



Demographic analysis of cyanobacteria based on the mutation rates estimated from an ancient ice core

Takahiro Segawa^{1,2} · Nozomu Takeuchi³ · Koji Fujita⁴ · Vladimir B. Aizen⁵ · Eske Willerslev^{6,7,8} · Takahiro Yonezawa⁹

Received: 13 September 2017 / Revised: 12 November 2017 / Accepted: 20 November 2017 / Published online: 5 January 2018
© The Genetics Society 2018

Abstract

Despite the crucial role of cyanobacteria in various ecosystems, little is known about their evolutionary histories, especially microevolutionary dynamics, because of the lack of knowledge regarding their mutation rates. Here we directly estimated cyanobacterial mutation rates based on ancient DNA analyses of ice core samples collected from Kyrgyz Republic that dates back to ~12,500 cal years before present. We successfully sequenced the 16S rRNA and 16S–23S internal transcribed spacer (ITS) region. Two cyanobacterial operational taxonomic units (OTUs) were detected from the ancient ice core samples, and these OTUs are shared with those from the modern glacier surface. The mutation rate of ITS region was estimated by comparing ancient and modern populations, and were at the magnitude of 10^{-7} substitutions/sites/year. By using a model selection framework, we also demonstrated that the ancient sequences from the ice sample were not contaminated from modern samples. Bayesian demographic analysis based on coalescent theory revealed that cyanobacterial population sizes increased over Asia regions during the Holocene. Thus, our results enhance our understanding of the enigmatic timescale of cyanobacterial microevolution, which has the potential to elucidate the environmental responses of cyanobacteria to the drastic climatic change events of the Quaternary.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41437-017-0040-3>) contains supplementary material, which is available to authorized users.

✉ Takahiro Segawa
tsegawa@yamanashi.ac.jp

¹ Center for Life Science Research, University of Yamanashi, 1110, Shimokato, Chuo, Yamanashi, Japan

² National Institute of Polar Research, 10–3 Midori-cho, Tachikawa, Tokyo, Japan

³ Department of Earth Sciences, Graduate School of Science, Chiba University, 1–33 Yayoi-cho, Inage-ku, Chiba, Japan

⁴ Graduate School of Environmental Studies, Nagoya University, Chikusa-ku, Nagoya, Japan

⁵ Department of Geography, University of Idaho, Moscow, ID 83844, USA

⁶ Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen K 1350, Denmark

⁷ Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

⁸ Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK

⁹ School of Life Sciences, Fudan University, SongHu Rd. 2005, Shanghai 200438, China

Introduction

Diverse microbes inhabit glacier surfaces worldwide, even though snow and ice are extreme environments that have low temperatures and nutrient-poor conditions for microbes. Cyanobacteria, which are photosynthetic autotrophs, are the major group of microbes that live on glaciers. They help sustain other heterotrophic organisms, such as bacteria and microfauna in glacier ecosystems (e.g., Langford et al. 2010; Edwards et al. 2014). Cyanobacteria on glaciers are usually found on the melting ice surface, particularly in cryoconite holes, which are small melt water ponds formed on the ice surface. Cyanobacteria grow in the holes during the melting season and form granular aggregations called cryoconite granules (Langford et al. 2014; Segawa et al. 2014; Cook et al. 2016).

Previous studies revealed that cyanobacteria have strong interactions with glacier environments. Cyanobacterial growth largely depends on glacier surface conditions, and requires sunlight, melt water, and certain amounts of nutrients (Vincent, 2000; Langford et al. 2010). Thus, cyanobacterial abundance and species compositions may vary spatially and geographically on glaciers under different physical or chemical conditions (Langford et al. 2014;

Uetake et al. 2016). Cyanobacterial abundance also affects the glacier itself; cyanobacteria usually live in cryoconite holes and form granules, but they disperse and cover the entire glacier surface on some glaciers (Takeuchi et al. 2001; Segawa and Takeuchi, 2010). Because of the dark coloration of cryoconite granules, their dispersion on the glacier surface can reduce surface albedo and accelerate the melting of snow and ice (Takeuchi et al. 2001; Takeuchi and Li, 2008; Yallop et al. 2012). Particularly, recent increase of cryoconite coverage on the Greenland Ice Sheet has been suggested to be a potential impact on mass loss of the ice sheet (Wientjes and Oerlemans, 2010; Shimada et al. 2016). Therefore, understanding of the ecological traits of glacial cyanobacteria, such as life history, dispersal processes, and demographic history are important to evaluate the interaction between physical and biological aspects of glaciers.

Recently, phylogeographic patterns of cyanobacteria have been resolved by enthusiastic studies (Bahl et al. 2011; Dvořák et al. 2012). It has been speculated that cyanobacterial spatial distribution patterns in the cryosphere are essentially global at the species level by 16S rRNA gene analysis (Jungblut et al. 2010; Harding et al. 2011). According to 16S rRNA sequences analysis by Christmas et al. (2015), several clades of cold-dwelling cyanobacteria are common to cold regions, and 20 clades have common ancestors that were highly likely adapted to cold environments. However, that analysis focused upon publicly available sequences deposited in DNA database. Although many cosmopolitan microorganisms in Antarctica are frequently dispersed by aerial exchange with the rest of the world (Vincent, 2000), recent work demonstrated evidence of regional differentiation and adaptations at the population level by internal transcribed spacer (ITS) sequences (Segawa et al. 2017). Segawa et al. (2017) suggested that cyanobacteria on glaciers are fundamentally endemic at the population level and there is currently limited migration among regions. Therefore, it remains unclear how, when, and to what extent cyanobacteria expand their distributions on glaciers, because limited information on their mutation rates makes it difficult to elucidate demographic dynamics and establish their phylogeographic patterns in cold environments.

Ice cores drilled from glaciers can provide a means of direct analysis of microbes that grew in the past. Ice cores are usually drilled in the accumulation area of glaciers, where surface snow rarely melts. Because snow remains year-round without melting, annual snow layers continuously accumulate in this area. Ice cores are vertically drilled in these areas and analyzed to obtain continuous past to present inter-annual climate information. For example, studies discovered microorganisms in Antarctic ice cores by 16S rRNA gene analyses (Bulat et al. 2004; Christner et al.

2006), and past flora and fauna from the Greenland ice cores by analyzing chloroplast and mitochondrial DNA (Willerslev et al. 2007). In contrast to the polar ice cores, the ice cores drilled from glaciers located in low latitudes often contain melt features, which are refrozen ice layers that are formed during warm climate periods. Microbes rarely grow on the snow surface without melt, but can grow on melting snow surfaces (Uetake et al. 2006; Kohshima et al. 2007). Therefore, microbes that grew on the snow surface in the past can be found in the ice cores of low-latitude glaciers. In fact, abundant snow algae and cyanobacteria that grew on the glaciers have been found in ice cores drilled from glaciers in the Asian high mountains, including the Himalayas and Tianshan Mountains (Takeuchi et al. 2009; Chen et al. 2016). These studies suggested that such microbes can be used as more reliable markers for identifying annual layer boundaries and inferring other environmental information. Application of molecular techniques to analyzing microbes in the historical glacier layers could provide direct comparison of past and present microbes. The knowledge obtained by this study will illuminate their evolutionary history, especially the micro-evolutionary processes, such as demographic movements and regional differentiation.

In this study, we directly estimated cyanobacterial mutation rate by analyzing past cyanobacterial communities from ice core samples collected from the Gregoriev Glacier, Kyrgyz Republic, which date back to approximately 12,500 cal years before present (yr BP). We successfully sequenced nucleotide data from the 16S rRNA gene and ITS regions. We also describe a new method in a model selection framework to evaluate whether an observed sequence from an ancient sample is real ancient DNA or contaminant DNA that originated from modern environments. Here we report cyanobacterial operational taxonomic units (OTUs) from ancient ice cores, estimates of their mutation rates, and dynamics of their population sizes based on these data. Results of this study regarding cyanobacterial demography on glaciers will provide important information for understanding glacial ecosystems, changes in glacier albedo, and potential future population dynamics.

