Fottea, Olomouc, 19(2): 115–131, 2019
DOI: 10.5507/fot.2019.003

Evaluating amplicon high–throughput sequencing data of microalgae living in melting snow: improvements and limitations

Stefanie Lutz[1*], Lenka Procházková[2*], Liane G. Benning[1, 3,4], Linda Nedbalová[2,5] & Daniel Remias[6]

[1]*GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany; *Corresponding author e–mail: stefanie.lutz@agroscope.admin.ch; current address: Agroscope, Müller-Thurgau-Strasse 29, 8820 Wädenswil, Switzerland*
[2] *Department of Ecology, Faculty of Science, Charles University, Viničná 7, 128 44 Prague 2, Czech Republic; *Corresponding author e–mail: lenkacerven@gmail.com*
[3] *School of Earth & Environment, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK*
[4] *Department of Earth Sciences, Free University of Berlin, 12249 Berlin, Germany*
[5] *The Czech Academy of Sciences, Institute of Botany, Dukelská 135, 379 82 Třeboň, Czech Republic*
[6] *University of Applied Sciences Upper Austria, Stelzhamerstr. 23, 4600 Wels, Austria*

Abstract: Melting snowfields are dominated by closely related green algae. Although microscopy–based classification are evaluable distinction tools, they can be challenging and may not reveal the diversity. High–throughput sequencing (HTS) allows for a comprehensive community evaluation but has been rarely used in such ecosystems. We found that assigning taxonomy to DNA sequences strongly depends on the quality of the reference databases. Furthermore, for an accurate identification, a combination of manual inspection of automated assignments, and oligotyping of the abundant 18S OTUs and ITS2 secondary structure analyses were needed. The use of one marker can be misleading because of low variability (18S) or the scarcity of references (ITS2). Our evaluation reveals that HTS outputs need to be thoroughly checked when the organisms are poorly represented in databases. We recommend an optimized workflow including consistent sampling, a two–molecular marker approach, light microscopy–based guidance, generation of appropriate reference sequences and a final manual verification of taxonomic assignments as a best approach for accurate diversity analyses.

Key words: 18S rDNA, ITS2 rDNA, high–throughput sequencing, Illumina, oligotyping, OTU clustering, red snow, Sanger, secondary structure, snow algae

Introduction

In alpine and polar regions, psychrophilic microalgae can cause a distinct colouration of melting snow from green to different shades of yellow, orange and red (Anesio et al. 2017; Hoham & Duval 2001; Kol 1968; Komárek & Nedbalová 2007; Leya 2013; Lutz et al. 2016). Snow algae have evolved a range of adaptive strategies to overcome a multitude of environmental stresses including low temperatures, freezing, desiccation, nutrient scarcity and extreme irradiation. Thus, they are of general interest to study a wide range of fundamental cellular processes. These eukaryotic photoautotrophs mostly belong to the Chlamydomonadaceae (Chlorophyceae) and their carotenoid–rich immotile stages (cysts), which are adapted to harsh conditions, are predominately found throughout the melt season (Remias et al. 2010).

Traditionally, the red snow phenomenon has been associated with *Chlamydomonas nivalis* (F.A. Bauer) Wille (Kol 1968). Yet, a plethora of further species can be found in melting snow including *Chlainomonas* sp. Christen (Remias et al. 2016), *Chloromonas nivalis* (Chodat) Hoham et Mullet (Procházková et al. 2018a; Remias et al. 2010) or *Chloromonas brevispina* (F.E. Fritsch) Hoham, Roemer et Mullet (Matsuzaki et al. 2015). Nonetheless, we are only scratching the surface of snow algal diversity characterization and the above–mentioned species likely constitute a small proportion of the true diversity. For many taxa, no strains are available because the germination of cysts collected in the field was not successful (Procházková et al. 2018a; own observations). Moreover, cells of one species transgress through a variety of morphological and physiologic changes during their life cycle. This poses another challenge for the microscopy–based identification and classification.

In contrast, high–throughput sequencing (HTS) allows for a comprehensive assessment of the microbial

community composition of a natural ecosystem. With its broad application in many other environments (Grossmann et al. 2016), it is striking how rarely such approaches have been used for psychrophilic algae. So far a few studies have targeted snow algal communities in the Arctic (Lutz et al. 2015a; Lutz et al. 2015b; Lutz et al. 2016, 2017; Segawa et al. 2018), in the US (Brown et al. 2016), in Japan (Terashima et al. 2017) and in Antarctica (Segawa et al. 2018). However, HTS data on European Alpine communities is completely absent in the literature.

The nature of HTS to produce large datasets in a mostly automated way comes at a cost; i.e., some of the data processing steps are to a certain degree a 'black box'. In addition, several technical biases must be taken in account. These may include effects due to extreme GC/AT ratios (Oyola et al. 2012), the choice of primers and library preparation methods (Schirmer et al. 2015), annealing temperature (Schmidt et al. 2013), DNA polymerase (Brandariz–Fontes et al. 2015), amplicon size variability (Schirmer et al. 2015) or from the sequencing technology itself (Schloss et al. 2011).

