiak J, Sasin-Kurowska J, Borsuk P, Gromadka R, Górniak D, et al. Evidence of adaptation, niche separation and microevolution within the genus Polaromonas on Arctic and Antarctic glacial surfaces. Extremophiles 2016;20:403–13.
13. Hotaling S, Shain DH, Lang SA, Bagley RK, Tronstad LM, Weisrock DW, et al. Long-distance dispersal, ice sheet dynamics and mountaintop isolation underlie the genetic structure of glacier ice worms. Proc R Soc B 2019;286:20190983.
14. Liu Q, Liu H-C, Zhou Y-G, Xin Y-H. Microevolution and Adaptive Strategy of Psychrophilic Species Flavobacterium bomense sp. nov. Isolated From Glaciers. Front Microbiol. 2019;10:1069.
15. Vimercati L, Darcy JL, Schmidt SK. The disappearing periglacial ecosystem atop Mt. Kilimanjaro supports both cosmopolitan and endemic microbial communities. Sci Rep. 2019;9:10676.
16. Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, Shapiro B, et al. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science 2003;300:791–5.
17. Segawa T, Ushida K, Narita H, Kanda H, Kohshima S. Bacterial communities in two Antarctic ice cores analyzed by 16S rRNA gene sequencing analysis. Polar Sci. 2010;4:214–27.
18. Miteva V, Rinehold K, Sowers T, Sebastian A, Brenchley J. Abundance, viability and diversity of the indigenous microbial populations at different depths of the NEEM Greenland ice core. Polar Res. 2015;34:25057.
19. Miteva V, Sowers T, Schüpbach S, Fischer H, Brenchley J. Geochemical and Microbiological Studies of Nitrous Oxide Variations within the New NEEM Greenland Ice Core during the Last Glacial Period. Geomicrobiol J. 2016;33:647–60.
20. Zhong Z-P, Solonenko NE, Gazitúa MC, Kenny DV, Mosley-Thompson E, Rich VI, et al. Clean Low-Biomass Procedures and Their Application to Ancient Ice Core Microorganisms. Front Microbiol. 2018;9:1094.
21. Liu Y, Priscu JC, Yao T, Vick-Majors TJ, Michaud AB, Sheng L. Culturable bacteria isolated from seven high-altitude ice cores on the Tibetan Plateau. J Glaciol. 2019;65:29–38.
22. Zhong Z-P, Tian F, Roux S, Gazitúa MC, Solonenko NE, Li Y-F, et al. Glacier ice archives nearly 15,000-year-old microbes and phages. Microbiome 2021;9:160.
23. Fujita K, Takeuchi N, Nikitin SA, Surazakov AB, Okamoto S, Aizen VB, et al. Favorable climatic regime for maintaining the present-day geometry of the Gregoriev Glacier, Inner Tien Shan. Cryosphere. 2011;5:539–49.
24. Takeuchi N, Fujita K, Aizen VB, Narama C, Yokoyama Y, Okamoto S, et al. The disappearance of glaciers in the Tien Shan Mountains in Central Asia at the end of Pleistocene. Quat Sci Rev. 2014;103:26–33.
25. Segawa T, Takeuchi N, Rivera A, Yamada A, Yoshimura Y, Barcaza G, et al. Distribution of antibiotic resistance genes in glacier environments. Environ Microbiol Rep. 2013;5:127–34.
26. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 2013;499:74–8.
27. Coleman AW, Suarez A, Goff LJ. Molecular delineation of species and syngens in volvocacean green algae (Chlorophyta). J Phycol. 1994;30:80–90.
28. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012;40:e3.
29. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–i90.
30. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26:2460–1.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
32. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.
33. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
35. Abarenkov K, Henrik Nilsson R, Larsson K-H, Alexander IJ, Eberhardt U, Erland S, et al. The UNITE database for molecular identification of fungi – recent updates and future perspectives. N Phytologist. 2010;186:281–5.
36. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl Environ Microbiol. 2009;75:7537–41.
37. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
38. Coleman AW. Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. Mol Phylogenet Evol. 2009;50:197–203.

39. Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M. 5.8S–28S rRNA interaction and HMM-based ITS2 annotation. Gene 2009;430:50–7.
40. Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R, et al. The ITS2 Database III—sequences and structures for phylogeny. Nucleic Acids Res. 2010;38:D275–D9.
41. Hamada M, Kiryu H, Sato K, Mituyama T, Asai K. Prediction of RNA secondary structure using generalized centroid estimators. Bioinformatics 2009;25:465–73.
42. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA Websuite. Nucleic Acids Res. 2008;36:W70–W4.
43. Coleman AW. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet. 2003;19:370–5.
44. Schultz J, Maisel S, Gerlach D, MüLler T, Wolf M. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 2005;11:361–4.
45. Coleman AW. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucleic Acids Res. 2007;35:3322–9.
46. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32:268–74.
47. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48.
48. Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol. 1992;9:552–69.
49. Tajima F. The effect of change in population size on DNA polymorphism. Genetics 1989;123:597–601.
50. Larribe F, Fearnhead P. On composite likelihoods in statistical genetics. Statistica Sin. 2011;21:43–69.
51. Ness RW, Morgan AD, Colegrave N, Keightley PD. Estimate of the Spontaneous Mutation Rate in Chlamydomonas reinhardtii. Genetics 2012;192:1447–54.
52. Williamson CJ, Cameron KA, Cook JM, Zarsky JD, Stibal M, Edwards A. Glacier Algae: A Dark Past and a Darker Future. Front Microbiol. 2019;10:524.
53. Onuma Y, Takeuchi N, Tanaka S, Nagatsuka N, Niwano M, Aoki T. Observations and modelling of algal growth on a snowpack in north-western Greenland. Cryosphere. 2018;12:2147–58.
54. Demchenko E, Mikhailyuk T, Coleman AW, Pröschold T. Generic and species concepts in Microglena (previously the Chlamydomonas monadina group) revised using an integrative approach. Eur J Phycol. 2012;47:264–90.
55. Nakada T, Misawa K, Nozaki H. Molecular systematics of Volvocales (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic analyses. Mol Phylogenet Evol. 2008;48:281–91.
56. Kol E. Kryobiologie. Biologie und Limnologie des Schnees und Eises. I. Kryovegetation. In: Elster H-J, Ohle W, editors. Die Binnengewässer 24. Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung (Nägele u. Obermiller); 1968. p. 1–216 with 16 pls. German.
57. Procházková L, Leya T, Křížková H, Nedbalová L. Sanguina nivaloides and Sanguina aurantia gen. et spp. nov. (Chlorophyta): the taxonomy, phylogeny, biogeography and ecology of two newly recognised algae causing red and orange snow. FEMS Microbiol Ecol. 2019;95:fiz064.
58. Lester SE, Ruttenberg BI, Gaines SD, Kinlan BP. The relationship between dispersal ability and geographic range size. Ecol Lett. 2007;10:745–58.
59. Paul JR, Morton C, Taylor CM, Tonsor SJ. Evolutionary Time for Dispersal Limits the Extent but Not the Occupancy of Species' Potential Ranges in the Tropical Plant Genus Psychotria (Rubiaceae). Am Naturalist. 2009;173:188–99.
60. McCauley SJ, Davis CJ, Werner EE, Robeson MS II. Dispersal, niche breadth and population extinction: colonization ratios predict range size in North American dragonflies. J Anim Ecol. 2014;83:858–65.
61. Sheard C, Neate-Clegg MHC, Alioravainen N, Jones SEI, Vincent C, MacGregor HEA, et al. Ecological drivers of global gradients in avian dispersal inferred from wing morphology. Nat Commun. 2020;11:2463.
62. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 2013;29:1682–4.
63. Yakimovich KM, Gauthier NPG, Engstrom CB, Leya T, Quarmby LM. A Molecular Analysis of Microalgae from Around the Globe to Revise Raphidonema (Trebouxiophyceae, Chlorophyta). J Phycol. 2021;57:1419–32.
64. Allaby M. A Dictionary of Ecology. 4th ed. Oxford: Oxford University Press; 2010.
65. Wakely J. Coalescent Theory: An Introduction. New York: Roberts and Company Publishers; 2016.
66. Stibal M, Elster J. Growth and morphology variation as a response to changing environmental factors in two Arctic species of Raphidonema (Trebouxiophyceae) from snow and soil. Polar Biol. 2005;28:558–67.
67. Komárek J, Nedbalová L. Green Cryosestic Algae. In: Seckbach J, editor. Algae and Cyanobacteria in Extreme Environments. Dordrecht: Springer Netherlands; 2007. p. 321–42.
68. Korshikov C. Viznachnik prisnovodnihk vodorostey Ukrainsykoi RSR [Vyp] V. Pidklas Protokokovi (Protococcineae). Bakuol'ni (Vacuolales) ta Protokokovi (Protococcales) [The Freshwater Algae of the Ukrainian SSR. V. Sub-Class Protococcineae. Vacuolales and Protococcales]. Kyjv [Kiev]: Akad. NAUK URSR;1953. Ukrainian.
69. Hoham RW. Pleiomorphism in the snow alga, *Raphidonema nivale* Lagerh. (Chlorophyta), and a revision of the genus *Raphidonema* Lagerh. Syesis. 1973;6:255–63.
70. Novis PM. New records of snow algae for New Zealand, from Mt Philistine, Arthur's Pass National Park. NZ J Bot. 2002;40:297–312.
71. Leya T. Feldstudien und genetische Untersuchungen zur Kryophilie der Schneealgen Nordwestspitzbergens. Berlin, German: Humboldt-Universität zu Berlin; 2004.
72. Hindák F, Komárek J. Cultivation of the cryostonic alga Koliella tatrae (Kol) Hind. Biol Plant. 1968;10:95–7.
73. Hoham RW. Optimum Temperatures and Temperature Ranges for Growth of Snow Algae. Arct Alp Res. 1975;7:13–24.
74. Stewart A, Rioux D, Boyer F, Gielly L, Pompanon F, Saillard A, et al. Altitudinal Zonation of Green Algae Biodiversity in the French Alps. Front Plant Sci. 2021;12:679428.
75. Hoham RW, Remias D. Snow and Glacial Algae: A Review1. J Phycol. 2020;56:264–82.
76. Takeuchi N. The altitudinal distribution of snow algae on an Alaska glacier (Gulkana Glacier in the Alaska Range). Hydrol Process. 2001;15:3447–59.
77. Li X, Ding Y, Han T, Kang S, Yu Z, Jing Z. Seasonal controls of meltwater runoff chemistry and chemical weathering at Urumqi Glacier No.1 in central Asia. Hydrol Process. 2019;33:3258–81.
78. Segawa T, Takeuchi N. Cyanobacterial communities on Qiyi glacier, Qilian Shan, China. Ann Glaciol. 2010;51:135–44.

ACKNOWLEDGEMENTS

This study was supported by Grants-in-Aid for Scientific Research (no. 17H01854, 18H04139, 19H01143, 21H03588) from the Japan Society for the Promotion of Science, by a grant from the Institute for Fermentation, Osaka (G-2020-2-133 and Y-2019-008), and by the Arctic Challenge for Sustainability II (ArCS II), Program Grant Number JPMXD1420318865. The ice core was drilled as part of the Ili Projects (R-03, 2006–2012) funded by the Research Institute for Humanity and Nature, Japan. We thank the Spanish Polar Program for the support provided by the Juan Carlos I Station on Livingston Island. Project CTM2014-56473-R from the Spanish Polar Program contributed to the funding of the field work on Livingston Island. Snow and ice samples from Siberia were collected in field campaigns of the Suntar-Khayata Observational Research Project with support from the Melnikov Permafrost Institute of the Russian Academy of Sciences (2012–2014). We acknowledge the National Institute of Polar Research through General Collaboration Project no. 2-31. Some of the computations were performed using a supercomputer of the National Institute of Genetics, Japan.