Materials and methods

Ice core samples

The analyzed Gregoriev Glacier ice core was drilled in 2007 at the top of the Gregoriev Ice Cap (41°58'33"N, 77°54'48"E) in Kyrgyz Republic, Central Asia (Fig. 1). The Gregoriev Ice Cap is located in the Teskey Ala-Too Range in Inner Tien Shan, which is an area that runs southward from Lake Issyk-Kul. The ice cap is a simple ice cap that covers a

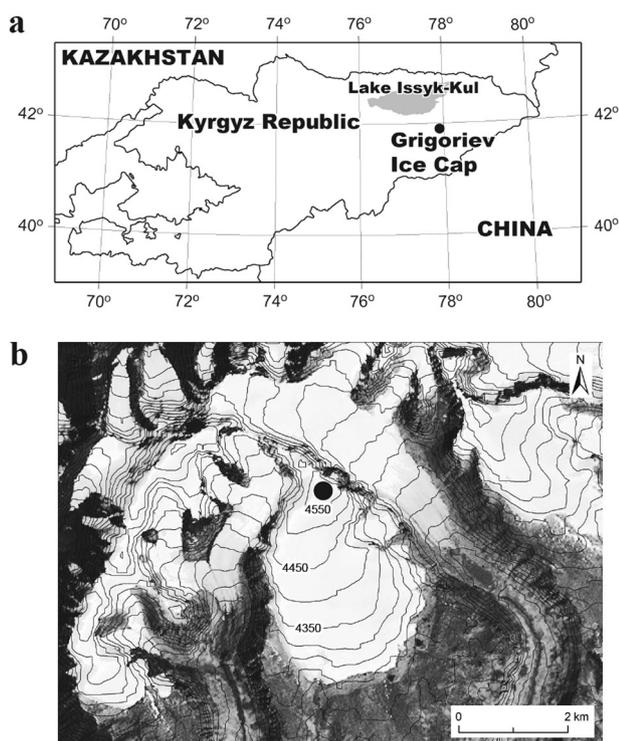


Fig. 1 Map that shows the locations **a** and the drilling site **b** of the ice core samples, the Grigoriev Ice Cap in Kyrgyz Republic, Central Asia. White area indicates present glacial coverage based on a satellite image (Landsat) acquired on 1999

small mountain that is approximately 8 km² in area. The elevation of the drill site was 4,563 m above sea level. Meteorological observations at the site showed that annual mean air temperature was -11.1°C , and occasionally exceeded the freezing point in July and August, which indicates occasional surface snow melt that allows microbial growth at the site (Fujita et al. 2011). The length of the ice core from the surface to the bottom was 86.87 m. Borehole temperature measurements revealed that the ice temperature was -2.68°C at 10 m deep and generally decreased with depth down to -3.92°C at the bottom (Takeuchi et al. 2014). The negative temperature at the bottom indicates basal ice has not melted.

Two samples from the ice core were analyzed in this study. This first sample was collected from the ice layers at the 85.54 m in depth (the Ice-dust sample), which is a prominent dust layer that contains abundant mineral dust and organic matter. The high concentration of the particles in the layer indicates that significant melt of the glacier surface occurred at the time. The second sample was frozen organic soil collected from the bottom of the borehole of the ice core at 86.87 m (the Bot sample). This sample also contained abundant mineral dust and particulate organic matter. These samples have been dated with radiocarbon measurement and were 8,155–8,002 cal. yr BP for the Ice-

dust sample, and 12,656–12,434 cal. yr BP for the Bot sample (Takeuchi et al. 2014).

Melting of ice core sample

All pre-PCR amplification steps were carried out in the ancient DNA clean lab (class 1,000) at the National Institute of Polar Research, Japan. The ancient DNA clean room undergoes nightly UV irradiation, is regularly cleaned with bleach, and physically separated from other modern biological laboratories to avoid contamination from amplified products. All of the equipment and reagents used for DNA extractions were cleaned before use with bleach, DNA AWAY (Thermo Fisher Scientific, USA), and/or UV treatment at 254 nm for 10 min (UVP cross-linker, CL-1000). During analysis in the clean room, gamma ray-sterilized full bodysuits, facemasks, and disposable gloves were used.

The ice-dust sample was melted in a clean type freezer (class 100) (Seiken Medical, Japan) using a melting device developed by our group that enabled us to obtain water from only the inner portion of the cores and allowed the 10-mm thick outer layer to remain intact (Segawa et al. 2013). Complete separation of the inner and outer cores is required to avoid contamination by microorganisms that can adhere to the cores during drilling and storage. To determine whether any contamination occurred during handling or due to ice core cracks, a solution of approximately 2 ng/ μl bacterial plasmid vector (pCR4-TOPO vector, Life Technologies, USA) was applied to the surface of the core samples. The vector contaminant was not amplified from the DNA extract of the inner part of the ice core samples, and only amplified with the outer layer of ice core samples by 45 cycles of PCR with vector-specific primers (M13-Forward, M13-Reverse). Therefore, we proceeded with further DNA analyses of the inner portion of the ice core samples. For the Bot sample, the sample was melted in sterile 50-ml tubes (Eppendorf, Germany) at 4°C , and was used to the following analysis.

DNA extraction

DNA extractions were conducted in a class 100 clean bench (MHE-130AB3, Panasonic, Japan), and the subsequent PCR set up procedures were carried out in another class 100 clean bench in the ancient clean room. To avoid contamination from PCR products, all of the procedures before PCR were performed in a room separate from that which contained thermal cyclers and PCR products. After melting, the samples were transferred to sterilized tubes and centrifuged at $10,000\times g$ for 10 min. The pellet was then transferred into separate 2.0 mL Matrix-E tubes (Qbiogene, USA), and DNA extraction was performed as previously

described (Willerslev et al. 2003) except for the bead-beating procedure; we used a multi-bead shocker at 2500 r.p.m for 30S (Yasui Kikai, Japan).

16S rRNA–ITS long reads by PCR amplification and sequencing

Although high-throughput sequencing is commonly applied to 16S *rRNA* gene and 16S–23S ITS amplicon analyses, contiguous 16S–ITS rRNA sequences from the same template molecules cannot be produced by Second generation sequencing technologies. Because of the lack of ITS sequence data in the DNA database, information regarding the sequences detected on the glacier is limited at present (e.g., only 5% of cyanobacterial OTUs on the glacier surface had over 95% similarity to sequences in the NCBI-nt database) (Segawa et al. 2017); therefore, it is difficult to estimate taxonomic information based only on ITS sequences. To perform parallel broad-scale (16S *rRNA* gene) and fine-scale (16S–23S ITS) analyses on the same individuals, we conducted conventional Sanger-based sequencing of clone libraries.

PCR amplification of the cyanobacterial 16S *rRNA* gene to 23S rRNA ITS regions was performed using high-fidelity PrimeStar polymerase (Takara, Japan) and primers 27FHT (Isenbarger et al. 2008) and 23S30R (Lepère et al. 2000). The PCR conditions were as follows: 3 min of initial denaturation at 94 °C, 30 cycles at 98 °C for 10S, 57 °C for 15S, and 72 °C for 4 min, and a final extension at 72 °C for 7 min. To minimize PCR drift, 3–5 independent reactions were conducted. The pooled PCR products were purified using the MinElute PCR Purification Kit (Qiagen, USA) and cloned using the pCR4Blunt-TOPO vector with a Zero Blunt TOPO PCR Cloning Kit for Sequencing (Life Technologies). *Escherichia coli* HST08 Premium Competent Cells (Takara) were transformed with the cloning vector. We sequenced 200 clones in each sample. Sequencing was achieved using Big Dye Terminator v3.1 and an ABI 3130xl automatic sequencer (Life Technologies) with primers M13F, M13R, and 907 F (Lane et al. 1985). As negative controls, we tested DNA extractions and PCRs with blank samples, as previously described.