Several DNA markers, which are being employed for algal species delimitation, have been summarized in Leliaert et al. (2014). Comprehensive 18S rDNA marker reference databases such as Silva (Quast et al. 2012), PR2 (Guillou et al. 2013) or EukRef (Campo et al. 2018) exist. However, the 18S rDNA marker is not sufficiently variable to distinguish among closely related taxa (Hall et al. 2010). In contrast, the internal nuclear rDNA transcribed spacer 2 (ITS2 rDNA) inherits high taxonomic resolution, but also in some cases high intragenomic variation (Thornhill et al. 2007; Simon & Weiss 2008; Alanagreh et al. 2017), which may affect OTU (Operational Taxonomic Unit) clustering and taxonomic identification. Since the rRNA cassette can vary in copy numbers per organism and the ITS regions are free to independently drift within the same organism, a potential overestimation of OTUs may occur, especially in some fungal species (i.e., one species can split into several OTUs (Lindner & Banik 2011; Lindner et al. 2013). Nevertheless, intergenic variation of ITS2 in algae is generally considered low compared to fungi. Moreover, there is a dearth of appropriate ITS2 rDNA references sequences (Yao et al. 2010; Buchheim et al. 2011) that can be used for algae. Hence, for any accurate diversity evaluation of cryophilic algae diversity in environmental samples at least a two–marker approach is advisable (Chase & Fay 2009).

In a methodologically motivated approach, we evaluated the application of HTS for the characterization of snow algal communities in the extreme habitat of melting European Alpine snowfields. We present a case study that aims to improve the application of HTS techniques for accurate diversity assessments in such 'less common' and 'less well–studied' ecosystems. To do this we (1) investigated the suitability of the two markers 18S and ITS2 rDNA for amplicon high–throughput sequencing; (2) evaluated the importance of completion and curation of reference databases for correct taxonomic assignments of environmental sequences; (3) complemented our HTS data with traditional Sanger sequencing data to gain more and longer reference sequences; (4) cross–correlated the sequencing data with traditional microscopic observations; (5) tested different strategies for taxonomic assignments (QIIME, Blast and a final manual refinement) in order to reveal potential differences; and (6) to delineate cryptic diversity of dominant species, we performed oligotyping of the most abundant 18S rDNA OTUs and assessed species boundaries by ITS2 rDNA transcript secondary structures comparison.

Material and Methods

The overall workflow we followed in this study is shown in Fig. 1.

Field work and sample preparation. The samples were collected from a non–permanent, flat snow field in the Kühtai region of the Tyrolean Alps in Austria (Table 1). The site was dominated by an alpine meadow covered by a melting snow pack with characteristic reddish surface coloration. For HTS, two field samples (sample 1, sample 2) containing mixed communities of several snow algae were harvested in the summers of 2015 and 2016 (Table 1). For Sanger sequencing cells from virtually monospecific patches were collected and identified by light microscopy. The Sanger samples were each dominated by one of the locally abundant taxa: *Chlamydomonas nivalis* (sample 3 collected in 2016 and described in Procházková et al. 2018b), *Scotiella cryophila* (sample 4 collected in 2009 and described in Remias et al. 2018) and *Chloromonas brevispina* (sample 5; collected in 2016 and described here). The 2016 Sanger and HTS samples (samples 1, 2 and 5; Table 1) were collected about two to three weeks earlier in the melting season than sample 1 collected in 2015 or sample 4 collected in 2009. Cell harvest was performed as previously described by Procházková et al. (2018a) using a sterilized stainless steel shovel, putting the snow into sterile sampling bags and keeping it cold/frozen until returning to the laboratories for microscopic analyses and DNA extractions. The presence of algae and the species composition were evaluated using an Evolution field microscope (Pyser SGI, USA). For HTS, we intentionally used two different sampling approaches: In 2015 (sample 1) sampling included the complete snow column from the snow surface to the soil layer (approximately 30 cm; Fig. S1), whereas in 2016 (sample 2) surface and ground snow were not harvested (Fig. S2) to avoid allochthonous organisms (airborne and soil algae can occur in snow, Stibal & Elster 2005).

Light Microscopy. Cells were analyzed in the laboratory with a Nikon Eclipse 80i light microscope equipped with a Plan Fluor 1.3 100× objective and a Nikon DS–5M digital camera.

Sanger sequencing of locally abundant taxa. 18S rDNA and ITS2 rDNA sequences of *Chloromonas brevispina* K–2 (sample 5) were gained in course of this study, whereas Sanger sequences from two other species were recently published – *Chlamydomonas nivalis* DL07 (sample 3; Procházková et al. 2018b) and *Scotiella cryophila* K–1 (sample 4; Remias et al. 2018). These three sequences were used for the generation of a custom reference sequence database and are available at

Table 1. Overview of snow algae samples from Kühtai in the Austrian Alps with three locally abundant taxa (causing virtually monospecies bloom) harvested for Sanger sequencing and two samples of mixed cryoflora communities collected for HTS. Sample number (code), collection date, sampling altitude (m a.s.l.) and geographic position (GPS) are shown.