AUTHOR CONTRIBUTIONS

TS, RM, HM, TY, and NT conceived of the project and designed the research. TS conceived the hypothesis. TS, NT, VA, FN, ZH, and KF collected snow and ice samples and managed the glacier expedition for ice-core drilling. AA and TS supplied the sequences. RM, HM, TY. TS and SM analyzed the sequencing data. TS, SM, RM, TY, HM, FN, and NT wrote the paper. All authors gave final approval for publication.

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/s41396-023-01359-3.

Correspondence and requests for materials should be addressed to Takahiro Segawa or Takahiro Yonezawa.

Supplemental Information

Evolution of snow algae, from cosmopolitans to endemics, revealed by DNA analysis of ancient ice

Takahiro Segawa, Takahiro Yonezawa, Ryo Matsuzaki, Hiroshi Mori, Ayumi Akiyoshi, Francisco Navarro, Koji Fujita, Vladimir B. Aizen, Zhongqin Li, Shuhei Mano, Nozomu Takeuchi

This PDF file includes:

Supplementary Results

Demographic analysis

Demographic parameters (t, N_0, and N_1) are summarized in Table S14. The tMRCAs (time of the most recent common ancestors) were also calculated as $2 \times N_1$ generations ago under the assumption of the haploidy populations (if t years > $2 \times N_1$ generations). If $2 \times N_1$ generations > t years, as the coalescent process under the constant population size no longer holds, we assumed tMRCA as $t + 2 \times N_0$ generations ago.

Taking into account the 95% confidence interval of the mutation rate ($1.09–3.74 \times 10^{-10}$/site/year; [S1]) as well as the range of the generation interval (11–36 days; [S2, S3], the minimum–maximum range of the mutation rate was assumed to be 4.21×10^{-7} to 4.73×10^{-6}/sequence/year. The ranges of the demographic parameters (t, N_0, N_1, and tMRCA) were also estimated within this framework.

As mentioned in the main text, endemics were derived from the cosmopolitans $1.9 \times 10^5 – 9.2 \times 10^6$ years ago in Groups A and C, corresponding to the Late Miocene to the Middle Pleistocene. A global cooling event began in the Late Miocene and has continued to the present day, and it characterizes this period through the Cenozoic Era in terms of the inception of ice-sheets in both hemispheres [S4]. It is possible that environmental diversity in the cryosphere increased during this period, and this may have increased the prevalence of endemism as a result of local adaptations.

We also estimated the demographic histories of the cosmopolitans and endemics for each of Groups A, B, and C based on the Bayesian Skyline Plot [S5] with BEAST ver. 1.10.4 [S6]. The drift of the molecular evolutionary rate was modeled by the uncorrelated log-normal relaxed (UCLD) clock. The UCLD mean was assumed to be $2.08 \times 10^{-10} \times (365/24) = 3.16 \times 10^{-9}$/site/year. The standard deviation of the UCLD mean was calculated from the combinations of the minimum–maximum range of the mutation rate per generation [S1] and the generation

interval [S2, S3] and was assumed to be 4.94×10^{-9}/site/year. A normal distribution was assumed and truncated to 1.24×10^{-8}/site/year for the upper boundary and 1.11×10^{-9}/site/year for the lower boundary. The HKY + I + Γ model was used for the nucleotide substitution model. MCMC was conducted under 200 million generations, and trees were sampled every 10,000 generations.

The tMRCAs estimated by the Bayesian Skyline Plot using BEAST are summarized in Table S14. tMRCAs estimated by BEAST and by the ML method were fundamentally consistent. Although the Bayesian Skyline Plot can flexibly estimate the demographic histories (e.g., expanding, contracting, stable), they can be traced back only to the tMRCA. In contrast, our ML method, which assumes an expanding or stable population size model, can capture the demographic signal before the tMRCA. Accordingly, the demographic histories estimated by the ML and BEAST were not directly comparable for Group A cosmopolitans and endemics and Group C cosmopolitans because of $t >$ tMRCA. However, as $t <$ tMRCA in Group C endemics, those results were comparable (Table S14). The demographic histories of the cosmopolitans and endemics for each of Group A, B, and C as estimated by the Bayesian Skyline Plots are shown in Fig. S16. The timing of the population size expansion in Group C was ~200,000 years ago for the ML method (51,000–570,000 years ago, considering the range of mutation rates and generation intervals). In contrast, BEAST also showed a significant population size increase ~300,000 years ago, and thus the two results are consistent.

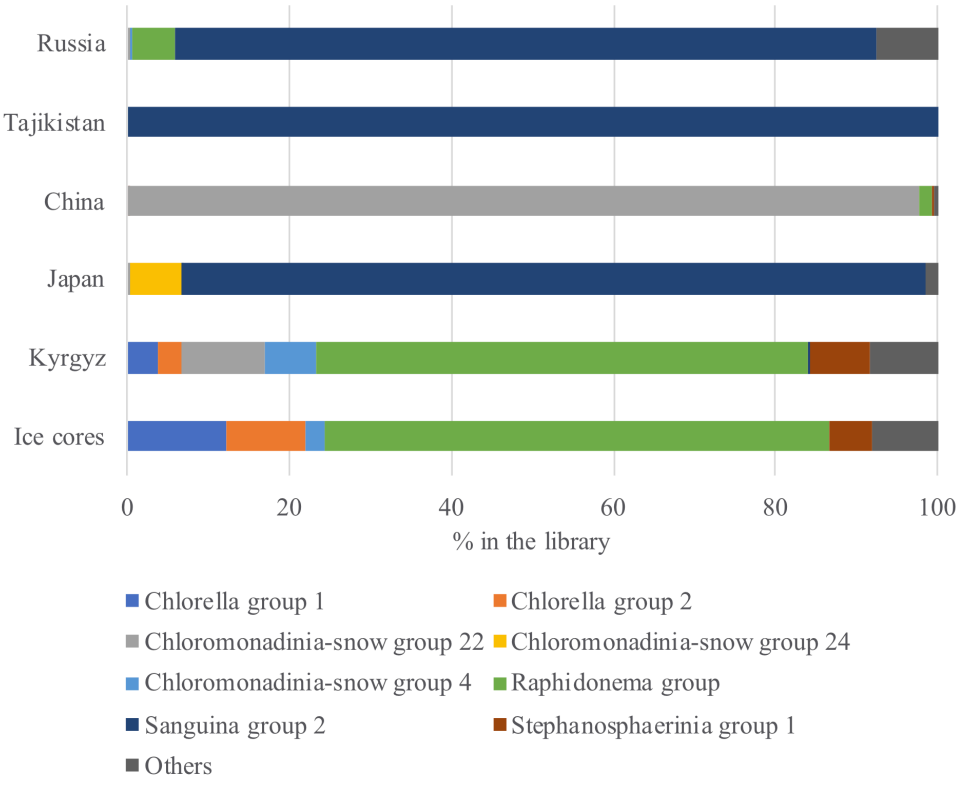

Figure S1: Algal taxonomic composition of the reads in snow and ice samples based on ITS2 sequences. The bar chart presents the average community composition for each region based on the five major ITS2 groups and low-abundance groups (grouped as Others).

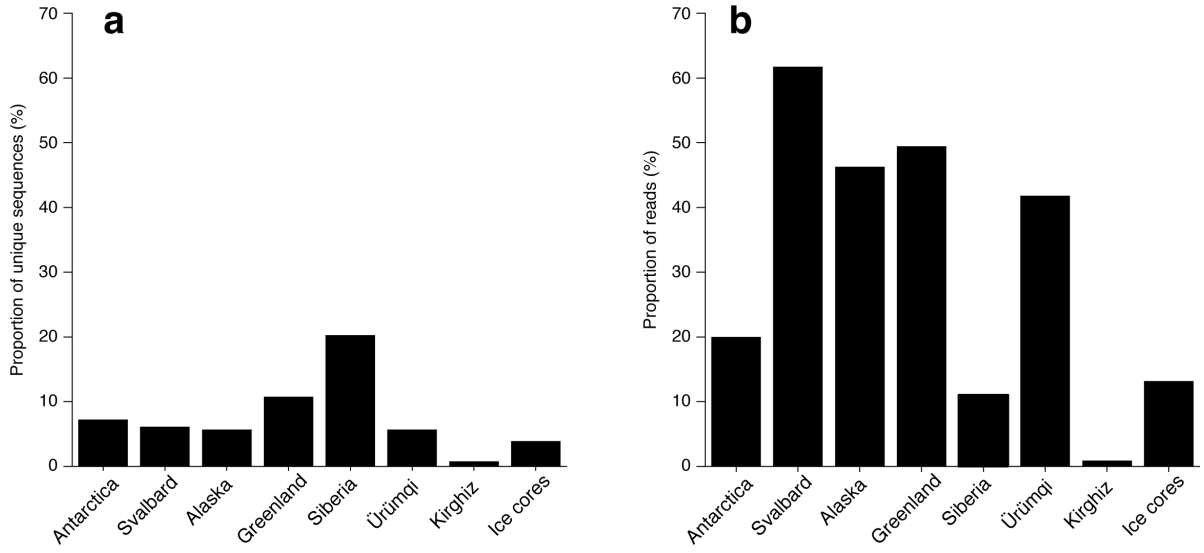

Figure S2: Distribution of the both poles and mid-latitude cosmopolitan in the *Raphidonema* group obtained from each region based on unique ITS2 sequences from high-throughput sequencing. Unique sequences and sequencing read numbers are shown. **(a)** Unique sequences. **(b)** Number of sequencing reads of the unique sequences.

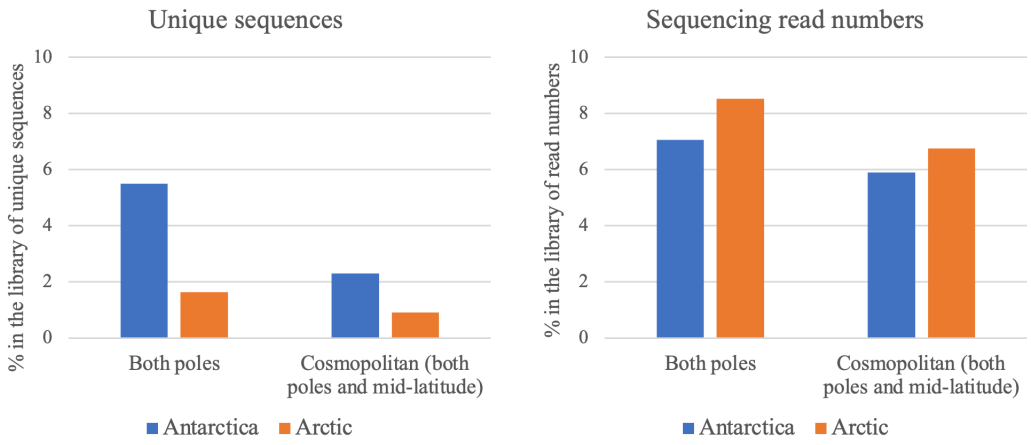

Figure S3: Distribution of the both poles (detected from Antarctica and the Arctic, namely Svalbard, Greenland, and Alaska) and cosmopolitan (both poles and mid-latitude) snow algae in the *Raphidonema* group obtained from each region based on unique ITS2 sequences. Unique sequences and sequencing read numbers are shown. **Left** Unique sequences. **Right** Number of sequencing reads. 56% unique sequences and 79.9% sequencing reads of the both poles distribution were observed in the cosmopolitan distribution (both poles and mid-latitudes).