BLASTN searches were performed to identify the closest sequences in the NCBI-nt database. The sequences that matched the cyanobacterial sequences were used for the following analysis. The sequencing data were analyzed using Mothur v. 1.38.1 (Schloss et al. 2009). Chimeric sequences were identified and removed using chimera.uchime (Edgar et al. 2011). Mothur data were used to calculate coverage with the formula $1 - n/N$, where n is the number of singletons and N is the total number of sequences.

Taxonomic classification of metagenomic shotgun library reads composition based on 16S and 18S rRNA gene analysis

Because PCR biases can result in misestimation of the relative abundances of cyanobacteria, we also directly estimated cyanobacterial diversity from shotgun metagenomic data. To build a whole-genome shotgun library, we conducted a whole-genome amplification (WGA) technique as previously described (Hongoh et al. 2008) using the Illustra GenomiPhi HY DNA Amplification Kit (GE Healthcare, UK). To evaluate potential contamination during DNA extraction and WGA, blank WGA with DNA extraction with blanks instead of DNA input was included in the workflow; we confirm that there was no amplification (contamination) in the negative controls.

Sequencing libraries were prepared using the TruSeq DNA PCR-Free Sample Prep Kit (Illumina, USA). Paired-end reads were generated on an Illumina HiSeq platform in the 100-bp paired-end mode. Adapter sequences were removed from the generated reads using Cutadapt v. 1.11 and quality trimming using Prinseq-lite v. 0.20.4 (Martin, 2011; Schmieder and Edwards, 2011). We discarded the reads that contained < 50 bp or were associated with an average Phred-like quality score < 20. As a result, 24,440,041 and 28,149,442 read pairs were generated, and 24,226,994 and 27,905,908 read pairs were obtained after trimming the Ice-dust and Bot sequences, respectively.

Shotgun metagenomic data were used for all taxonomic analyses by small subunit (SSU) *rRNA* gene sequences. High-quality reads were used for taxonomic assessment by screening for SSU *rRNA* genes with Metaxa2 (Bengtsson-Palme et al. 2015) to extract bacterial 16S *rRNA* gene and eukaryote 18S rRNA sequences from shotgun metagenomic data. Chimeric sequences were identified and removed using chimera.uchime, and extracted SSU gene fragments were matched to Silva v. 123 database sequences using the classify.seqs command in Mothur v. 1.38.1. Taxonomy was then assigned to the matching Silva sequences.

Phylogenetic analysis, mutation rate and demographic dynamic inference

We use the Sanger sequences in the following analysis. We initially clustered the OTUs at the 98% nucleotide identity level based on only the 16S *rRNA* gene sequences. We then sub-clustered unique OTUs based on the ITS sequences in each sample. We searched tRNA sequences using tRNADB-CE (Abe et al. 2014), and removed tRNA sequences for the following analysis. The representative sequences of ITS sequences for each haplotype were aligned with cyanobacterial ITS sequences from glaciers

using MAFFT (Kato and Toh, 2010), and carefully checked by eye; all ambiguous sites and clones were manually deleted. Maximum likelihood trees were reconstructed using RAxML v. 7.2.8 (Stamatakis, 2006) with the GTR + I + Γ models. Haplotype networks were reconstructed based on the ITS sequences with the median-joining method (Bandelt et al. 1999) using NETWORK v. 4.6.11 (<http://www.fluxus-engineering.com/sharenet.htm>).

To estimate the mutation rate of the ITS region and demographic dynamics, BEAST v. 1.7.4 (Drummond and Rambaut, 2007) was used under the HKY + Γ model with the log-normal relaxed molecular clock. The tip-date method was used to estimate the absolute timescale based on the age of the ancient ice core samples. The calibrated isotopic age of the ice core samples was determined to be 12,500 (± 106) cal. yr BP. Therefore, we assumed that the age of the Bot sample was 12,500 years before present, whereas the modern samples were 0 years before present. The results were summarized and visualized using Tracer v. 1.5 (<http://beast.bio.ed.ac.uk/tracer>). Parameter convergence was confirmed by checking that all effective sample sizes were > 200 . To estimate mutation rates, only unique sequences were included to reduce the computational burden. A constant population size was assumed.

The demographic dynamics were inferred using the Bayesian skyline plot (BSP) method (Drummond et al. 2005). Because this method assumes the coalescent process, all individuals should be included in this analysis. In this analysis, we assumed one clone sequence as one individual, and all sequences were included. However, as that approach was computationally too expensive, the mutation rates were fixed a priori in accordance with the estimated mutation rates mentioned above.

The maximum likelihood (ML) method was also used to infer the demographic dynamics and the time of the most recent common ancestor using GENETREE v. 9.0 (Griffiths, 2007). This method is based on the coalescent process, and assumes an infinite alleles model and no recombination. Exponential growth of the population sizes was assumed. The above-mentioned estimated mutation rates were applied. Because little is known about the generation intervals of cyanobacteria in wild environments, four different generation intervals of 3, 29, 44, 300 generations/year were considered because they were consistent with the range of bacterial generation times reported in previous studies (Nadeau and Castenholz, 2000; Vincent, 2007; Anesio et al. 2010; Schieruf and Wiuf, 2010).

Model selection approaches for molecular clock model comparison

To ensure that the observed sequence of the ancient Bot sample was not the contamination that originated from

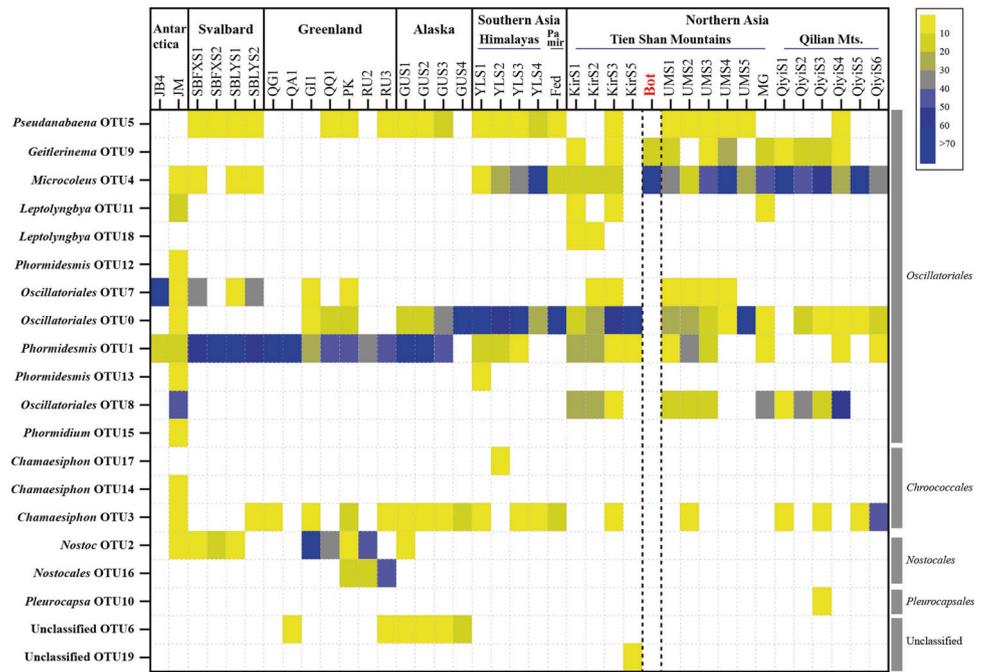
modern environments, we compared four models (Supplementary Figure S1) using Akaike Information Criterion through Markov chain Monte Carlo (AICM) (Baele et al. 2012) with 100 bootstraps: model A, in which the sequences from the ancient sample were assigned to 12,500 cal. yr BP, and the modern samples were assigned to 0 cal. yr BP; model B, in which the sequences from the ancient sample were assigned to 0 cal. yr BP, and randomly selected N haplotypes from modern glacier surface samples were assigned to 12,500 cal. yr BP, where N is the number of haplotypes from the Bot sample.; model C, in which the sequences from the ancient sample were assigned to 0 cal. yr BP, and all modern glacier surface sequences were assigned to 12,500 cal. yr BP; and model D, all samples were assigned to be modern (0 years before present). Model A represents a situation in which all ancient samples are actually 12,500 cal. yr BP, there was no contamination from the modern samples. On the other hand, model D is assuming all DNA sequences are from the modern samples. If the AICM of model D is smaller than that of model A, it is a signal of that the DNA sequences from the ancient samples are contamination from the modern samples. Models B and C represent situations in which there was contamination from modern samples. Cases in which the modern DNA was mistakenly assigned as ancient DNA can be examined by comparison of these models and model A. To reduce the computational burden, one sequence from one haplotype was used in this analysis. If haplotypes are shared by the modern surface samples and the Bot sample, two sequences (one is from the Bot and another is from the current glacier surface) from one haplotypes were involved in this analysis. Appropriate effective sample size in the analyses were checked with Tracer v. 1.6 (<http://beast.bio.ed.ac.uk/tracer>).