Sample (code)	Date	Altitude (m)	GPS	Sequencing	Reference
Sample 1 (WP79)	11.06.2015	2300	N47°13.422 E11°01.310	HTS	this study
Sample 2 (WP99)	31.05.2016	2299	N47°13.416 E11°01.260	HTS	this study
Sample 3 (DL07)	28.05.2016	2380	N47°13.709 E11°00.949	Sanger	Procházková et al. 2018b
Sample 4 (K–1)	05.06.2009	2432	N47°13.748 E11°00.704	Sanger	Remias et al. 2018
Sample 5 (K–2)	28.05.2016	2430	N47°13.753 E11°00.737	Sanger	this study

Table 2. Overview of HTS data. Number of 18S rDNA and ITS2 sequences before and after quality filtering, as well as the number and percentage of sequences assigned to green algal taxa (the remaining sequences were mostly assigned to Fungi, Alveolata and Rhizaria; data not shown).

Marker	Samples	No. of sequences before quality filtering	No. of sequences after quality filtering	No. and percentage of sequences assigned to green algae
18S	Sample 1	221837	184921	88795 (48.0%)
	Sample 2	294484	248269	119722 (48.2%)
ITS2	Sample 1	187543	116446	81022 (69.6%)
	Sample 2	204224	156958	108889 (69.4%)

NCBI under the accession numbers listed in Table S3. Total genomic DNA was isolated from the *Chlamydomonas nivalis* dominated sample 3 with the DNeasy Plant Mini kit (Quiagen) as previously described (Procházková et al. 2018a). Sample 55, containing *Chloromonas brevispina* was lower in biomass (<20 mg wet weight), and thus, DNA was extracted using the Instagene Matrix (Bio–Rad Laboratories, USA) followng the protocols described in Remias et al. (2016). Isolated DNA was diluted to a concentration of 5 ng.µl^{-1} and the 18S and ITS2 rDNA regions were amplified using existing primers (Table S1). Polymerase chain reactions (PCR) were performed according to Procházková et al. (2018a). The PCR products were stained with bromophenol loading dye, quantified on a 1.5% agarose gel and stained with GelRed (Biotium). The amplification products were purified and sequenced on the Applied Biosystems automated sequencer (ABI 3730×l) at Macrogen (Netherlands). Chromatogram data of forward and reverse sequences of both markers were visually inspected and edited in the program FinchTV 1.4.0 (Geospiza, USA). The contig of each marker was assembled in SeqMan 5.06 (DNASTAR Inc., USA).

High–throughput sequencing. DNA was extracted from both field samples (sample 1 and 2) using the PowerSoil® DNA Isolation kit (MoBio Laboratories). The 18S and ITS2 rDNA amplicons were prepared according to the Illumina "16S Metagenomic Sequencing Library Preparation" guide (Illumina). In brief, 18S rDNA genes were amplified using the eukaryotic primers 528F (5' GCGGTAATTCCAGCTCCAA) and 706R (5' AATCCRAGAATTTCACCTCT; Cheung et al. 2010) spanning the V4–V5 hypervariable regions. ITS2 rDNA genes were amplified using the primers 5.8SbF (5' GATGAAGAACGCAGCG; Mikhailyuk et al. 2008) and ITS4R (5' TCCTCCGCTTATTGATATGC; White et al. 1990). All primers were tagged with the Illumina adapter sequences. PCR was performed using KAPA HiFi HotStart ReadyMix. Initial denaturation at 95 °C for 3 min was followed by 25 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and elongation at 72 °C for 30 s. Final elongation was at 72 °C for 5 min. All PCRs were carried out in reaction volumes of 25 µl containing 12.5 µl of ReadyMix, each 5 µl of the forward and reverse primer and 12.5 ng of DNA template in 2.5 µl. All pre–amplification steps were done in a laminar flow hood with DNA–free certified plastic ware and filter tips. Amplicons were barcoded using the Nextera XT Index kit. The pooled library was sequenced on the Illumina MiSeq using paired 300 bp (base pairs) reads at the University of Bristol Genomics Facility. 18S and ITS2 rDNA raw sequences have been deposited to the European Nucleotide Archive (ENA) under accession number PRJEB24479.

Quality filtering of HTS sequences and ITS2 extraction. The sequencing quality of each de–multiplexed fastq file was analyzed using the FastQC software (http://www.bioinformatics.

babraham.ac.uk/projects/fastqc/). The low quality 3' ends of all reads were trimmed. All forward reads were trimmed by 20 bp and all reverse reads by 100 bp. All other processing steps were performed in Qiime (CAPORASO et al. 2010). The trimmed paired end reads were joined before further processing and additionally filtered only allowing a minimum Phred quality score of Q20. Reads that could not be joined or were below the quality cut–off were excluded from the analysis. Chimeric sequences were removed using USEARCH 6.1.

The software ITSx (BENGTSSON–PALME et al. 2013) was used to extract the ITS2 rDNA regions from all sequences to avoid the inclusion of the highly conserved neighbouring genes (i.e., 5.8S and 28S). Inclusion of these regions in the identification process would otherwise lead to misleading results. HMMER (EDDY 1996) was used to predict the origin of the sequences (e.g., Chlorophyta, Fungi) based on Hidden Markov Models.