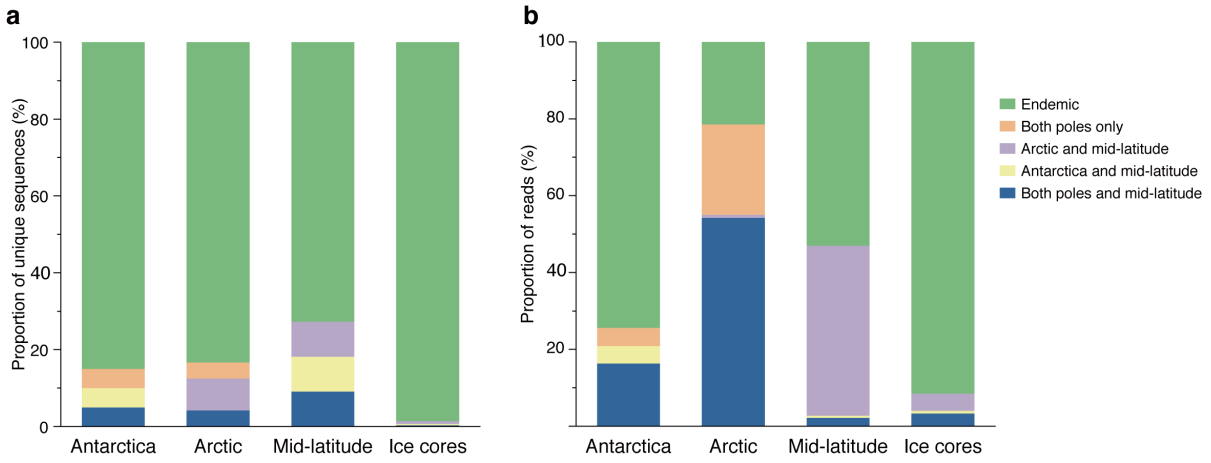

Figure S4: Distribution types of the *Raphidonema* group obtained from each region and the ice core based on ITS2 unique sequences by DADA2 analysis. **(a)** Unique sequences from surface snow and ice-core samples. **(b)** Number of sequencing reads from surface snow and ice core samples.

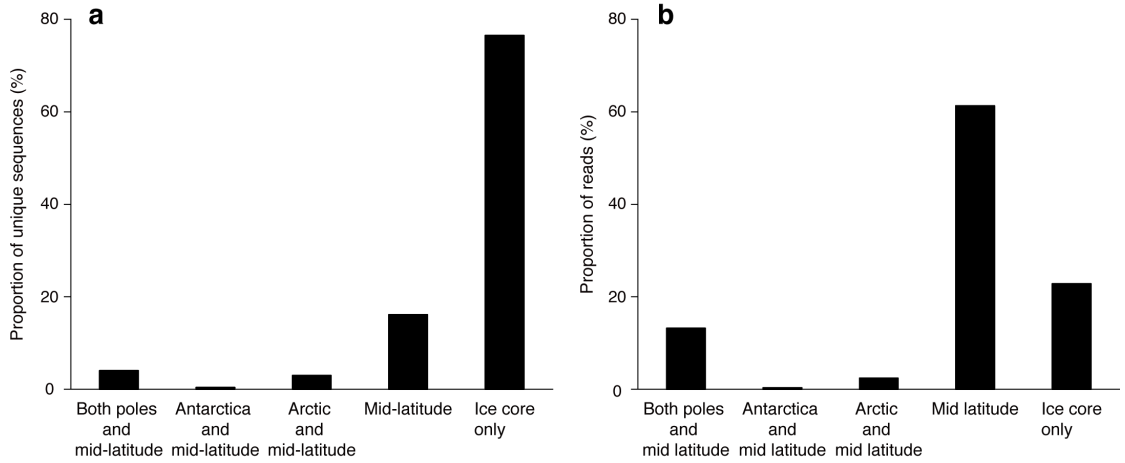

Figure S5: Average distribution categories in the *Raphidonema* group obtained from ice-core samples based on unique ITS2 sequences from high-throughput sequencing. Unique sequences and sequencing read numbers are shown. **(a)** Unique sequences. **(b)** Number of sequencing reads.

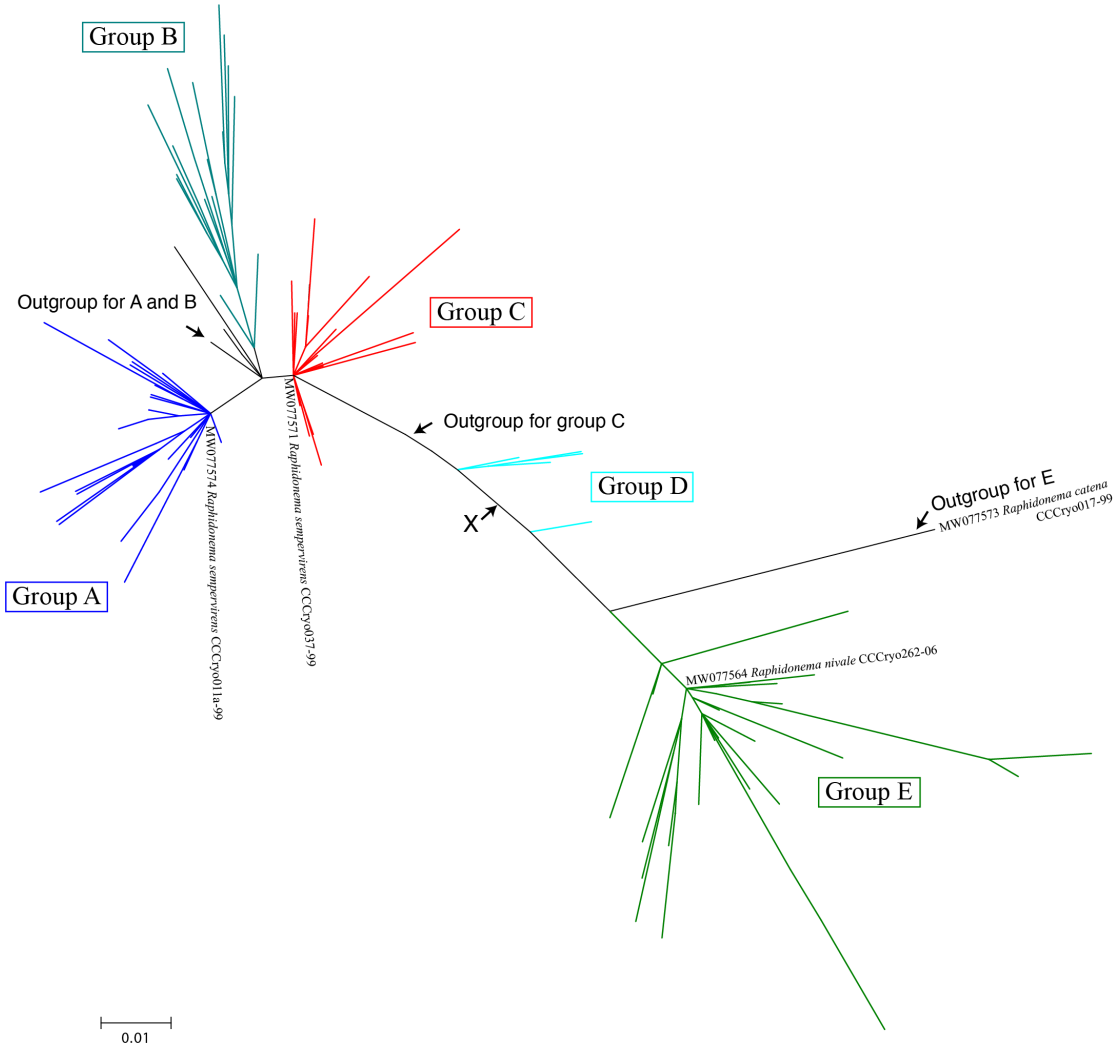

Figure S6: Phylogenetic relationship of 170 operational taxonomic units (OTUs) (98% OTU) in the *Raphidonema* group based on ITS2 sequences. Alignment among the OTUs was carried out based on ITS2 secondary structures. A maximum likelihood tree was constructed with 1000 bootstrap replications using IQ-TREE version 1.6.12 with the TVMe+G4 model. The groups are labeled and distinguished based on color. Detailed information about the sequences used for species identification is provided [S7]. The root position of this tree (indicated by'×') was assigned using the MinVar method [S8]. The MAD method [S9] showed essentially the same result. By locating this root position, the ancestral (closer to the root) and the derived (farther from the root) nodes can then be defined. The sequences derived from the closest ancestral node from each MRCA (most recent common ancestor) of Groups A–E was used for the outgroups for Groups A–E (Fig. 3, Fig. S12, S13).

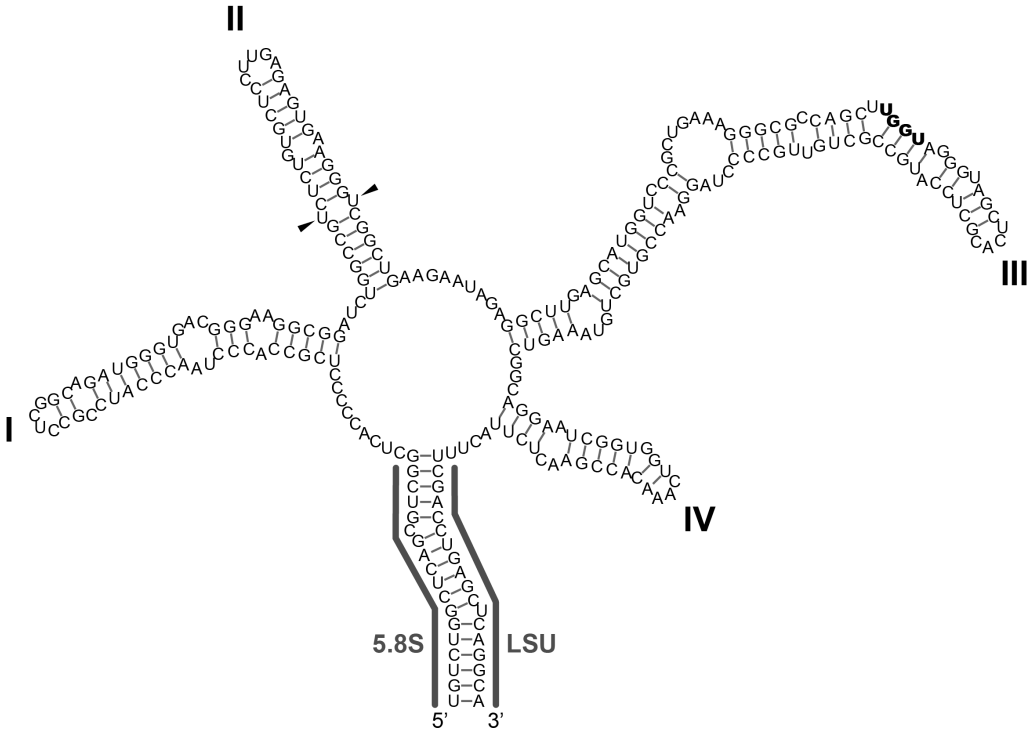

Figure S7: Predicted secondary structure of the nuclear rDNA ITS2 transcript of *Raphidonema* Group A (Unique sequence no. Uniq1360). The 3' end of the 5.8S ribosomal RNA (rRNA) and the 5' end of the large subunit of rRNA (LSU rRNA) are indicated by dark lines. Note the U-U mismatch in helix II (arrowheads) and the YGGY motif (UGGU) on the 5' side of helix III (bold type).