Results

Cyanobacterial community structure based on 16S-ITS PCR and a shotgun metagenome approach

We performed continuous 16S-ITS sequencing and metagenome shotgun libraries to analyze the cyanobacterial community structures in the ice cores. Using a Sanger-based clone library approach, the Good's coverage values based on the 16S *rRNA* gene of cyanobacteria were 100% in the clone libraries for both Ice-dust and Bot samples. The samples for the clone libraries sufficiently covered most of the cyanobacterial diversity. The 16S *rRNA* gene sequences from Ice-dust and Bot samples were grouped into just one and two cyanobacterial OTUs, respectively. The 16S *rRNA* gene of the Ice-dust sample was close to *Chroococcidiopsis* sp. CC1 (DQ914863), with 94% similarity. However, this

Fig. 2 Geographic distribution of cyanobacteria from glacier surface (Segawa et al. 2017) and ice core samples of this study by 16S rRNA gene analysis. A representative heat map of clone percentage in the samples is shown. Blue and yellow indicate high and low expression, respectively. White indicates no detectable signal



OTU is not closely related to cyanobacterial 16S sequences on glaciers. In addition, metagenome shotgun sequences of the Ice-dust sample were not similar to any previously deposited cyanobacterial 16S *rRNA* gene sequences. Therefore, it was suggested that the Ice-dust sample did not contain cyanobacteria that typically grow on glaciers. However, the Bot sample contained two cyanobacterial OTUs classified to Oscillatoriales based on Sanger-based analysis. The 16S *rRNA* genes in the Bot sample were affiliated with *Microcoleus* (85%) and *Geitlerinema* (15%). These two OTUs are closely related to cyanobacteria that are known to occur on glaciers, *Microcoleus* OTU4 and *Geitlerinema* OTU9 classified by Segawa et al. (2017), respectively (Fig. 2).

Using a metagenomic shotgun library approach, 69 reads were assigned to the cyanobacterial 16S *rRNA* gene in the Bot sample. A broader range of cyanobacterial taxa were found in the shotgun library compared with the Sanger library. The majority of the 16S *rRNA* genes in the shotgun library were affiliated with Oscillatoriales (75%); the other genes were affiliated with Synechococcales (10%), Nostocales (13%), and Chroococcales (1%). The frequently detected families were Microcoleaceae (43%) and Coleofasciculaceae (17%) (Supplementary Figure S2), consistent with the clone library analysis. The metagenomic analyses also revealed microbial community composition in the ice cores other than cyanobacteria (Supplementary Figure S3).

Cyanobacterial analysis of ITS regions

Overall, we observed geographically widespread OTUs by 16S *rRNA* gene analysis. However, the ITS region, with a higher evolution rate, often shows contrasting cyanobacterial distribution patterns. The haplotype networks based on the ITS sequences showed that both OTUs have regional biogeographical distribution patterns (Fig. 3). The haplotype network revealed that the endemic haplotypes independently derived from ubiquitous ancestral haplotypes multiple times. The result of our ITS analyses showed that both OTUs were fundamentally endemic with regional distributions at the strain level.

Cyanobacterial mutation rate estimation

To accurately and directly estimate the cyanobacterial evolutionary rate, the nucleotide sequence data of the modern and ancient samples enclosed in the ice core excavated from the Bot sample were analyzed. We conducted tip-date analysis under the coalescent process to estimate mutation rates. Figure 4 indicates the posterior distributions of the mutation rates. The posterior means of the 16S–23S ITS region mutation rate in OTUs 4 and 9 were 5.7×10^{-7} and 2.3×10^{-7} substitutions/site/year, respectively.

To ensure the observed sequences of the ancient Bot sample were not contaminated by modern environments, we compared four dating models using AICM (models A, B, C,

Fig. 3 Haplotype network analysis of 16S–23S ITS regions of OTU4 and OTU9. The haplotypes were colored based on geographical region, and the areas of each haplotype are proportional to the clone numbers of the unique sequence. Each color represents a haplotype from the ice core of this study and glacier samples (Segawa et al. 2017) as follows: red, ice core sample; blue, Antarctic glacier; green, Svalbard glacier; black, northern Asian glacier; and yellow, southern Asian glacier

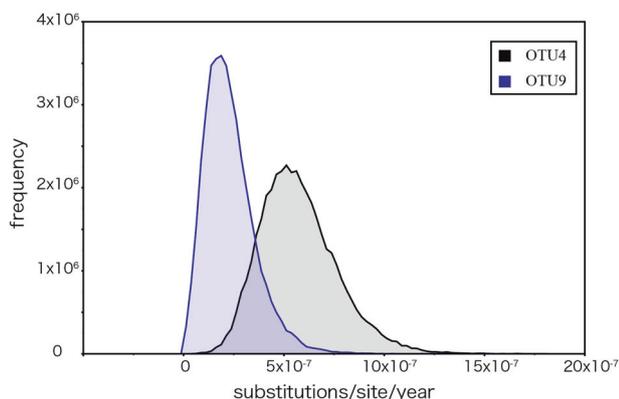
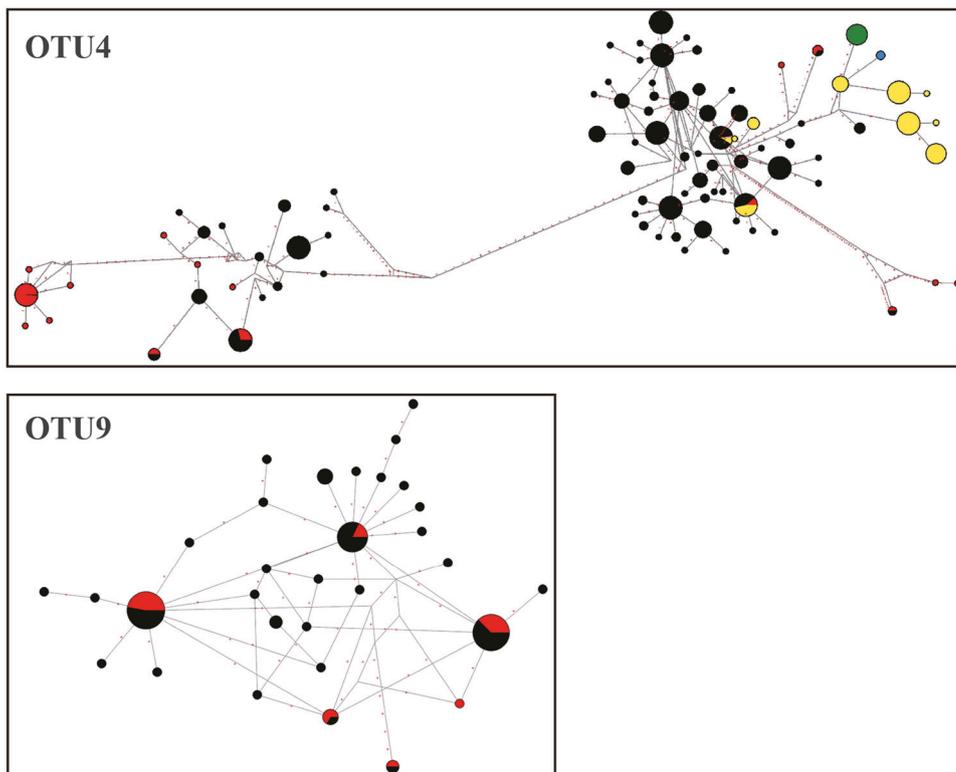