Clustering sequences into OTUs and creation of OTU table. In general, fragments of 18S rDNA sequences of phytoplankton assemblages and prokaryotic and eukaryotic alpine permafrost communities are clustered into OTUs at 97% similarity in HTS studies (FREY et al. 2016; TRAGIN et al. 2017). A far stricter threshold for clustering and species assignment is required for this marker when snow algal communities dominated by Chlamydomonadales are investigated. Several species in this group are very closely related and they differ in some cases by only one bp over the length of the amplicon (e.g. *Chloromonas fukushimae* GsCl–11 (AB906342) and *Chloromonas tughillensis* UTEX SNO91 (AB906348)). Therefore, OTUs were picked *de novo* and clustered at 99.5% similarity for the conserved 18S rDNA marker.

In contrast, ITS2 is more variable and was thus clustered at 94.0% similarity. The chosen threshold is in par with several findings on the level of identity of algal ITS2 rDNA. In *Gonium pectorale* less than 5% of the nucleotide positions differ in pairwise comparisons and less than 7% vary between all clones (COLEMAN et al. 1993). Similarly, only few nucleotide differences have been reported among strains of *Chloromonas reticulata* (3.4–4.1%), and of *Chlamydomonas reinhardtii* NIES–2463 and SAG 11–32a (3.3%), with the latter one being able to cross and produce zygotes (MATSUZAKI et al. 2015). A similar low identity threshold for OTU picking (95%) as in our case was also successfully applied during Illumina barcoding of soil fungal communities (SCHMIDT et al. 2013).

Singletons (OTUs containing only 1 sequence, likely derived from sequencing errors) were removed from both the 18S and ITS2 rDNA data sets prior to further analysis. The OTU tables were created by counting the number of times an OTU appeared in each sample and adding the taxonomic predictions to each OTU.

Identification of OTUs. The objective of this final process was to define species boundaries. In order to do so, the representative sequences (i.e., the cluster seeds in the OTU picking process) of the 50 most abundant 18S and ITS2 rDNA OTUs (comprising >85% and >98% of the total community, respectively) were used.

Three different strategies for taxonomic assignments of each OTU from environmental samples were tested:

Strategy A (basic version): The BLAST (KENT 2002) assignment method implemented in Qiime was used with the default minimum percent similarity of 90% to consider a database match a hit (unless a customized script is being used to overwrite the default setting and to increase the similarity threshold). The publicly available and Qiime–compatible Silva database (release 128) (QUAST et al. 2012) was used for the assignment of the 18S rDNA data set and extended, with 223 additional sequences of psychrophilic algae kindly provided by Dr. Thomas Leya from the CCCryo – Culture Collection of

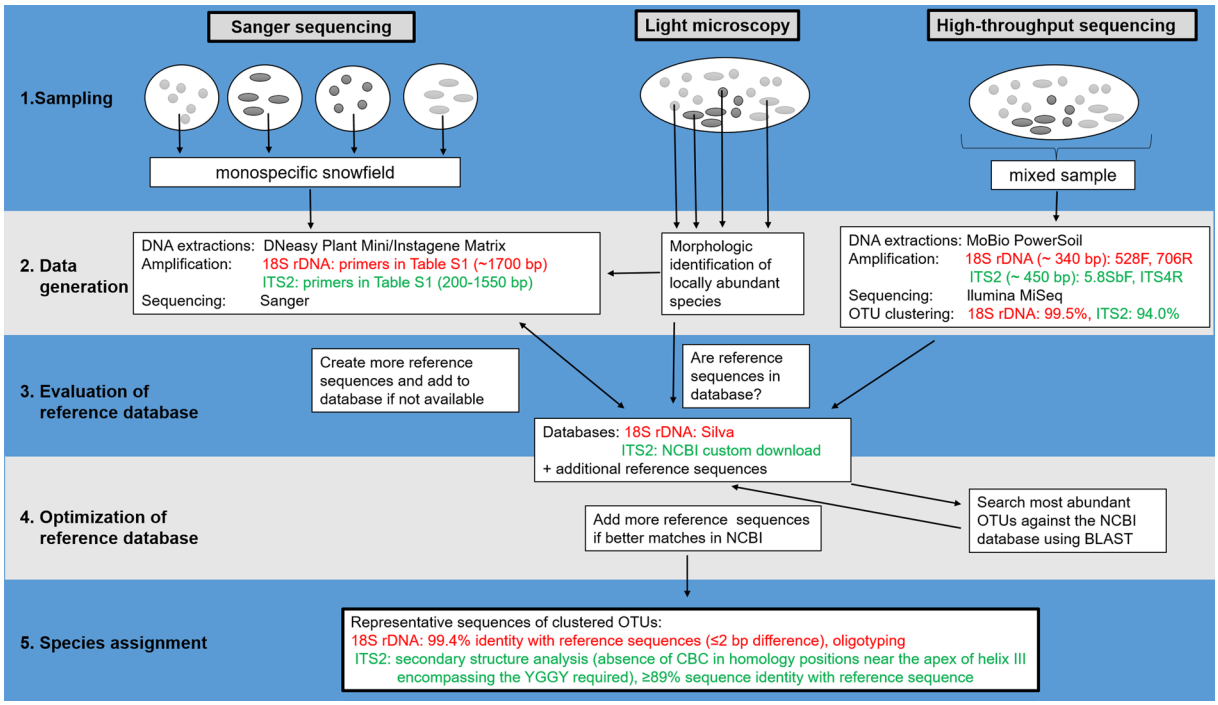

Fig. 1. Schematic workflow for the optimised molecular evaluation of snow algal community structures and diversity, which heavily relies on the combined power of light microscopy, Sanger sequencing and amplicon high–throughput sequencing. All details on the markers 18S rDNA and ITS2 are highlighted in red and green, respectively.