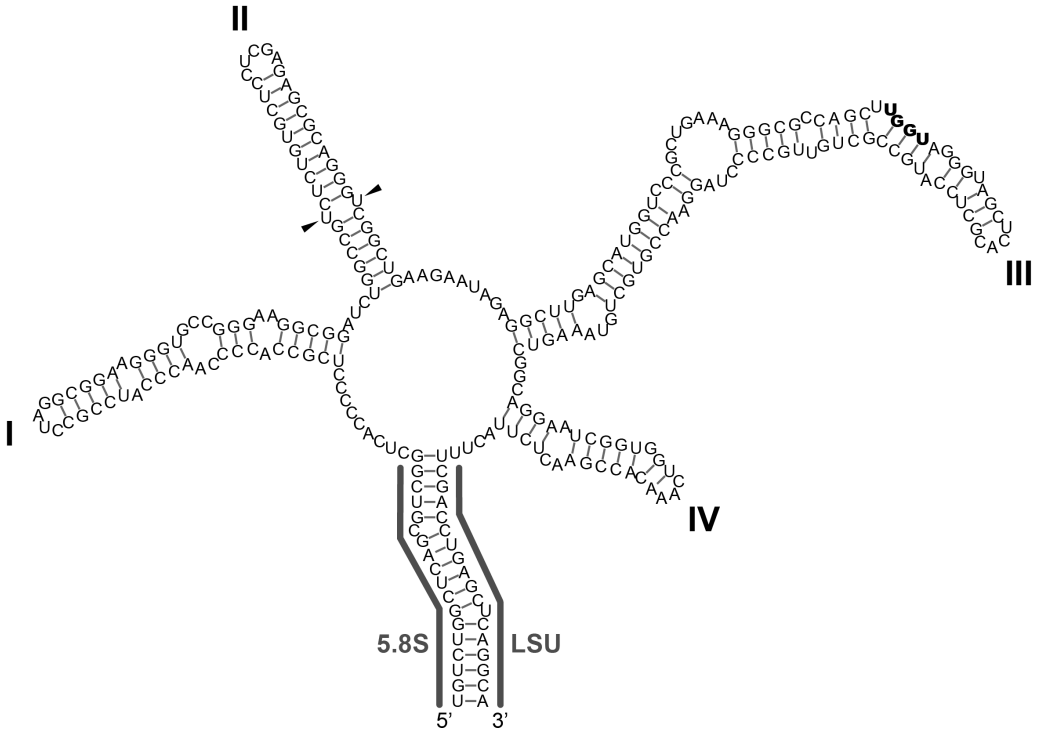

Figure S8: Predicted secondary structure of the nuclear rDNA ITS2 transcript of *Raphidonema* Group B (Unique sequence no. Uniq2588). The 3' end of the 5.8S ribosomal RNA (rRNA) and the 5' end of the large subunit of rRNA (LSU rRNA) are indicated by dark lines. Note the U-U mismatch in helix II (arrowheads) and the YGGY motif (UGGU) on the 5' side of helix III (bold type).

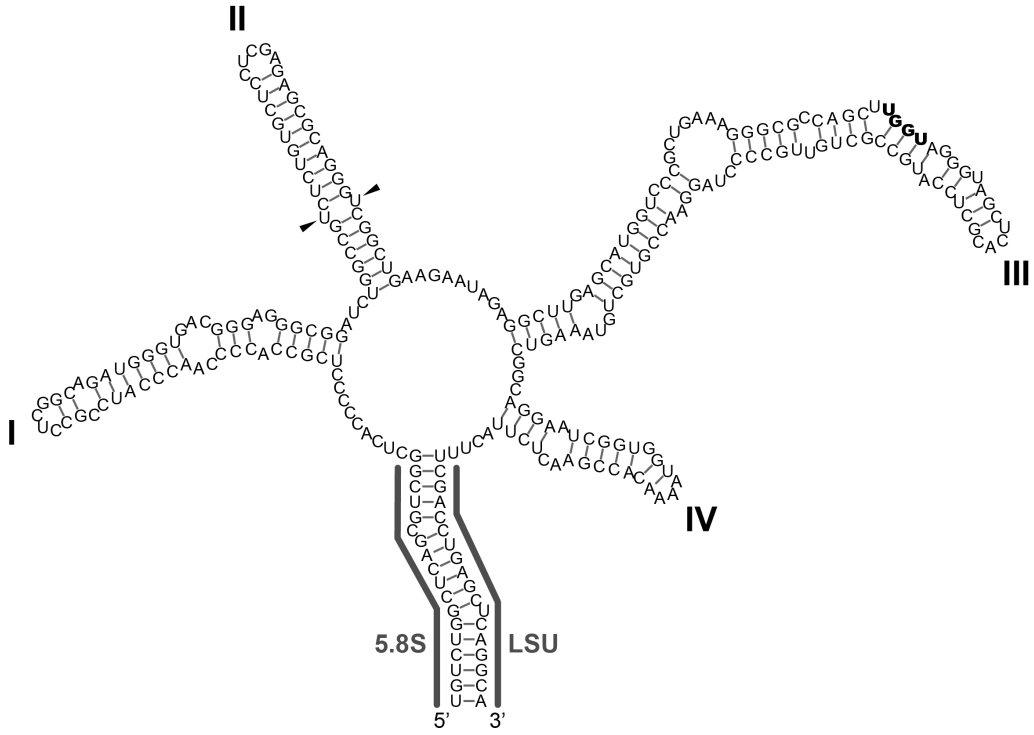

Figure S9: Predicted secondary structure of the nuclear rDNA ITS2 transcript of *Raphidonema* Group C (Unique sequence no. Uniq1782). The 3' end of the 5.8S ribosomal RNA (rRNA) and the 5' end of the large subunit of rRNA (LSU rRNA) are indicated by dark lines. Note the U-U mismatch in helix II (arrowheads) and the YGGY motif (UGGU) on the 5' side of helix III (bold type).

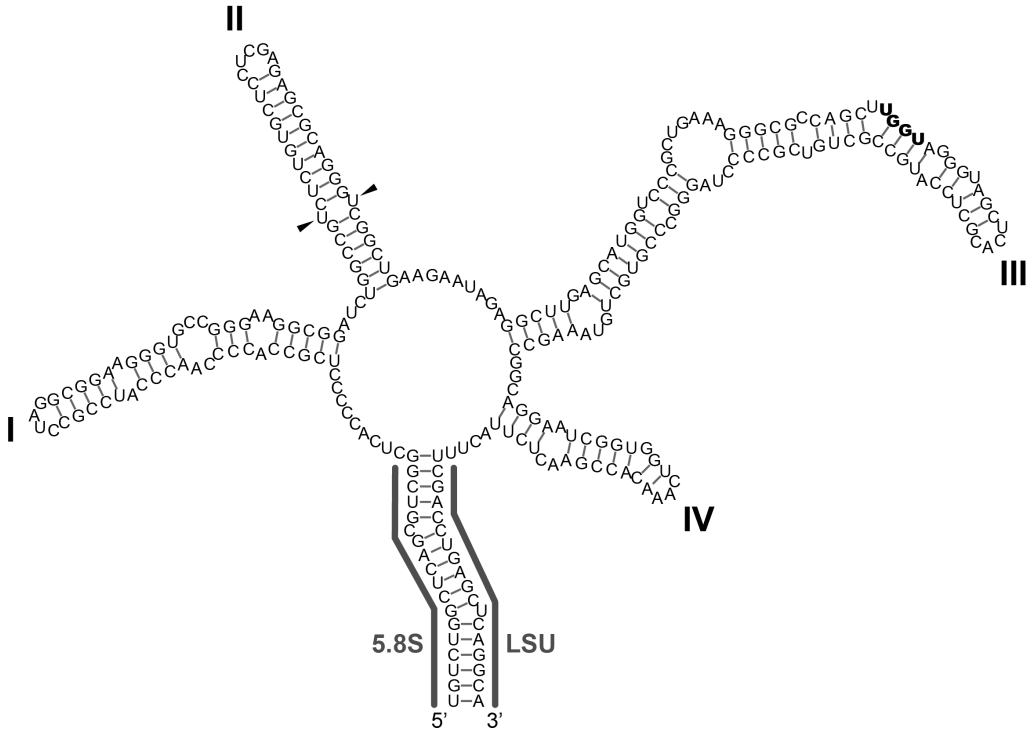

Figure S10: Predicted secondary structure of the nuclear rDNA ITS2 transcript of *Raphidonema* Group D (Unique sequence no. Uniq1367). The 3' end of the 5.8S ribosomal RNA (rRNA) and the 5' end of the large subunit of rRNA (LSU rRNA) are indicated by dark lines. Note the U-U mismatch in helix II (arrowheads) and the YGGY motif (UGGU) on the 5' side of helix III (bold type).

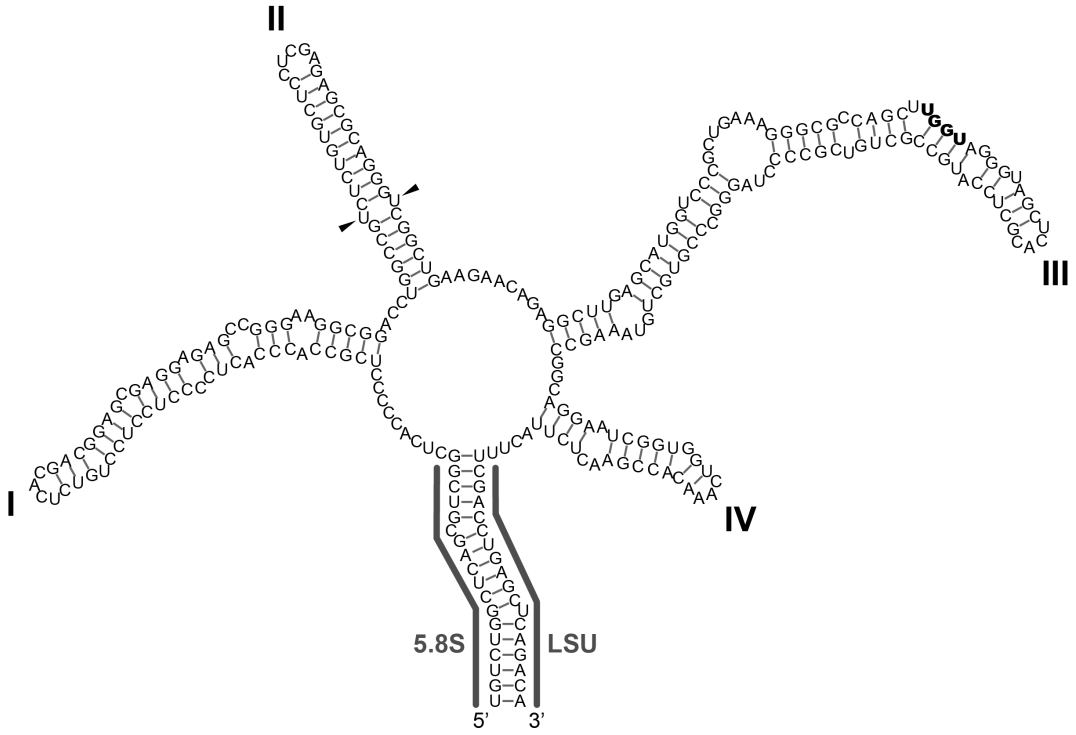

Figure S11: Predicted secondary structure of the nuclear rDNA ITS2 transcript of *Raphidonema* Group E (Unique sequence no. Uniq2918). The 3' end of the 5.8S ribosomal RNA (rRNA) and the 5' end of the large subunit of rRNA (LSU rRNA) are indicated by dark lines. Note the U-U mismatch in helix II (arrowheads) and the YGGY motif (UGGU) on the 5' side of helix III (bold type).