Fig. 4 Estimated rate of molecular evolution. The posterior probability densities of the evolutionary rate (substitutions/site/year) for 16S–23S ITS regions of cyanobacterial OTU4 and OTU9 under a relaxed molecular clock

and D). We computed the AICM for each model, and the results revealed model A, that assigned the ancient DNA from the Bot samples are 12,500 cal. yr BP had the smallest AICM values. The comparisons of AICM of each model were as follows: OTU4, $A < C < B < D$; OTU9, $A < D < B < C$ (Table 1). It suggests that the ancient sequences from the Bot sample were not contaminated with modern samples.

Table 1 Demographic and molecular clock model comparison analyzed by investigated Akaike's information criterion (AICM)

	OTU 4		OTU 9	
	AICM	S.E.	AICM	S.E.
Model A	7454.385	± 0.272	1516.732	± 0.038
Model B	7473.122	± 0.238	1520.227	± 0.040
Model C	7465.633	± 0.194	1529.998	± 0.091
Model D	7476.156	± 0.152	1518.917	± 0.072

Model A: The haplotypes from the ice core were assigned to 12,500 years before present

Model B: The haplotypes from the ice core were assigned to 0 years before present, and randomly selected N haplotypes from modern glacier surface samples were assigned to 12,500 cal. yr BP, where N is the number of haplotypes from the ice core sample

Model C: The haplotypes from the ice core were assigned to 0 years before present, k and all haplotypes from current glacier surface were assigned to 12,500 years before present

Model D: All haplotypes were assigned to be modern (0 years before present)

The comparisons of these AICM values, especially the model A vs. the model B, as well as the model A vs. the model D, also indicate the number of mutations occurred in the cyanobacterial ITS regions for 12,500 years are statistically significant. In other words, the mutation rates estimated by the comparisons of the modern samples and the

Bot samples are evolutionary significant, but not the artifact caused by the molecular variances in one chronological point.

Demographic reconstruction

We estimate the demographic movements of cyanobacterial OTU4 and OTU9 based on ITS regions using BSP analysis. Effective population sizes (N_e) of cyanobacterial OTU4 and OTU9 were estimated over time (Fig. 5). BSP revealed different tendencies of the demographic movements of cyanobacterial OTUs. The population size of OTU4 has been relatively stable over time but with a weak increase in N_e over the past 10,000 years, whereas that of OTU9 showed dynamic expansion, especially for approximately the past 10,000 years.

The ML estimation of the demographic movements with GENETREE v. 9 under the exponential growth model is shown in Supplementary Figure S4. OTU9 showed remarkably increased N_e , whereas OTU4 did not show increased N_e over time.

Discussion

Cyanobacterial DNA in the ice core samples

In this study, we successfully analyzed continuous 16S rRNA–ITS regions from the ancient ice core samples. The heat map of the geographic distribution of cyanobacterial OTUs from glacier and ice core samples based on the 16S rRNA gene analysis is shown in Fig. 2. On the basis of the detection rate of 16S rRNA gene in geographic regions, these OTUs from the Bot sample can be categorized as cosmopolitan (OTU4) and Asian that is dominant in northern Asia sites (OTU9) (Segawa et al. 2017). Although OTU4 observed on both polar and Asian glaciers, OTU4 dominates in Northern Asia. Therefore, both OTUs are abundant in Northern Asia. Since the Bot sample was frozen organic soil under the glacier, cyanobacteria in the sample may be different from those of glacier surface biota. However, according to Kaštovská et al. (2005), most of the cyanobacterial species detected in cryoconite on glaciers have been found in soil surrounding the glaciers. Furthermore, there were some cyanobacterial OTUs shared between the soil and cryoconites on glaciers (Segawa et al. 2014). Therefore, we assume that the Cyanobacteria frequently migrate between the glacial surface and surrounding soil, and thus their populations in the soils and ice core are not necessary to be distinguished.

In general, DNA is fragmented in ancient samples, such as bones, ice cores, and permafrost from the past 10,000 years. The degree of DNA strand cleavages depends on

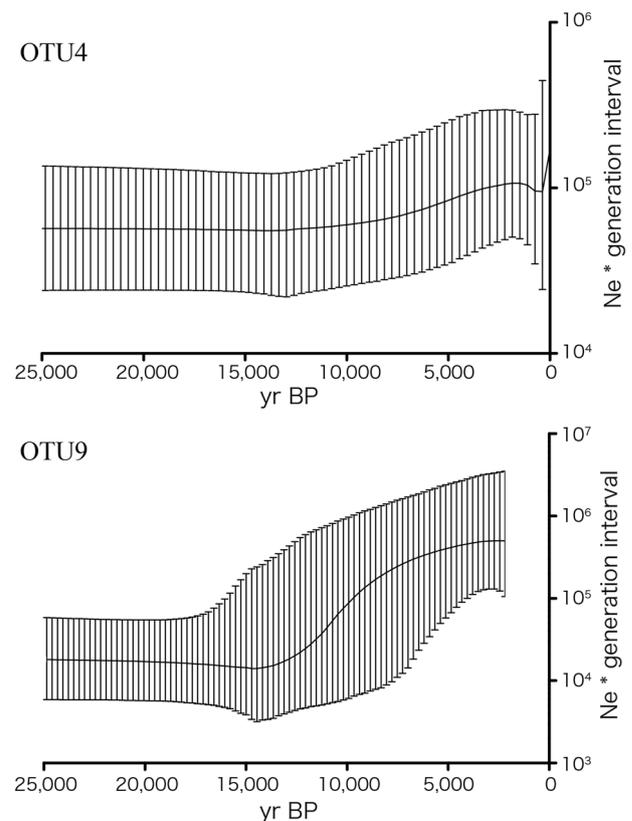


Fig. 5 Bayesian skyline plots that show the $N_e \times t$, where N_e stands for the effective population size and t stands for the generation intervals over time in cyanobacterial OTU4 and OTU9. Lines represent the median parameter estimates, with dashed lines representing the 95% highest posterior density interval. The x axis indicates years before present; the y axis indicates the effective population size multiplied by the generation interval ($N_e \times t$)

temperature and pH, and the short fragments of DNA could be present for a very long time (Allentoft et al. 2012). Allentoft et al. (2012) modeled the decay of a 500-bp mtDNA fragment preserved in bone, and demonstrated a half-life period is 9500 years at -5°C . Indeed, Willerslev et al. (2004) successfully reported DNA concentrations in permafrost samples, and demonstrated the dynamics of the taxonomic diversity for the past 600,000 years. In this study, the ice samples might have been kept frozen under the glacier, and age of the bottom sample is 12,500 cal. yr BP. DNA can be preserved at low temperatures over a relatively short geological time. In addition, both cyanobacterial OTUs detected from the ice core samples have cell walls (Segawa et al. 2017). It is possible that this structure helps DNA from strand cleavage, and this may be the reason why we successfully amplified a continuous 16S–ITS region.