Cryophilic Algae (Fraunhofer IZI–BB). Sequences assigned to Opisthokonta, Amoebozoa, Alveolata and Rhizaria were removed from the OTU table. For the taxonomic assignment of the ITS2 rDNA sequences, a custom database with the limited number of available reference sequences for (psychrophilic) green algae was downloaded from NCBI (Table S2).

Strategy B (extended version): The basic version was improved by adding reference sequences of the locally abundant taxa derived from Sanger sequencing (see above) to the custom reference database.

Strategy C (further extended version): All the steps were done as in the extended version. Additionally, manual comparisons were carried out for the representative sequences of the OTUs with their respective reference sequences (pairwise blast). A manual search of each OTU representative sequence against NCBI was performed (megablast). Additional reference sequences were added to the custom based database, if the search resulted in a better sequence identity match than the one with its respective reference sequence. A verification of sequence identities for 18S (including oligotyping) and ITS2 (including secondary structures comparisons) was performed. Hereafter, we define these unique ITS2 sequences among one species as "haplotypes".

Manual identification of OTUs (Strategy C [further extended version] in detail). Representative sequences of the most abundant 18S and ITS2 OTUs were manually submitted to the BLAST (Kent 2002) web server to search NCBI for close hits to algal taxa. The used BLAST nucleotides parameters were the following: megablast (highly similar sequences), ´others´ as database search set, uncultured/environmental sequences were included, other algorithm parameters were kept with default values.

In case of 18S rDNA, an identity threshold of ~99.4% (i.e., 2 bp nucleotide difference in a 342 bp sequence) had to be passed in order to be considered as a database match. Sequences below this threshold were recorded as "no blast hit". A stricter identity threshold could inflate the diversity due to potential sequencing errors (Bradley et al. 2016).

To discover cryptic diversity in the 18S rDNA data, the three most abundant OTUs of this marker were further subjected to oligotyping, a high–resolution method that uses Shannon entropy to evaluate the most information–rich nucleotide position in an amplicon data set (Eren et al. 2013; Lutz et al. 2018). All sequences contained in one OTU were extracted individually and trimmed to the same length of 340 bp using Fastx Trimmer (http://hannonlab.cshl.edu/fastx_toolkit/). The number of components (i.e., nucleotide position with the highest entropy) to be used was chosen based on the entropy analysis of the sequence alignment. Noise filtering was carried out using a minimum substantive abundance of 50.

In order to assess species boundaries using ITS2 rDNA, three steps were carried out: (1) A minimum similarity