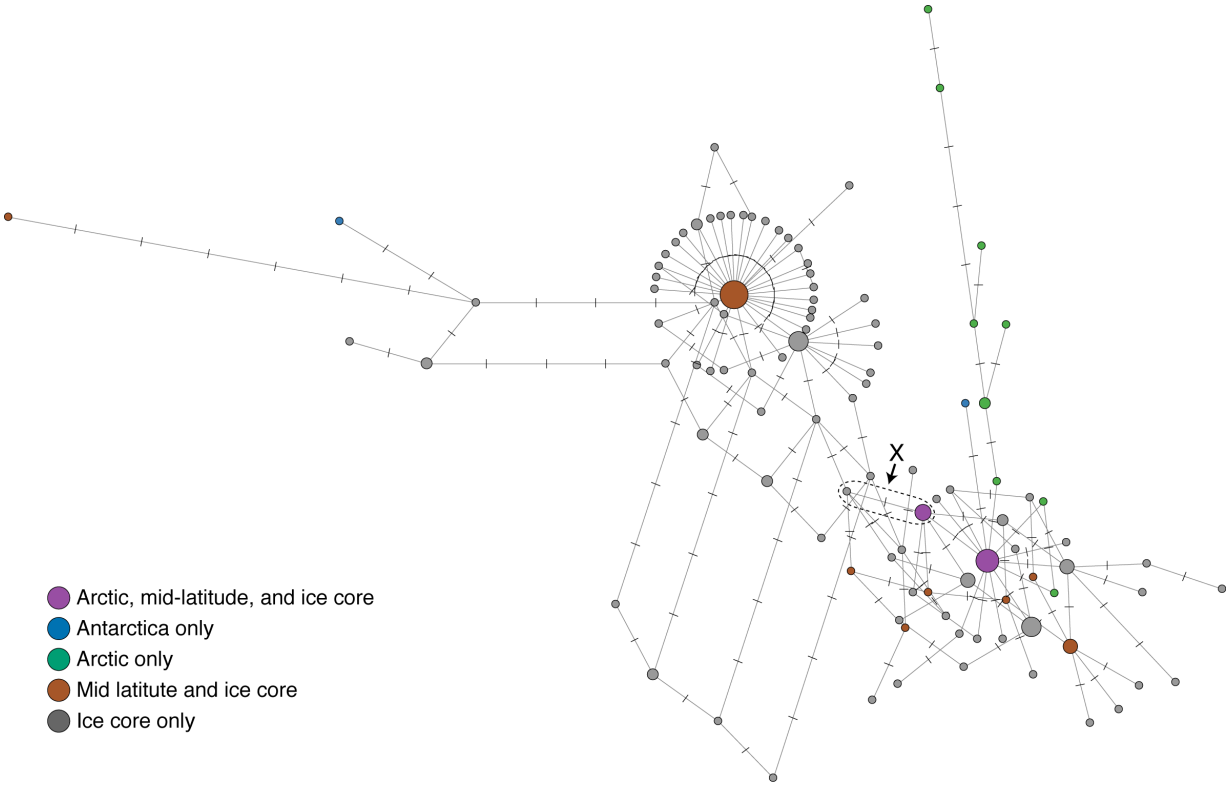

Figure S12: Phylotype networks for ITS2 sequences of *Raphidonema* Group D in this study. The median-joining method was used. Circles indicate phylotypes; the size of each circle is proportional to the number of unique sequences. Each notch on the edges represents a mutation. Phylotypes are colored according to geographic region. The area enclosed in a dashed-lined oval shape indicates the approximate position of the root in Figure S6.

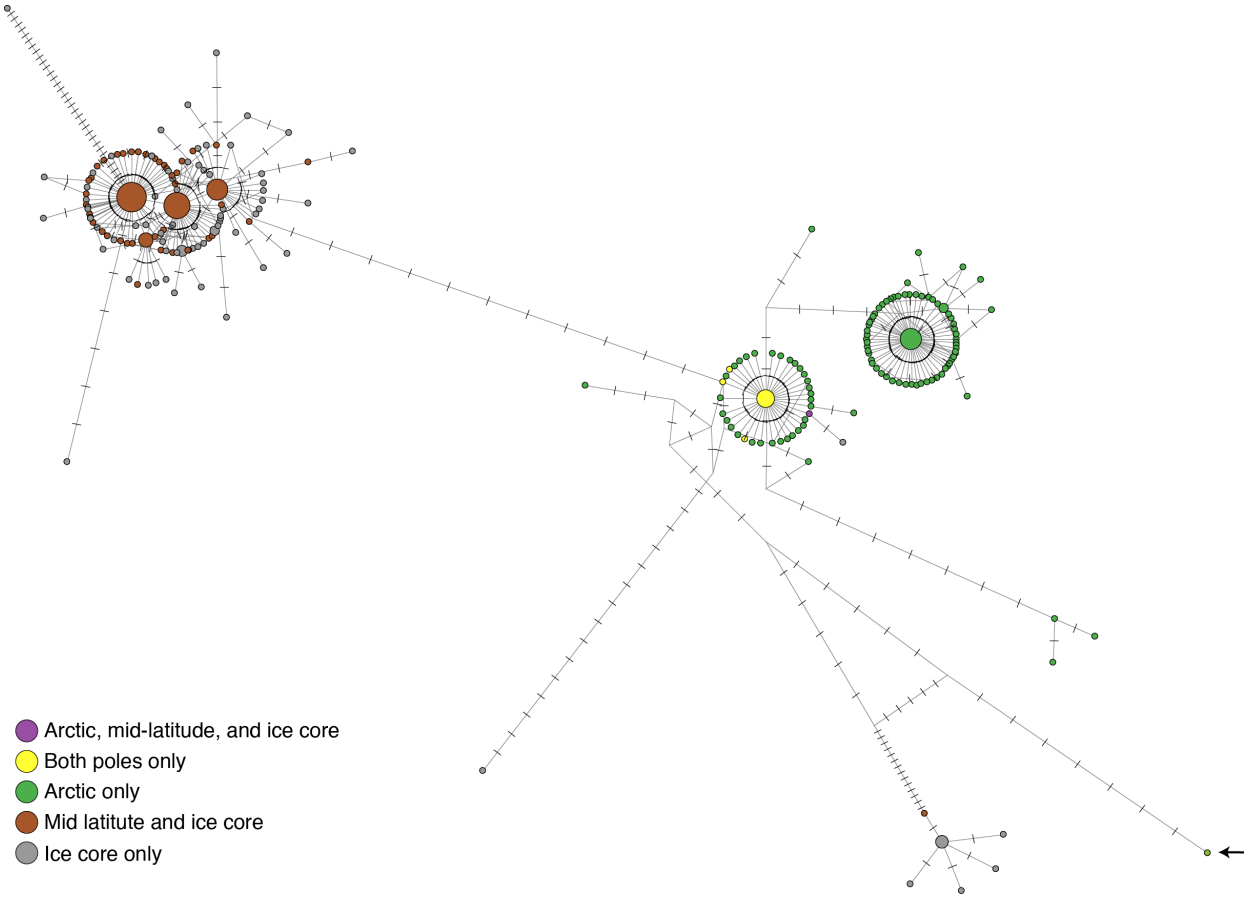

Figure S13: Phylotype networks for ITS2 sequences of *Raphidonema* Group E in this study. The median-joining method was used. Circles indicate phylotypes; the size of each circle is proportional to the number of unique sequences. Each notch on the edges represents a mutation. Phylotypes are colored according to geographic region. The arrow represents the phylotype in the outgroup.

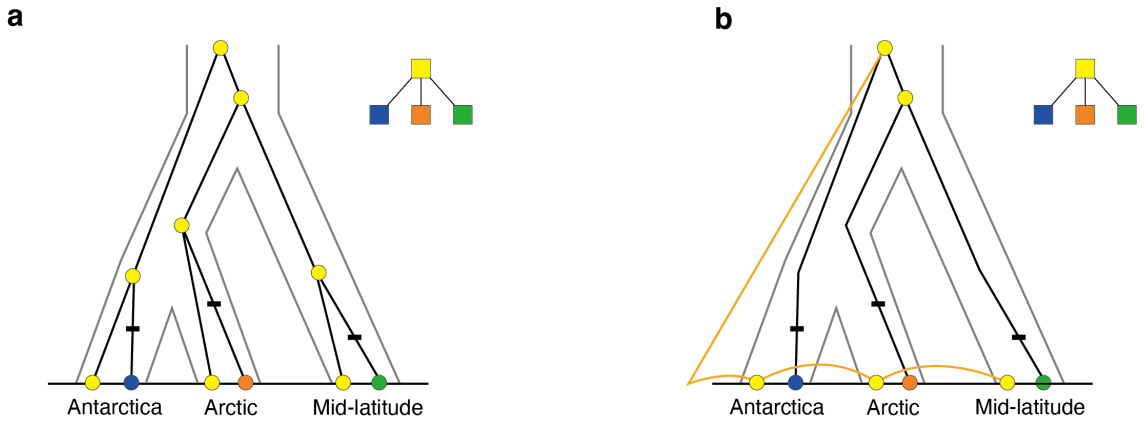

Figure S14: Conceptual diagram of "apparent cosmopolitan" and "true cosmopolitan". When the cosmopolitan phylotype is the ancestral type, any of the following two possibilities can occur: "apparent cosmopolitan" and "true cosmopolitan". **(a)** Apparent cosmopolitan. Genetically distantly related individuals share the same "ancestral" phylotype, detected in several different localities. **(b)** True cosmopolitan. Genetically closely related individuals move between regions within a short period (the geographic origin of true cosmopolitans is unknown). Black bold lines on the branches of the phylogenetic tree indicate mutations. Color points represent distribution type: cosmopolitans–yellow circles, endemics–blue, orange and green circles. If the ITS2 sequences lack sufficient phylogenetic information, the phylogenetic network of **(a)** and **(b)** can be the same in terms of haplotype divergence patterns (**a**, **b** upper right). Therefore, these two possibilities cannot be distinguished merely from the phylogenetic network.