The possibility of contamination from modern DNA samples was also evaluated using a molecular evolutionary approach. Jin et al. (2016) demonstrated that the

appropriateness of fossil calibrations for divergence time estimation can be evaluated in selection model selection framework. We extended this idea to ancient DNA studies to determine whether there was any contamination from modern samples. The four models (models A–D; Materials and Methods section) were constructed to depict situations with and without contamination from modern samples. Model A depicts no contamination from modern DNA, and models B, C and D depict DNA sequences from the Bot sample that were contaminated modern DNA. In cases of both *Microcoleus* OTU4 and *Geitlerinema* OTU9, model A was selected as the best model. This finding indicates that DNA sequences from the Bot sample really originated from ancient Cyanobacteria that date back to 12,500 cal. yr BP.

Mutation rate of the cyanobacteria ITS region

Our previous study demonstrated that the continuous 16S rRNA and ITS region data provide crucial information to reveal the microbial taxonomic diversity at both broad and fine scales, because the lower mutation rate of 16S rRNA is suitable for the inter-species comparison over a long timescale, and the higher mutation rate of the ITS region is suitable for intra-species comparison over a short timescale (Segawa et al. 2017).

In this study, we successfully estimated mutation rates of the cyanobacterial ITS region in *Microcoleus* OTU4 and *Geitlerinema* OTU9, and the posterior means of the mutation rate were 5.7×10^{-7} and 2.3×10^{-7} substitutions/site/year, respectively. These rates are of similar magnitude to the mitochondrial mutation rates of tetrapods (10^{-7} substitutions/site/year), which were previously inferred by ancient DNA analyses (Ho et al. 2007; Molak et al. 2013), but are of dissimilar magnitude compared with mitochondrial DNA of penguins and wild boars (10^{-6} substitutions/site/year). The estimated mutation rates for the ITS region in cyanobacteria are also largely similar to the genomic mutation rate of cultured bacteria. According to Ford et al. (2011), the mutation rate of *Mycobacterium tuberculosis* during latent infection under laboratory incubation was $\sim 8 \times 10^{-8}$ substitutions/site/year.

However, cyanobacterial molecular evolutionary rates estimated with paleontological evidence are substantially slower than our estimates. According to Bahl et al. (2011), the mean evolutionary rate of cyanobacterial ITS regions calibrated by the fossil record is 3.8×10^{-11} substitutions/site/year. Schirmermeister et al. (2013) also carried out divergence time estimation of cyanobacteria using reliable fossil records. Although Schirmermeister et al. (2013) did not report cyanobacterial molecular evolutionary rates, our recalculated evolutionary rate based on their data and their estimated divergence times is approximately 2.0×10^{-11} substitutions/site/year for 16S rRNA (data not shown). This

value is similar to the estimate by Bahl et al. (2011) for cyanobacterial 16S rRNA (1.9×10^{-11} substitutions/site/year).

Indeed, OTUs analyzed in this study and in the study by Bahl et al. (2011) belong to different orders (Oscillatoriales and Chroococciopsidales, respectively); however, the mutation rate estimated in this study is four magnitudes higher than their estimation. Because all ambiguously aligned sites were removed a priori in analyses by Bahl et al. (2011), the actual evolutionary rate of cyanobacterial ITS regions may be much higher than 10^{-11} substitutions/site/year. ITS regions in cyanobacteria contain conserved regions and numerous in-del regions that is difficult to be aligned even within the order; therefore, evolutionary rates were quite different between these regions.

It is possible this discrepancy can be explained by the time dependency of the evolutionary rates (Ho et al. 2005; Ho and Larson, 2006). Ho et al. (2005; 2006) demonstrated the higher-molecular evolutionary rates in the short term (e.g., < 1 million years ago) and lower evolutionary rates in the long term (e.g., > 1 million years ago); they approximated this time-dependent decline in molecular evolutionary rates using an exponential curve. The former correspond to the (non-lethal) mutation rates, and the latter correspond to substitution rates. Mutation rate refers to new genetic mutations that spontaneously occur during the replication and transmission of genetic information from parent to offspring, whereas substitution rate is some portion of new mutations that fix in an evolving lineage over time, which typically depends on both the mutation rate and the effects of natural selection (Barrick and Lenski, 2013).

According to the nearly neutral theory (Ohta and Tachida, 1990), lethal or deleterious mutations are immediately swept from populations, but it takes a long time to completely eradicate slightly deleterious mutations from populations. Consequently, more mutations can be observed with intra-species variation than those expected from inter-species comparisons (Yonezawa and Hasegawa, 2014). Therefore, observed molecular evolutionary rates are higher in the intra-species comparison than those of the inter-species comparison (Ho et al. 2005).

Typically, divergence time estimates for animals and plants are inferred using the ages of fossilized ancestors as calibration points (Benton et al. 2009). Because these fossil calibrations can be typically applied among higher taxa, the molecular evolutionary rates estimated with the fossil calibration are the substitution rates. However, it is difficult to estimate mutation rates within species. Ancient DNA samples can facilitate direct measurement of mutation rates (Ho et al. 2007; Molak et al. 2013).

Recently, mutation rate estimation in animals, such as from bison bones recovered from permafrost, has been reported in several studies (Orlando et al. 2013; Der

Sarkissian et al. 2015). However, mutation rate estimation of microorganisms has only been carried out under laboratory conditions (MacLean et al. 2013); therefore, microorganism mutation rates in natural environments remain unclear (Ho et al. 2007). Mutation rates provide insight into past population dynamics (Orland et al. 2013) and mathematical models regarding speciation dynamics (Gavrilets, 2004); thus, it is important to understand mutation rates in these natural environments.

Timescale of Cyanobacteria demography and phylogeography

Cyanobacteria are the only organisms that have evolved oxygenic photosynthesis, and are some of the most important life forms that influenced Earth environments in the Precambrian (Lane, 2005; Schirmermeister et al. 2015). Cyanobacteria changed the Earth environments from anoxic to aerobic. In particular, the rapid accumulation of oxygen in the atmosphere 2.5–2.3 billion years ago is referred to as the Great Oxidation Event; this caused the largest shifts of biota in earth history and might have triggered the emergence of eukaryotes (Lane, 2005). In addition, the chloroplast in plants and algae was formed by symbiosis of cyanobacteria and eukaryotes, which caused further diversification of eukaryotes. Therefore, the evolution of the cyanobacteria in Earth's history has been vigorously investigated, especially in the fields of molecular evolution and paleontology, and the impact of cyanobacteria on major geological events has been extensively argued (e.g., Schirmermeister et al. 2015; Nutman et al. 2016).

Nevertheless, cyanobacterial evolutionary history during the short term, such as at the intra-species level, has not been well studied. Although several previous studies carried out demographic studies on cyanobacteria, these were not argued in the absolute timescale (e.g., Takuno et al. 2012). Therefore, although the repetitive glacial and interglacial periods in the Quaternary period may have potentially deeply impacted cyanobacterial evolution, little is known about this relationship. This issue mainly stems from difficulty in estimating accurate mutation rates.