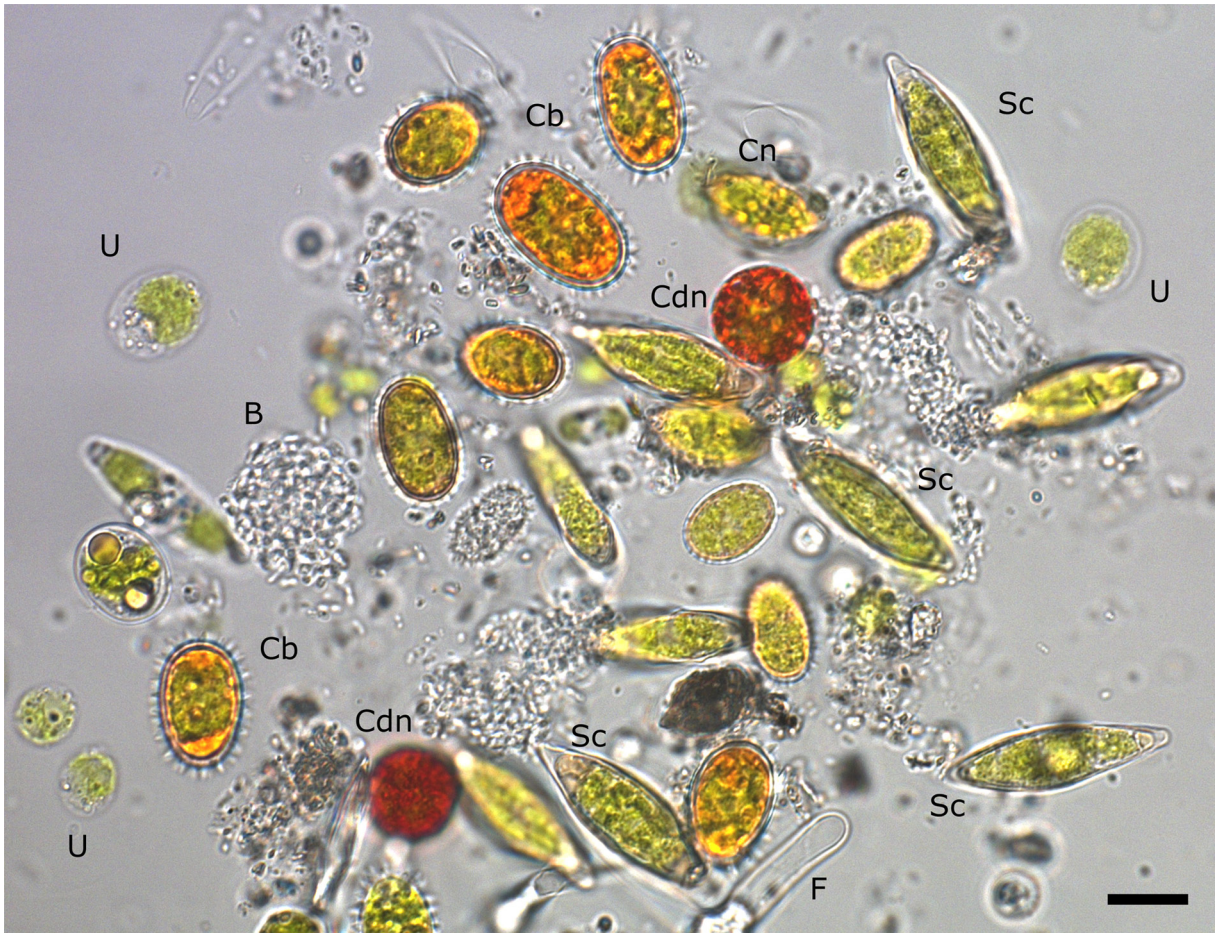

Fig. 2. Light micrograph of cells in field sample 1. The three locally abundant snow algae identified using morphological features were *Cr. brevispina* (Cb), *Scotiella cryophila* (Sc) and *Cd. nivalis* (Cdn). Other cells observed in this sample were *Cr. nivalis* (Cn), unknown unicellular green alga (U), fungus (F) or bacteria (B). Scale bar 10 μm.

Table 3. Algal community structure based on the 18S rDNA data set. The ten most abundant OTUs were selected (>78% of the community). The table shows the discrepancies between OTU assignments using three strategies: (A) basic version using Qiime and the Silva database, (B) extended version using Qiime and additional reference sequences of the locally abundant taxa (underlined) and (C) further extended version using final manual verification of taxa assignments at NCBI, only allowing up to 2 bp nucleotide difference to the respective reference sequence (sequences below this threshold were recorded as "no blast hit"). A comprehensive list of the 50 most abundant OTUs with corresponding OTU identification numbers can be found in Table S4.

OTU ID	Sample 1 (%)	Sample 2 (%)	(A) Qiime + Silva	(B) Qiime + Silva + local references	(C) Qiime + Silva + local references + manual verification
denovo14334	33.0	66.8	*Chloromonas* sp. Gassan–A LC012753.1	*Chloromonas brevispina* K–2	Ambiguous hits: *Chloromonas brevispina* K–2, *Scotiella cryophila* K–1, *Chloromonas sp.* TA AB902996, *Chloromonas* sp. Gassan–B LC012714.1
denovo45654	18.7	0.1	*Mesotaenium* sp. AG–2009–1 FM992335.1	*Ancylonema nordenskioeldii* AF514397.2	*Ancylonema nordenskioeldii* AF514397.2
denovo36485	0.9	13.6	Uncultured Chlamydo–monadaceae AB902971.1	*Chlamydomonas nivalis* DL07	*Chlamydomonas nivalis* DL07
denovo40226	8.2	0	*Botrydiopsis constricta* AJ579339.1	*Botrydiopsis constricta* AJ579339.1	*Botrydiopsis constricta* AJ579339.1
denovo15070	4.6	1.0	Uncultured *Chloromonas* AB903008.1	*Chloromonas* cf. *alpina* CCCryo 033–99 HQ404865.1	*Chloromonas platystigma* strain CCCryo 020–99
denovo20542	4.6	0	Uncultured Dunaliellaceae EF023287.1	Uncultured Dunaliellaceae EF023287.1	*Chloroidium saccharophilum* isolate HST10K KX024691.1
denovo101	3.2	1.1	*Chloromona*s sp. D–CU581C AF517086.1	*Chloromonas* cf. *rostafinskii* CCCryo 025–99 AF514402.1	Ambiguous hits: *Chloromonas* sp. NIES–2379 AB906350.1, *Chloromonas rostafinskii* strain CCCryo 025–99 AF514402.1
denovo23251	3.0	<0.1	*Chloromonas* sp. Gassan–A LC012753.1	*Chloromonas brevispina* K–2	Ambiguous hits: *Chloromonas brevispina* K–2, *Chloromonas* sp. Hakkoda–1 LC012710.1, *Chloromonas* sp. Gassan–A LC012709.1
denovo30051	0.4	1.9	Uncultured *Chloromonas* AB902984.1	Uncultured *Chloromonas* AB902984.1	No blast hit
denovo36086	2.1	0	*Prasiola furfuracea* AF189073.1	*Prasiola furfuracea* AF189073.1	No blast hit
	21.3	15.5	Other	Other	Other

Table 4. Oligotyping of the three most abundant 18S rDNA OTUs (sequences shown in Table S5). The table shows individual oligotypes that were conflated in the three most abundant 18S rDNA OTUs (Table 3), but were not be detected by conventional OTU clustering. The refined taxonomic assignments and their respective relative abundances were then used for the final description of the snow algal community composition in Figure 5. For instance, OTU 'denovo14334' was assigned to *Chloromonas brevispina* K–2 (Sample 1: 33%). However, oligotyping revealed that only one oligotype within this OTU corresponded to this species, which decreased its relative abundance from 33.0% to 23.4%.