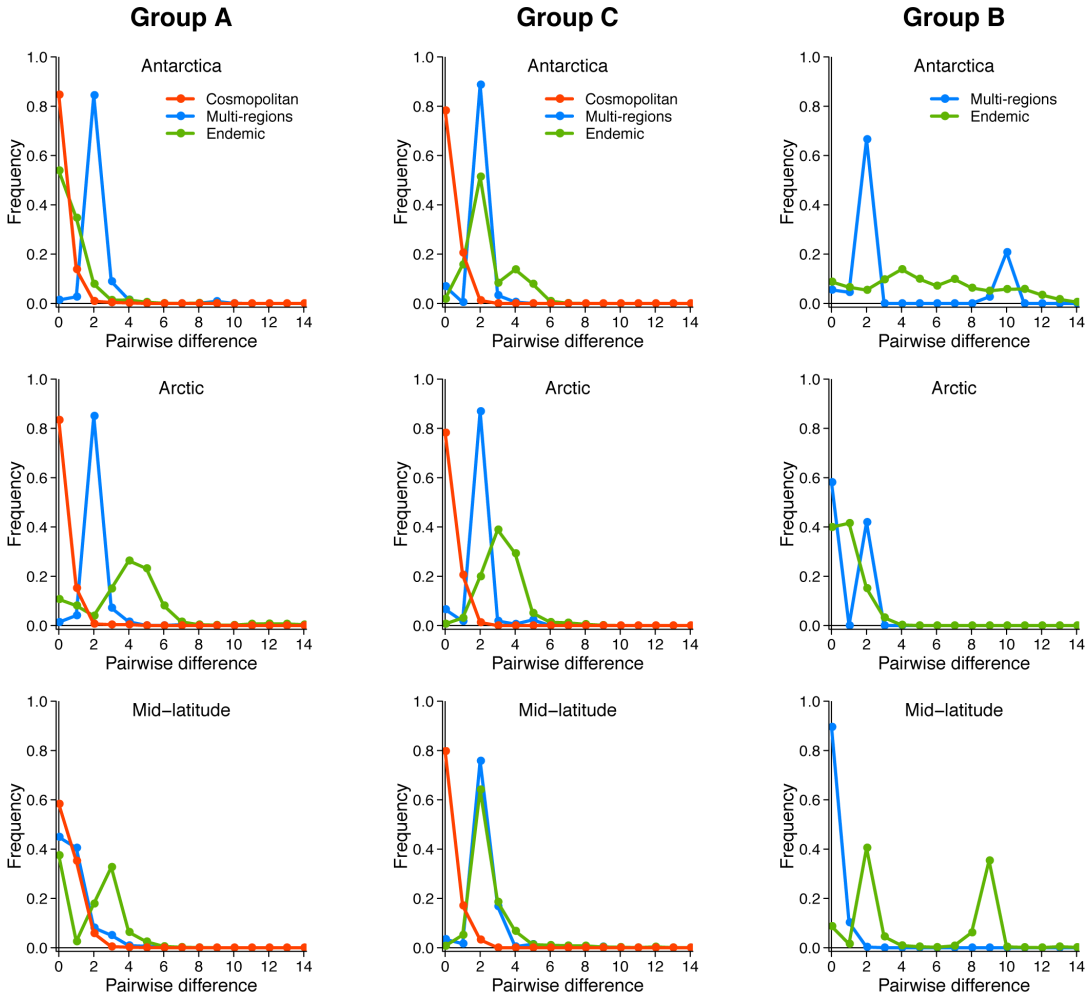

Figure S15: Mismatch distribution based on the number of pairwise differences in each *Raphidonema* Group (A–C) from each region. The lines represent the distribution of the observed number of pairwise differences of each *Raphidonema* group. Regions (Antarctica, Arctic, and mid-latitude) and distribution types (cosmopolitan, multi-regions, endemic) are distinguished. Calculations were performed for all distribution types of *Raphidonema* Groups A and C, for which various cosmopolitan phylotypes were detected and had a star-like phylogeny. On the other hand, calculations for only multi-regions and endemic phylotypes were performed for *Raphidonema* Group B, because no variation was found in cosmopolitan phylotypes.

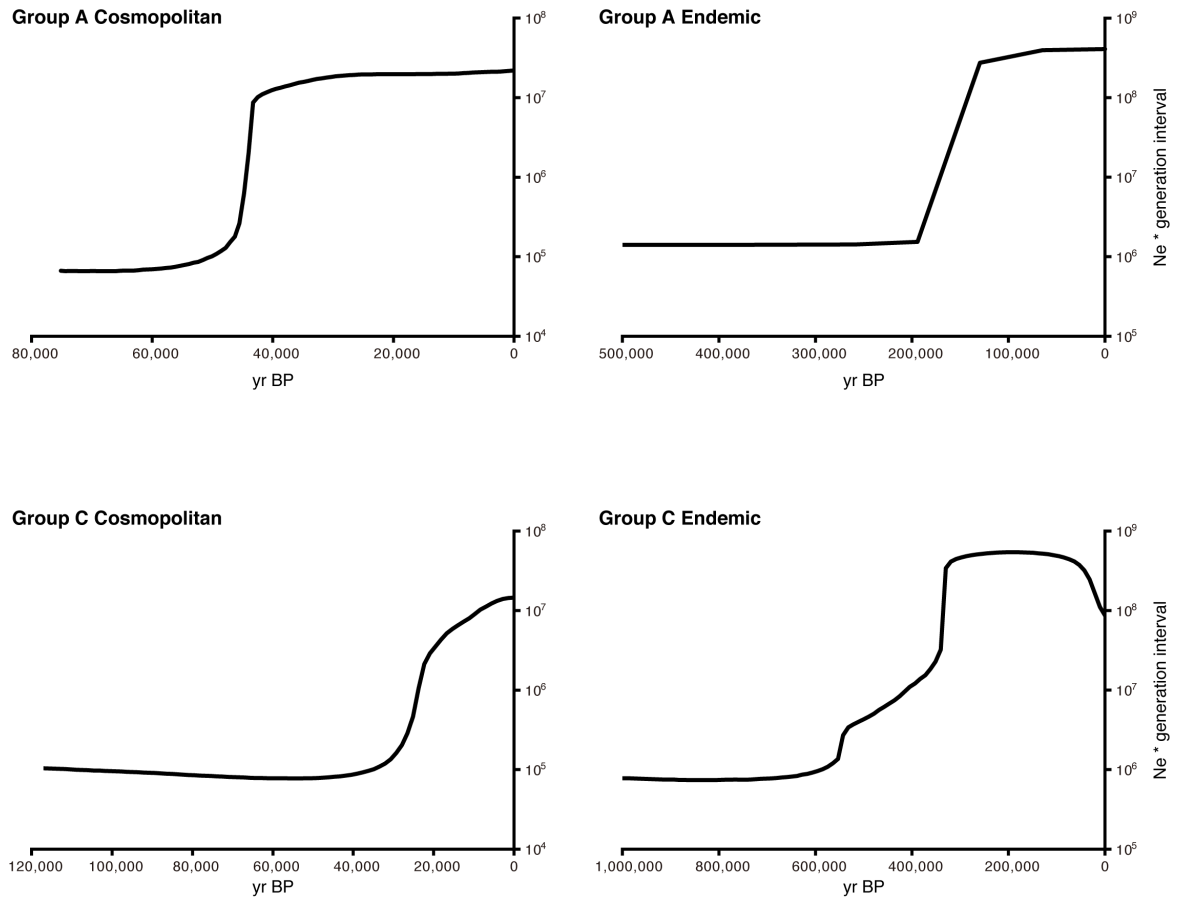

Figure S16: Bayesian skyline plots that show the Ne*t, where Ne stands for the effective population size and t stands for the generation intervals over time in each *Raphidonema* Group (A and C). Lines represent the median parameter estimates. The x axis indicates years before present; the y axis indicates the effective population size multiplied by the generation interval (Ne*t).

Table S1: List of ice core samples usedno in this study.

Ice core no.	Ice core no.	Depth (m)	Age	Successful amplification by PCR of ITS2 region
No.1	GGC2-27	3.85	AD 2001	no
No.2	GGC2-156	11.08	AD 1,984	no
No.3	GGC2-281	17.30	AD 1,967	no
No.4	GGC2-300	18.25	AD 1,965	no
No.5	GGC2-334	19.99	AD 1,963	yes
No.6	GGC2-353	20.99	AD 1,962	no
No.7	GGC2-449	25.80	AD 1,950	no
No.8	GGC2-491	27.86	AD 1,943	no
No.9	GGC2-528	29.69	AD 1,934	yes
No.10	GGC2-629	34.89	AD 1,918	yes
No.11	GGC2-992	54.05	AD 1,833	yes
No.12	GGC2-1193	63.46	AD 1,781	yes
No.13	GGC1-2202	68.60	–	yes
No.14	GGC1-2436	75.97	–	no
No.15	GGC1-2612	80.61	–	yes
No.16	GGC1-2672	81.83	7,969 – 8,017 cal yr bp	yes
No.17	GGC1-2684	82.07	–	yes
No.18	GGC1-2715	82.69	–	yes
No.19	GGC1-2854	85.52	8,016 – 8,137 cal yr bp	yes

Ages of samples No.1–12 were determined by the annual layers of pollen grains [S10]. Ages of samples No.16 and No.19 were determined by radio-carbonisotope [S11]. Ages of samples No.13–15, 17, and 18 were not determined, but appears to be close to the age of No. 16.

Table S2: List of modern snow samples used in this study.

Region	Sample location	Site	Lat	Long	Altitude (m)	Year	Site condition	Reference
Greenland	Greenland Ice sheet, Tugto Glacier	IF13	77.89473	-68.77335	1001	2012.July	red snow on the glacier	[S12]
Greenland	Greenland Ice sheet	IH2	69.56110	-49.34715	1107	2011.August	red snow on the glacier	[S12]
Svalbard	Foxfonna glacier	FX S3	78.13747	16.15794	642	2011.August	red snow on the glacier	[S12]
Svalbard	Foxfonna glacier	FX S4	78.12687	16.20279	755	2011.August	red snow on the glacier	[S12]
Svalbard	Longyearbreen glacier	LY S3	78.17579	15.49680	417	2011.August	red snow on the glacier	[S12]
Svalbard	Longyearbreen glacier	LY S4	78.16565	15.46344	580	2011.August	red snow on the glacier	[S12]
Svalbard	Austre Brøggerbreen	SP	78.89904	11.82895	130	2008.July	red snow on the glacier	[S12]
Alaska	Juneau ice field	TK1	58.64446	-134.21068	1040	2001.July	red snow on the glacier	[S12]
Alaska	Gulkana glacier	GU S5	63.28771	-145.40551	1680	2001.August	red snow on the glacier	[S12]
Alaska	Gulkana glacier	GU S4	63.28068	-145.41242	1585	2005.June	red snow on the glacier	[S12]
Alaska	Harding ice field	HD	60.15278	-149.77992	1150	2014.August	red snow on the glacier	[S12]
Antarctica	Riiser-Larsen	Riiser-Larsen	-66.78765	50.56543	500	2008.February	non-glacier based seasonal snow	[S12]
Antarctica	Yukidori Valley	Yukidori Valley	-69.23935	39.76112	50	2008.January	non-glacier based seasonal snow	[S12]
Antarctica	Johnsons Glacier, Livingston Island	CH1	-62.65595	-60.36640	5	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Johnsons Glacier, Livingston Island	CH5	-62.65588	-60.36582	5	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Johnsons Glacier, Livingston Island	CH6	-62.65563	-60.36594	5	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Johnsons Glacier, Livingston Island	CH7	-62.65604	-60.36558	5	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Glaciar Rocoso, Livingston Island	RC4	-62.71219	-60.40879	35	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Glaciar Rocoso, Livingston Island	RC5	-62.71250	-60.40926	30	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Glaciar Rocoso, Livingston Island	RC7	-62.71508	-60.41130	10	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Hurd Glacier-Sally Rocks lobe, Livingston Island	SA.1R	-62.70138	-60.41831	35	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Hurd Glacier-Sally Rocks lobe, Livingston Island	SA.2R	-62.70153	-60.41784	12	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Hurd Glacier-Sally Rocks lobe, Livingston Island	SA.3R	-62.70243	-60.41806	0	2015.January	non-glacier based seasonal snow	[S12]
Antarctica	Hurd Glacier-Sally Rocks lobe, Livingston Island	SA.4R	-62.70299	-60.41846	0	2015.January	non-glacier based seasonal snow	[S12]
Mid-latitude	No. 31 Glacier, Suntar-Khayata Mountains, Russia	SHA	62.59513	140.90641	2611	2012.July	red snow on the glacier	This study
Mid-latitude	Ürümqi Glacier No. 1, Tien Shan Mountains, China	UM S5	43.10679	86.80594	4010	2014.August	red snow on the glacier	This study
Mid-latitude	Gregoriev Glacier, Tien Shan Mountains, Kyrgyz	Kir S1	41.94847	77.91836	4250	2006.August	surface ice	This study
Mid-latitude	Gregoriev Glacier, Tien Shan Mountains, Kyrgyz	Kir S2	41.95168	77.91848	4297	2006.August	surface ice	This study
Mid-latitude	Gregoriev Glacier, Tien Shan Mountains, Kyrgyz	Kir S3	41.95818	77.91647	4367	2006.August	surface ice	This study
Mid-latitude	Gregoriev Glacier, Tien Shan Mountains, Kyrgyz	Kir S5	41.96773	77.91317	4469	2006.August	snow on the glacier	This study
Mid-latitude	Fedchenko Glacier, Pamir, Tajikistan	Fed	38.700	72.333	4040	2009.August	red snow on the glacier	This study
Mid-latitude	Tateyama Mountains, Japan	TA R2	36.58838	137.60390	2450	2011.June	non-glacier based seasonal snow	This study
Mid-latitude	Tateyama Mountains, Japan	TA Raicho	36.58838	137.60390	2450	2012.June	non-glacier based seasonal snow	This study
Mid-latitude	Tateyama Mountains, Japan	TA RA1	36.58838	137.60390	2450	2018.June	non-glacier based seasonal snow	This study