Ancient DNA provides crucial information to directly estimate mutation rates (Ho et al. 2007). In this study, the cyanobacterial mutation rates in the ITS region were estimated to be $\sim 2.3\text{--}5.7 \times 10^{-7}$ substitutions/site/year. We further estimated the demographic dynamics of cyanobacterial OTU4 and OTU9 by BSP analysis. Estimated N_e of cyanobacterial OTU4 and OTU9 over time were shown (Fig. 5). These two cyanobacterial OTUs showed different tendencies with regard to demographic dynamics. The N_e of OTU4 has been stable over time, whereas, the N_e of OTU9 showed rapid increase with particularly steep increases during the Holocene. ML estimation of the N_e

with GENETREE also showed similar tendencies: the N_e for OTU9 exponentially increased over time; however, the N_e for OTU4 did not increase (Supplementary Figure S4). These findings indicate the different environmental responsiveness of the cyanobacterial OTUs. However, although it was not detected in the GENETREE ML analysis, OTU4 also showed a weak increase in N_e since approximately 10,000 years before present in the BSP analysis. This suggests that the population sizes of the cyanobacterial OTUs have been more or less increased in Holocene compared with the last glacial period. It is possible that the increase of the solar irradiance and the atmospheric temperature based on the Milankovitch cycles somewhat affect the demographic dynamics of the cyanobacterial OTUs, but the sensitivities of these factors are different in each OTU.

We demonstrated that ancient DNA analyses provide crucial information for estimating mutation rates and changing in geographic distributions. The ancient DNA data also provide new insight into cyanobacterial biogeography. Recently, Segawa et al. (2017) demonstrated the endemic geographic distribution of cyanobacterial OTUs at the population level. However, the evolutionary timescale in which geographic endemism was established was not investigated; the mutation rates obtained in this study make this possible to estimate. Our findings underscore the importance of understanding both the ecology of cyanobacteria in glacial ecosystems and the formation of the cryoconite, which reduces the ice surface albedo of glaciers. This knowledge will help elucidate cyanobacterial micro-evolutionary patterns, such as demographic dynamics, the formation of geographic patterns, genetic differentiations, and natural selection.

Data accessibility

The Raw Illumina sequence data sets have been submitted to the DDBJ Short Read Archive under accession number DRA005575. The nucleotide sequences have been deposited in DDBJ/EMBL/GenBank under the accession numbers LC257535-LC257585.

Acknowledgements We express special thanks to Prof. Hirohisa Kishino for discussion on mutation rate estimation and methods that potentially detect contamination in model selection framework. This study was supported by Grants-in-Aid for Scientific Research (no.17H01854, 26281017, 26247078, and 26241020) from the Japan Society for the Promotion of Science (JSPS), and the Transdisciplinary Research Integration Center (TRIC) of the Research Organization of Information and Systems. The ice core drilling was conducted as part of the Ili Projects (R-03, 2006–2012) funded by the Research Institute for Humanity and Nature (RIHN), Japan. This work was supported by Grants-in-Aid for Scientific Research (Innovative Area “Genome Science” 221S0002) from the JSPS, and computations were partially performed on the NIG (National Institute of Genetics) supercomputer

at ROIS (Research Organization of Information and Systems) National Institute of Genetics.

Author contributions T.S., T.Y., N.T. and E.W. conceived the ideas; K.F., V.A. and N.T. collected ice core samples; T.S. supplied the sequences; T.S. and T.Y. analyzed the data; T.S., T.Y. and N.T. wrote the manuscript; all authors gave final approval for publication.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

- Abe T, Inokuchi H, Yamada Y, Muto A, Iwasaki Y, Ikemura T (2014) tRNADB-CE: tRNA gene database well-timed in the era of big sequence data. *Front Genet* 5:114
- Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML et al. (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc Biol Sci* 279:4724–4733
- Anesio AM, Sattler B, Foreman C, Telling J, Hodson A, Tranter M et al. (2010) Carbon fluxes through bacterial communities on glacier surfaces. *Ann Glaciol* 51(56):32–40
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 29(9):2157–2167
- Bahl J, Lau MCY, Smith GJD, Vijaykrishna D, Cary SC, Lacap DC et al. (2011) Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nat Commun* 2:163
- Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1):37–48
- Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. *Nat Rev Genet* 14(12):827–839
- Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DGJ et al. (2015) metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour* 15(6):1403–1414
- Benton MJ, Donoghue PCJ, Asher RJ (2009) Calibrating and constraining molecular clocks. In: Hedges SB, Kumar S (eds) *The Timetree of Life*. Oxford University Press, Oxford, UK, p 35–86
- Bulat SA, Alekhina IA, Blot M, Petit J-R, de Angelis M, Wagenbach D et al. (2004) DNA signature of thermophilic bacteria from the aged accretion ice of Lake Vostok, Antarctica: implications for searching for life in extreme icy environments. *Int J Astrobiol* 3(01):1–12
- Chen Y, Li X-K, Si J, Wu G-J, Tian L-D, Xiang S-R (2016) Changes of the bacterial abundance and communities in shallow ice cores from dunde and muztagata glaciers, Western China. *Front Microbiol* 7:1716
- Christmas NAM, Anesio A, Sanchez-Baracaldo P (2015) Multiple adaptations to polar and alpine environments within cyanobacteria: a phylogenomic and Bayesian approach. *Front Microbiol* 6:1070
- Christner BC, Royston-Bishop G, Foreman CM, Arnold BR, Tranter M, Welch KA et al. (2006) Limnological conditions in Subglacial Lake Vostok, Antarctica. *Limnol Oceanogr* 51(6):2485–2501
- Cook J, Edwards A, Takeuchi N, Irvine-Fynn T (2016) Cryoconite. *Prog Phys Geogr* 40(1):66–111
- Der Sarkissian C, Ermini L, Schubert M, Yang Melinda A, Librado P, Fumagalli M et al. (2015) Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Curr Biol* 25(19):2577–2583
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Mol Biol Evol* 22(5):1185–1192
- Dvořák P, Hašler P, Poulíčková A (2012) Phylogeography of the *Microcoleus vaginatus* (Cyanobacteria) from Three Continents – A Spatial and Temporal Characterization. *PLoS ONE* 7(6):e40153
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16):2194–2200
- Edwards A, Mur LAJ, Girdwood SE, Anesio AM, Stibal M, Rassner SME et al. (2014) Coupled cryoconite ecosystem structure–function relationships are revealed by comparing bacterial communities in alpine and Arctic glaciers. *FEMS Microbiol Ecol* 89(2):222–237
- Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J et al. (2011) Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet* 43(5):482–486
- Fujita K, Takeuchi N, Nikitin SA, Surazakov AB, Okamoto S, Aizen VB et al. (2011) Favorable climatic regime for maintaining the present-day geometry of the Gregoriev Glacier, Inner Tien Shan. *Cryosphere* 5(3):539–549
- Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princeton Univ, Princeton
- Griffiths RC (2007). GENETREE version 9.0 <http://www.stats.ox.ac.uk/~griff/software.html>
- Harding T, Jungblut AD, Lovejoy C, Vincent WF (2011) Microbes in high arctic snow and implications for the cold biosphere. *Appl Environ Microbiol* 77(10):3234–3243
- Ho SY, Kolokotronis SO, Allaby RG (2007) Elevated substitution rates estimated from ancient DNA sequences. *Biol Lett* 3(6):702–705
- Ho SY, Larson G (2006) Molecular clocks: when times are a-changin'. *Trends Genet* 22(2):79–83
- Ho SY, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 22(7):1561–1568
- Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD et al. (2008) Genome of an endosymbiont coupling N₂ fixation to cellulolysis within protist cells in termite gut. *Science* 322(5904):1108–1109
- Isenbarger TA, Finney M, Rios-Velazquez C, Handelsman J, Ruvkun G (2008) Miniprimer PCR, a new lens for viewing the microbial world. *Appl Environ Microbiol* 74(3):840–849
- Jin H, Yonezawa T, Zhong Y, Kishino H, Hasegawa M (2016) Cretaceous origin of giant rhinoceros beetles (Dynastini; Coleoptera) and correlation of their evolution with the Pangean breakup. *Genes Genet Syst* 91(4):209–215
- Jungblut AD, Lovejoy C, Vincent WF (2010) Global distribution of cyanobacterial ecotypes in the cold biosphere. *ISME J* 4(2):191–202
- Kaštovská K, Elster J, Stibal M, Šantrůčková H (2005) Microbial Assemblages in Soil Microbial Succession After Glacial Retreat in Svalbard (High Arctic). *Microb Ecol* 50:396–407
- Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26(15):1899–1900
- Kohshima S, Takeuchi N, Uetake J, Shiraiwa T, Uemura R, Yoshida N et al. (2007) Estimation of net accumulation rate at a Patagonian glacier by ice core analyses using snow algae. *Glob Planet Chang* 59(1–4):236–244
- Lane D, Pace B, Olsen G, Stahl D, Sogin M, Pace N (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA* 82:6955–6959