OTU **Oligotype**	**Taxa assignment**	**Similarity (%)**	**Sample 1 (%)**	**Sample 2 (%)**
denovo14334	Ambiguous hits: *Chloromonas brevispina* K–2, *Scotiella cryophila* K–1, *Chloromonas sp.* TA AB902996, *Chloromonas* sp. Gassan–B LC012714.1	99.4	33.0	66.8
TTT	Uncultured snow algae LC371427.1, LC371425.1, LC371423.1, LC371419.1, LC371414.1	100	1.0	60.7
TCT	*Chloromonas brevispina* K–2, *Chloromonas* sp. Gassan–A LC012753.1 , *Chloromonas* sp. Hakkoda–1 LC012710.1	100	23.4	<0.1
CTT	*Chloromona*s sp. Gassan–B LC012714, uncultured *Chloromonas* sp. ANT1 AB903007.1 and *Chloromonas* sp. TA8 AB902996.1, *Chloromonas polyptera* JQ790556.1, uncultured Viridiplantae HQ188979.1	100	8.5	0.2
TTC	28 hits	>99	0.1	5.9
denovo45654	*Ancylonema nordenskioeldii* AF514397.2	100	18.7	0.1
A	*Ancylonema nordenskioeldii* AF514397.2	100	17.2	0.1
C	*Mesotaenium berggrenii* var. *alaskana* JF430424.1, *Mesotaenium* sp. AG–2009–1 FM992335.1	99.4	1.5	<0.1
denovo36485	*Chlamydomonas nivalis* DL07	100	0.9	13.6
T	*Chlamydomonas nivalis* DL07, 14 hits including several *Chlamydomonas nivalis* and uncultured snow algae strains	100	0.8	12.3
C	14 hits including several *Chlamydomonas nivalis* and uncultured snow algae strains	100	0.1	1.3

of ≥89.0% between an OTU and the reference sequence had to be passed to be considered as a database match. (2) If an OTU passed this identity threshold, it was retained for the creation of ITS2 rDNA transcript secondary structures. (3) The absence of a CBC in the homology positions of the ITS2 in comparison to an OTU with the reference sequence near the 5′– apex of helix III in the ITS2 secondary structure was required in order to be assigned to this reference taxon. A suggested schematic overview of taxonomic assignment of environmental ITS2 rDNA sequences from environmental samples is shown in Fig. S3. In detail, the ITS2 sequences were folded using the Mfold server (http://mfold.rna.albany.edu/?q5mfold; Zuker 2003; note: HTS delivers DNA based data, but during RNA folding, thymine ['T'] is converted to uracil ['U']). The model of the secondary structure with the minimum free energy that was consistent with the specific features of nuclear rDNA ITS2 and that contained four helixes and U–U mismatch in helix II (Coleman 2007) was selected. The ITS2 sequences and secondary structures were automatically and synchronously aligned (Schultz & Wolf 2009) using 4SALE (Seibel et al. 2006, 2008), and subsequently manually validated and corrected. First, structure based information (i.e., consensus of all secondary structures and all secondary structures displayed separately) was visually inspected to detect misaligned sequences. This was followed by the manual editing of the secondary structure in order to provide accurate sequence–structure alignments in the context sensitive editing mode (i.e., sequences and secondary structure information are used to validate whether a binding in the context is possible or not). The alignment consisted of the reference species and all OTUs assigned to this species based on preliminary pairwise comparisons in BLAST. Species delimitation was performed in 4SALE and was based on the detection of CBCs (Compensatory Base Changes), both nucleotides of a paired site mutate while the pairing itself stays stable (e.g., paired sites A–U mutated into G–C). A search for CBCs can only be performed in homologous positions of the ITS2 molecule, which can be unambiguously aligned. For Chlorophyceae, the consensus secondary structure model of ITS2 was identified and the conservation level of individual ITS2 sequence positions (i.e., a position is conserved above 70%) in the alignment was specified (Caisová et al. 2013). Comparisons of the ITS2 secondary structure prediction of the reference sequences gained from Sanger sequencing and those from the HTS data set that were preliminarily assigned to those reference sequences were performed. Based on these comparisons, species boundaries between OTUs and the number of haplotypes for each reference species was assessed. Even a single CBC in helices II and III of the ITS2 secondary structure may indicate sexual incompatibility as has been shown in crossing experiments

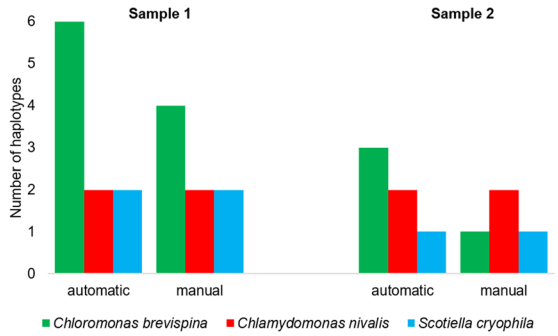

Fig. 3. Analysis of the 38 most abundant OTUs of both samples: Comparison of haplotype diversity of the three locally abundant species (*Chloromonas brevispina* K–2, *Chlamydomonas nivalis* DL07, *Scotiella cryophila* K–1; identified by light microscopy) between an evaluation based on automatic species assignment via Qiime and a manual assignment according to the CBC species concept (comparison of the ITS2 rDNA transcript secondary structures). A sequence identity of ≥89.0% and an absence of CBCs in homologic positions near the 5´apex of helix III were required in comparison to a reference species for an assignment to each taxon.