Table S3: Number of reads for each taxonomy group in the ice core and various regions.

		Ice cores	Japan	China	Tajikistan	Kyrgyz	Russia
Chlorophyceae	*Sanguina* group 1 [a]	0	0	0	0	0	2
	Sanguina group 2 [b]	0	70,908	0	211,858	94	26,287
	Chloromonadinia-snow group 1	0	63	0	0	0	0
	Chloromonadinia-snow group 2	0	0	0	0	0	13
	Chloromonadinia-snow group 3	15,463	0	14	0	88	0
	Chloromonadinia-snow group 4	16,374	7	0	0	1,527	68
	Chloromonadinia-snow group 5	20	0	0	0	0	0
	Chloromonadinia-snow group 6	0	0	0	0	0	366
	Chloromonadinia-snow group 7	0	0	0	0	0	237
	Chloromonadinia-snow group 8	4	0	0	0	0	0
	Chloromonadinia-snow group 9	0	539	0	0	0	0
	Chloromonadinia-snow group 10	7,000	0	33	0	289	389
	Chloromonadinia-snow group 11	8,102	0	38	0	0	0
	Chloromonadinia-snow group 12	0	0	0	0	0	2
	Chloromonadinia-snow group 13	0	3	0	0	0	5
	Chloromonadinia-snow group 14	3	0	0	0	0	0
	Chloromonadinia-snow group 15	116	0	0	0	13	0
	Chloromonadinia-snow group 16	486	0	0	0	0	0
	Chloromonadinia-snow group 17	468	0	0	0	0	0
	Chloromonadinia-snow group 18	0	0	0	0	0	31
	Chloromonadinia-snow group 19	0	0	0	0	0	265
	Chloromonadinia-snow group 20	0	19	0	0	12	0
	Chloromonadinia-snow group 21	0	0	0	0	0	9
	Chloromonadinia-snow group 22	0	287	21,734	0	2,454	107
	Chloromonadinia-snow group 23	0	467	0	0	0	0
	Chloromonadinia-snow group 24	0	4,790	0	0	0	0
	Chloromonadinia-snow group 25	0	0	0	0	0	4
	Chloromonadinia-snow group 26	0	0	0	0	0	361
	Chloromonadinia-snow group 27	0	0	0	0	0	600
	Moewusinia group	0	0	0	0	10	0
	Monadinia group	13,635	0	11	0	27	0
	Reinhardtinia group 1	1,158	0	0	0	516	0
	Reinhardtinia group 2	2,605	0	0	0	78	0
	Reinhardtinia group 3	2,687	0	0	0	0	0
	Stephanosphaerinia group 1	36,197	0	36	0	1,748	0
	Stephanosphaerinia group 2	3,745	0	0	0	61	0
	unnamed group	40	0	0	0	943	0
Trebouxiophyceae	*Chlorella* group 1	83,659	0	2	0	885	0
	Chlorella group 2	66,711	0	41	0	744	0
	Elliptochloris group	0	0	0	0	0	2
	Raphidonema group	425,945	0	373	0	14,666	1,614
	Trebouxia group 1	0	0	0	0	0	6
	Trebouxia group 2	12	0	18	0	0	0
	Trebouxia group 3	0	0	5	0	0	0
	Neocystis group	46	0	0	0	0	0
Ulvophyceae	*Chamaetrichon* group	77	0	0	0	0	0
	Planophila group	101	0	0	0	0	0
	Total	684,654	77,083	22,305	211,858	24,155	30,368

[a] Formerly known as *Chlamydomonas*-snow group 1

[b] Formerly known as *Chlamydomonas*-snow group 2

Table S4: Ratio for each taxonomy group in the ice core and various regions.

		Ice cores	Japan	China	Tajikistan	Kyrgyz	Russia
Chlorophyceae	*Sanguina* group 1 [a]	0	0	0	0	0	0.01
	Sanguina group 2 [b]	0	92.0	0	100	0.4	86.6
	Chloromonadinia-snow group 1	0	0.1	0	0	0	0
	Chloromonadinia-snow group 2	0	0	0	0	0	0.04
	Chloromonadinia-snow group 3	2.3	0	0.1	0	0.4	0
	Chloromonadinia-snow group 4	2.4	0.01	0	0	6.3	0.2
	Chloromonadinia-snow group 5	0.003	0	0	0	0	0
	Chloromonadinia-snow group 6	0	0	0	0	0	1.2
	Chloromonadinia-snow group 7	0	0	0	0	0	0.8
	Chloromonadinia-snow group 8	0.001	0	0	0	0	0
	Chloromonadinia-snow group 9	0	0.7	0	0	0	0
	Chloromonadinia-snow group 10	1.0	0	0.1	0	1.2	1.3
	Chloromonadinia-snow group 11	1.2	0	0.2	0	0	0
	Chloromonadinia-snow group 12	0	0	0	0	0	0.01
	Chloromonadinia-snow group 13	0	0.004	0	0	0	0.02
	Chloromonadinia-snow group 14	0.0004	0	0	0	0	0
	Chloromonadinia-snow group 15	0.02	0	0	0	0.1	0
	Chloromonadinia-snow group 16	0.1	0	0	0	0	0
	Chloromonadinia-snow group 17	0.1	0	0	0	0	0
	Chloromonadinia-snow group 18	0	0	0	0	0	0.1
	Chloromonadinia-snow group 19	0	0	0	0	0	0.9
	Chloromonadinia-snow group 20	0	0.02	0	0	0.05	0
	Chloromonadinia-snow group 21	0	0	0	0	0	0.03
	Chloromonadinia-snow group 22	0	0.4	97.4	0	10.2	0.4
	Chloromonadinia-snow group 23	0	0.6	0	0	0	0
	Chloromonadinia-snow group 24	0	6.2	0	0	0	0
	Chloromonadinia-snow group 25	0	0	0	0	0	0.01
	Chloromonadinia-snow group 26	0	0	0	0	0	1.2
	Chloromonadinia-snow group 27	0	0	0	0	0	2.0
	Moewusinia group	0	0	0	0	0.04	0
	Monadinia group	2.0	0	0.05	0	0.1	0
	Reinhardtinia group 1	0.2	0	0	0	2.1	0
	Reinhardtinia group 2	0.4	0	0	0	0.3	0
	Reinhardtinia group 3	0.4	0	0	0	0	0
	Stephanosphaerinia group 1	5.3	0	0.2	0	7.2	0
	Stephanosphaerinia group 2	0.5	0	0	0	0.3	0
	unnamed group	0.01	0	0	0	3.9	0
Trebouxiophyceae	*Chlorella* group 1	12.2	0	0.01	0	3.7	0
	Chlorella group 2	9.7	0	0.2	0	3.1	0
	Elliptochloris group	0	0	0	0	0	0.01
	Raphidonema group	62.2	0	1.7	0	60.7	5.3
	Trebouxia group 1	0	0	0	0	0	0.02
	Trebouxia group 2	0.002	0	0.1	0	0	0
	Trebouxia group 3	0	0	0.02	0	0	0
	Neocystis group	0.01	0	0	0	0	0
Ulvophyceae	*Chamaetrichon* group	0.01	0	0	0	0	0
	Planophila group	0.01	0	0	0	0	0
	Total	100	100	100	100	100	100

[a] Formerly known as *Chlamydomonas*-snow group 1

[b] Formerly known as *Chlamydomonas*-snow group 2

Table S5: Number of reads and unique sequences in total and in the *Raphidonema* group.

Samping site	name	Sequencing reads in total	Sequencing reads in *Raphidonema* group	Unique sequences in total	Unique sequences in *Raphidonema* group
Suntar-Khayata glacier	SHA	30,368	1,614	1,022	74
Raicho sawa	TA R2	25,122	0	1,005	0
Raicho sawa	TA Raicho	16,208	0	1,367	0
Raicho sawa	TA RA1	35,753	0	916	0
Ürümqi Glacier No. 1	UM S5	22,305	373	740	35
Fedchenko Glacier	Fed	211,858	0	3,968	0
Gregoriev Glacier	Kir S1	2,036	803	251	134
Gregoriev Glacier	Kir S2	10,135	7,662	1,057	863
Gregoriev Glacier	Kir S3	10,560	5,717	1,202	752
Gregoriev Glacier	Kir S5	1,424	484	151	84
Ice core on Gregoriev Glacier	No. 5	195,365	176,398	4,516	3,795
Ice core on Gregoriev Glacier	No. 9	8,580	7,317	690	585
Ice core on Gregoriev Glacier	No. 10	14,985	3,489	1,264	457
Ice core on Gregoriev Glacier	No. 11	6,970	1,271	848	237
Ice core on Gregoriev Glacier	No. 12	18,576	17,034	1,398	1,281
Ice core on Gregoriev Glacier	No. 13	2,681	128	117	20
Ice core on Gregoriev Glacier	No. 15	81,373	69,332	4,281	3,601
Ice core on Gregoriev Glacier	No. 16	131,255	9,951	5,632	1,020
Ice core on Gregoriev Glacier	No. 17	59,254	7,222	4,283	1,083
Ice core on Gregoriev Glacier	No. 18	43,281	16,361	3,192	1,242
Ice core on Gregoriev Glacier	No. 19	122,334	117,442	4,101	3,841
Total		1,050,423	442,598	25,039	8209

Table S6: Number of reads, unique sequences, and 98% OTU analysis of ITS2 sequences in the *Raphidonema* group in the various regions.

	Mid-latitude	Ice cores	Antarctica	Arctic	Total
Number of reads	16,653	425,945	139,972	311,079	893,649
98% OTU number	33	95	64	82	222
Unique sequence numbers	1,489	8,072	5,721	10,350	22,389

Table S7: PERMANOVA analysis of ITS2 sequences based on unique sequences in the *Raphidonema* group.