- Lane N (2005) *Power, Sex, Suicide. Mitochondria and the Meaning of Life*. Oxford University Press, Oxford
- Langford H, Hodson A, Banwart S, Boggild C (2010) The micro-structure and biogeochemistry of Arctic cryoconite granules. *Ann Glaciol* 51(56):87–94
- Langford HJ, Irvine-Fynn TDL, Edwards A, Banwart SA, Hodson AJ (2014) A spatial investigation of the environmental controls over cryoconite aggregation on Longyearbreen glacier, Svalbard. *Biogeosciences* 11(19):5365–5380
- Lepère C, Wilmotte A, Meyer B (2000) Molecular Diversity of *Microcystis* Strains (Cyanophyceae, Chroococcales) Based on 16S rDNA Sequences. *Syst Geogr Plants* 70(2):275–283
- MacLean RC, Torres-Barcelo C, Moxon R (2013) Evaluating evolutionary models of stress-induced mutagenesis in bacteria. *Nat Rev Genet* 14(3):221–227
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12
- Molak M, Lorenzen ED, Shapiro B, Ho SY (2013) Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. *Mol Biol Evol* 30(2):253–262
- Nadeau T-L, Castenholz RW (2000) Characterization of psychrophilic oscillatoriids (Cyanobacteria) from Antarctic meltwater ponds. *J Phycol* 36:914–923
- Nutman AP, Bennett VC, Friend CRL, Van Kranendonk MJ, Chivas AR (2016) Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* 537(7621):535–538
- Ohta T, Tachida H (1990) Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* 126(1):219–229
- Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M et al. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499(7456):74–78
- Schierup MH, Wiuf C (2010) The Coalescent of Bacterial Populations. In: Robinson DA, Falush D and Feil EJ (eds) *Bacterial Population Genetics in Infectious Disease*. John Wiley & Sons: Hoboken, NJ.
- Schirmmeister BE, de Vos JM, Antonelli A, Bagheri HC (2013) Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proc Natl Acad Sci USA* 110(5):1791–1796
- Schirmmeister BE, Gugger M, Donoghue PCJ (2015) Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* 58(5):769–785
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75(23):7537–7541
- Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863–864
- Segawa T, Ishii S, Ohte N, Akiyoshi A, Yamada A, Maruyama F et al. (2014) The nitrogen cycle in cryoconites: naturally occurring nitrification-denitrification granules on a glacier. *Environ Microbiol* 16(10):3250–3262
- Segawa T, Takeuchi N (2010) Cyanobacterial communities on Qiya glacier, Qilian Shan China. *Ann Glaciol* 51(56):135–144
- Segawa T, Takeuchi N, Rivera A, Yamada A, Yoshimura Y, Barcaza G et al. (2013) Distribution of antibiotic resistance genes in glacier environments. *Environ Microbiol Rep* 5(1):127–134
- Segawa T, Yonezawa T, Edwards A, Akiyoshi A, Tanaka S, Uetake J et al. (2017) Biogeography of cryoconite forming cyanobacteria on polar and Asian glaciers. *J Biogeogr* 00:1–13. <https://doi.org/10.1111/jbi.13089>
- Shimada R, Takeuchi N, Aoki T (2016) Inter-annual and geographical variations in the extent of bare ice and dark ice on the Greenland ice sheet derived from MODIS satellite images. *Front Earth Sci* 4:43
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690
- Takeuchi N, Fujita K, Aizen VB, Narama C, Yokoyama Y, Okamoto S et al. (2014) The disappearance of glaciers in the Tien Shan Mountains in Central Asia at the end of Pleistocene. *Quat Sci Rev* 103:26–33
- Takeuchi N, Fujita K, Nakazawa F, Nakawo M, Rana B (2009) A snow algal community on the surface and in an ice core of Rikha-Samba Glacier in Western Nepali Himalayas. *Bull Glaciol Res* 27:25–35
- Takeuchi N, Kohshima S, Seko K (2001) Structure, formation, darkening process of albedo reducing material (cryoconite) on a Himalayan glacier: a granular algal mat growing on the glacier. *Arct Antarct Alp Res* 33:115–122
- Takeuchi N, Li Z (2008) Characteristics of Surface Dust on Ürümqi Glacier No. 1 in the Tien Shan Mountains, China. *Arct Antarct Alp Res* 40(4):744–750
- Takuno S, Kado T, Sugino RP, Nakhleh L, Innan H (2012) Population genomics in bacteria: a case study of *Staphylococcus aureus*. *Mol Biol Evol* 29(2):797–809
- Uetake J, Kohshima S, Nakazawa F, Suzuki K, Kohno M, Kameda T et al. (2006) Biological ice-core analysis of Sofiyskiy glacier in the Russian Altai. *Ann Glaciol* 43:70–78
- Uetake J, Tanaka S, Segawa T, Takeuchi N, Nagatsuka N, Motoyama H et al. (2016) Microbial community variation in cryoconite granules on Qaanaaq Glacier, NW Greenland. *FEMS Microbiol Ecol* 92(9):fiw127
- Vincent WF (2000) Evolutionary origins of Antarctic microbiota: invasion, selection and endemism. *Antarct Sci* 12(03):374–385
- Vincent WF (2007) Cold tolerance in cyanobacteria and life in the cryosphere. In: J. S (ed) *Algae and Cyanobacteria in Extreme Environments*. Springer: Heidelberg, pp 287–301.
- Wientjes IGM, Oerlemans J (2010) An explanation for the dark region in the western melt zone of the Greenland ice sheet. *Cryosphere* 4(3):261–268
- Willerslev E, Cappellini E, Boomsma W, Nielsen R, Hebsgaard MB, Brand TB et al. (2007) Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland. *Science* 317(5834):111–114
- Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, Shapiro B et al. (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300(5620):791–795
- Willerslev E, Hansen AJ, Ronn R, Brand TB, Barnes I, Wiuf C et al. (2004) Long-term persistence of bacterial DNA. *Curr Biol* 14(1):R9–R10
- Yallop ML, Anesio AM, Perkins RG, Cook J, Telling J, Fagan D et al. (2012) Photophysiology and albedo-changing potential of the ice algal community on the surface of the Greenland ice sheet. *ISME J* 6(12):2302–2313
- Yonezawa T, Hasegawa M (2014) Extreme nearly neutral evolution in mitochondrial genomes of laboratory mouse strains. *Gene* 534(2):444–448