	Antarctica	Arctic	Mid-latitude	Ice cores
Antarctica				
Arctic	0.0002*			
Mid-latitude	0.0001*	0.0004*		
Ice cores	0.0001*	0.0001*	0.0088*	

* Statistically significant after Bonferroni's correction at $P < 0.01$.

Table S8: Classification of distribution type for this study.

Type	Distribution type	Detecting region
Cosmopolitan	Both poles and mid-latitude	Antarctica – Svalbard – Greenland – Alaska – Mid-latitude
Multi-regions	Antarctica and mid-latitude	Antarctica – Mid-latitude
	Arctic and mid-latitude	Svalbard – Greenland – Alaska – Mid-latitude
	Both poles only	Antarctica – Svalbard – Greenland – Alaska
Endemic	Arctic only	Svalbard – Greenland – Alaska
	One region	Antarctica
		Greenland
		Svalbard
		Alaska
		Mid-latitude

Table S9: Numbers and ratio for each distribution type in the *Raphidonema* group based on unique sequences.

Distribution type	Unique sequences numbers				Unique sequences (%)			
	Antarctica	Arctic	Mid-latitude	Ice cores	Antarctica	Arctic	Mid-latitude	Ice cores
Both poles and mid-latitude	323	323	21	319	5.6	3.1	1.4	4.0
Antarctica and mid-latitude	34	-	2	34	0.6	-	0.1	0.4
Arctic and mid-latitude	-	246	35	242	-	2.4	2.4	3.0
Both poles only	1,568	1,568	-	-	27.4	15.1	-	-
Antarctica only	3,796	-	-	-	66.4	-	-	-
Arctic only	-	8,213	-	-	-	79.4	-	-
Mid-latitude	-	-	1,431	1,302	-	-	96.1	16.1
Ice cores only	-	-	-	6,175	-	-	-	76.5
Total	5,721	10,350	1,489	8,072	100.0	100.0	100.0	100.0

Table S10: Numbers and ratio for each distribution type in the *Raphidonema* group based on read sequences.

Distribution type	Sequencing reads				Sequencing reads (%)			
	Antarctica	Arctic	Mid-latitude	Ice cores	Antarctica	Arctic	Mid-latitude	Ice cores
Both poles and mid-latitude	27,963	169,893	466	56,147	20.0	54.6	2.8	13.2
Antarctica and mid-latitude	142	-	46	1,468	0.1	-	0.3	0.3
Arctic and mid-latitude	-	3,410	3,387	10,220	-	1.1	20.3	2.4
Both poles only	7,760	69,132	-	-	5.5	22.2	-	-
Antarctica only	104,107	-	-	-	74.4	-	-	-
Arctic only	-	68,644	-	-	-	22.1	-	-
Mid-latitude	-	-	12,754	261,060	-	-	76.6	61.3
Ice cores only	-	-	-	97,050	-	-	-	22.8
Total	139,972	311,079	16,653	425,945	100.0	100.0	100.0	100.0

We found that only limited unique sequences are distributed across all of the regions (mid-latitudes, 1.4%; Antarctica, 5.6%; Arctic, 3.1%). They account for a large proportion of the sequencing reads in the polar regions, but for a small proportion in the mid-latitudes (mid-latitudes, 2.8%; Antarctica, 20.0%; Arctic, 54.6%). These results suggested that a limited number of snow algae in the *Raphidonema* group detected from mid-latitude regions was globally dispersed across the two poles. The mid-latitude samples analyzed in this study were located in the Northern Hemisphere, which may have resulted in a larger amount of the both poles and mid-latitude phylotypes being detected in the Arctic as compared with the Antarctic.

Table S11: Number of unique sequences for each phylogenetic group of *Raphidonema* and each distribution type.

	Group A	Group B	Group C	Group D	Group E	Others
Both poles and mid-latitude	205	1	116	0	0	1
Antarctica and mid-latitude	13	18	3	0	0	0
Arctic and mid-latitude	123	0	117	2	1	3
Both poles only	1,115	2	439	0	10	2
Antarctica only	1,321	2,263	179	1	30	2
Arctic only	3,279	38	3,178	11	1,433	274
Mid-latitude	255	13	28	7	1,073	55
Ice cores only	429	83	1,242	119	4117	185
Total reads	6,740	2,418	5,302	140	6,664	522

Table S12: Number of total sequences for each phylogenetic group of *Raphidonema* and each distribution type.

Distribution type	Group A	Group B	Group C	Group D	Group E	Others	Total
Both poles and mid-latitude	133,446	2,798	118,211	0	0	14	254,469
Antarctica and mid-latitude	212	1,343	101	0	0	0	1,656
Arctic and mid-latitude	13,218	0	3,553	190	21	35	17,017
Both poles only	42,459	15	28,103	0	6,296	19	76,892
Antarctica only	55,737	47,247	783	2	333	5	104,107
Arctic only	24,903	517	18,165	50	20,798	4,211	68,644
Mid-latitude	1,533	38	699	635	266,665	4,244	273,814
Ice cores only	5,109	796	19,879	1,070	68,553	1,643	97,050
Total sequence reads	276,617	52,754	189,494	1,947	362,666	10,171	893,649

Table S13: Maximum likelihood estimates of demographic parameters of the phylotypes based on the coalescent model.

Group A

Phylotypes	τ	θ_0	θ_1	Log-Likelihood
Cosmopolitans	33.8 – 33.9	0.108 – 0.010	0.217 $\pm$ 0.002	-39,824
Multi-regions	11.0	(2.19$\pm$1.64) $\times$ 10^{-5}	2.35$\pm$0.02	-85,267
Endemics	22.0	1.27 $\times$ 10^{-5}	3.35$\pm$0.02	-216,873

Group C

Phylotypes	τ	θ_0	θ_1	Log-Likelihood
Cosmopolitans	7.74$\pm$0.64	0.0179 $\pm$ 0.4723	0.698 $\pm$ 0.004	-76,979
Multi-regions	3.10 $\pm$ 0.03	(0.417$\pm$2.82)$\times$ 10^{-3}	280$\pm$11	-19,891
Endemics	0.48 $\pm$ 0.03	2.27 $\pm$ 0.03	133 $\pm$ 3	-45,236

Group B

Phylotypes	τ	θ_0	θ_1	Log-Likelihood
All	15.0	6.48$\pm$0.04	6.76$\pm$0.04	-196,070

Table S14: Demographic parameters taking account of the mutation rates and generation intervals based on the coalescent model.

Group A

Phylotypes	t	N_0	N_1	$tMRCA_{ML}$	$tMRCA_{BEAST}$
Cosmopolitans	1.4×10^7	$(0.63–6.8)\times10^5$	1.4×10^6	1.8×10^5	1.2×10^5
	$[3.6\times10^6–4.0\times10^7]$	$[3.5\times10^4–1.3\times10^6]$	$[7.6\times10^5–2.6\times10^6]$	$[4.6\times10^4–5.2\times10^5]$	$[1.5\times10^4–3.3\times10^5]$
Endemics	9.2×10^6	80	2.1×10^7	2.8×10^6	2.7×10^6
	$[2.3\times10^6–2.6\times10^7]$	[44–154]	$[1.2\times10^7–4\times10^7]$	$[7.1\times10^5–8.0\times10^6]$	$[4.4\times10^5–6.4\times10^6]$

Group C

Phylotypes	t	N_0	N_1	$tMRCA_{ML}$	$tMRCA_{BEAST}$
Cosmopolitans	3.2×10^6	5.6×10^5	2.2×10^6	5.8×10^5	2.1×10^5
	$[8.2\times10^5–9.2\times10^6]$	$[3.1\times10^5–1.1\times10^6]$	$[1.2\times10^4–4.2\times10^4]$	$[1.5\times10^5–1.7\times10^6]$	$[1.5\times10^4–6.2\times10^5]$
Endemics	1.9×10^5	7.2×10^6	4.2×10^8	2.1×10^6	1.5×10^6
	$[5.1\times10^4–5.7\times10^5]$	$[4.0\times10^6–1.4\times10^7]$	$[2.3\times10^8–8.0\times10^8]$	$[5.3\times10^5–6.0\times10^6]$	$[2.7\times10^5–3.7\times10^6]$

Group B

Phylotypes	t	N_0	N_1	$tMRCA_{ML}$	$tMRCA_{BEAST}$
All	6.2×10^6	2.0×10^7	2.1×10^7	5.6×10^6	3.1×10^6
	$[1.6\times10^6–1.8\times10^7]$	$[1.1\times10^7–3.9\times10^7]$	$[1.2\times10^7–4.1\times10^7]$	$[1.4\times10^6–1.6\times10^7]$	$[4.3\times10^5–7.9\times10^6]$

t: The population expansion time (in years ago).
N_0 : Effective population sizes before the population expansion.
N_1: Effective population sizes after the population expansion.
$tMRCA_{ML}$: Time of the most recent common ancestor (in years ago) by the ML estimation.
$tMRCA_{BEAST}$: Time of the most recent common ancestor (in years ago) by the Bayesian Inference using the BEAST.
Numbers in the braket indicate the ranges of the estimates taking account of the uncertainties of the mutation rates and the generation intervals (for ML) or 95% HPD (for the Bayesian Inference).

Supplementary References

S1. Ness RW, Morgan AD, Colegrave N, Keightley PD. Estimate of the Spontaneous Mutation Rate in *Chlamydomonas reinhardtii*. Genetics. 2012;192(4):1447-54.

S2. Williamson CJ, Cameron KA, Cook JM, Zarsky JD, Stibal M, Edwards A. Glacier Algae: A Dark Past and a Darker Future. Front Microbiol. 2019;10(524).

S3. Onuma Y, Takeuchi N, Tanaka S, Nagatsuka N, Niwano M, Aoki T. Observations and modelling of algal growth on a snowpack in north-western Greenland. The Cryosphere. 2018;12(6):2147-58.

S4. Zachos J, Pagani M, Sloan L, Thomas E, Billups K. Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. Science. 2001;292(5517):686-93.

S5. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Mol Biol Evol. 2005;22(5):1185-92.

S6. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evolution. 2018;4(1).

S7. Yakimovich KM, Gauthier NPG, Engstrom CB, Leya T, Quarmby LM. A Molecular Analysis of Microalgae from Around the Globe to Revise Raphidonema (Trebouxiophyceae, Chlorophyta). J Phycol. 2021;57(5):1419-32.

S8. Mai U, Sayyari E, Mirarab S. Minimum variance rooting of phylogenetic trees and implications for species tree reconstruction. PLoS One. 2017;12(8):e0182238.

S9. Tria FDK, Landan G, Dagan T. Phylogenetic rooting using minimal ancestor deviation. Nature Ecology & Evolution. 2017;1(7):0193.

S10. Takeuchi N, Sera S, Fujita K, Aizen VB, Kubota J. Annual layer counting using pollen grains of the Grigoriev ice core from the Tien Shan Mountains, central Asia. Arct Antarct Alp Res. 2019;51:299-312.

S11. Takeuchi N, Fujita K, Aizen VB, Narama C, Yokoyama Y, Okamoto S, et al. The disappearance of glaciers in the Tien Shan Mountains in Central Asia at the end of Pleistocene. Quat Sci Rev. 2014;103:26-33.

S12. Segawa T, Matsuzaki R, Takeuchi N, Akiyoshi A, Navarro F, Sugiyama S, et al. Bipolar dispersal of red-snow algae. Nature Communications. 2018;9(1):3094.