(COLEMAN 2000, 2009). One CBC in the most conserved part of helix III region of the ITS2 encompassing the YGGY motif (the most conserved region of the ITS2 secondary structures of eukaryotes; COLEMAN 2007) suggested the separation of two sister species which differ in their cell morphology, i.e., *Chloromonas reticulata* and *Chloromonas chlorococcoides* (MATSUZAKI et al. 2012). It has been shown that the probability of a CBC representing two distinct species is 93% (WOLF et al. 2013). The secondary structure of nuclear rDNA ITS2 was drawn using VARNA version 3.9 (DARTY et al. 2009).

Evaluation of the different strategies for taxonomic assignments. Nonmetric multidimensional scaling (NMDS) based on Bray–Curtis distances was performed using the program CANOCO 5 (TER BRAAK & ŠMILAUER 2012) to visualize the differences in taxonomic assignments based on 18S and ITS2 rDNA and among the three strategies used (A, B, C, see above).

RESULTS

Community composition based on light microscopy

The most abundant algal taxa identified in both samples 1 and 2 were *Chloromonas brevispina*, *Chlamydomonas nivalis* and *Scotiella cryophila.* All cells of these species were immotile cysts containing a secondary red carotenoid pigmentation, more or less masking the chlorophylls (Fig. 2, Fig. S4). The macroscopic appearance of the snow was red at the surface, turning greenish deeper in the snow at the spot where sample 1 was collected (Fig. S1) and yellowish where sample 2 was collected. Sample 1 additionally contained several other, unidentified unicellular green algae. The microscopic identification of the sample 5 revealed solely *Chloromonas brevispina.* The microscopic identifications of the dominant algae in samples 3 and 4 have been described previously (PROCHÁZKOVÁ et al. 2018b; REMIAS et al. 2018).

Output of the Sanger sequencing of the locally abundant taxa

Long sequences of multiple DNA regions containing 18S (about 1700 bp) and ITS2 rDNA (~200–1550 bp, Fig. 1) were obtained from the samples with virtually monospecific blooms of the three locally abundant taxa. The primers used (Table S1) amplified a larger fragment than ITS2. The actual length of the ITS2 rDNA for Chlorophyta (most common photosynthetic members of snow communities) varies between taxa in a range between 180 and 480 bp (BUCHHEIM et al. 2011). For instance, the AL1500af and LR3 primers (Table S1) are complementary to the end of 18S rDNA and 26S rDNA, and therefore resulting in the amplification of an approximately 1550 bp region.

Output of the 18S rDNA HTS data

A total of 433,190 18S rDNA sequences passed the quality control and 208,517 sequences could be assigned to green algal taxa (Table 2). The remainder of the sequences was assigned mostly to fungi, as well as Alveolata and Rhizaria (data not shown). 50 OTUs made up >87% of the total community composition (Table S4) and they were selected for the data evaluation and workflow optimization (Fig. 1). An overview of the ten most abundant OTUs (>78% of the total community) can be found in Table 3. The largest proportion of the sequences (sample 1: 33.0%, sample 2: 66.8%) was clustered in one OTU ‘denovo14334’ (99.4% similarity).

Evaluation of the different strategies for taxonomic assignments in 18S rDNA

Strategy A (basic version): The initial species assignment solely using the Qiime–compatible Silva database resulted in *Chloromona*s sp. Gassan–A LC012753.1 (Table 3 – column (1)). Other abundant OTUs were assigned to *Mesotaenium* sp. and several “uncultured *Chloromonas* and Chlamydomonadaceae*”* without a species affiliation (Table 3). Strategy B (extended version): The inclusion of the reference sequences of the locally abundant taxa into custom reference sequence database resulted in new assignments (26 out of the 50 most abundant OTUs) and in some cases in the clarification of the species assignment (Table S4 – column (2)). For instance, the “uncultured Chlamydomonadaceae” was identified as *Chlamydomonas nivalis* (sample 1: 0.9%, sample 2: 13.6%). Strategy C (further extended version): The representative sequences of the 50 most abundant OTUs in 18S rDNA were subjected to a manual BLAST search against NCBI GenBank (since the taxa assignments in Qiime (CAPORASO et al. 2010) uses a low default value of 90% minimum percent similarity to assign taxonomies to OTUs). The aim was to verify the actual percentage of identity, and whether closer hits not present in the Silva database occurred. Indeed, the manual verification step improved the taxonomic assignment of another eight taxa. However, it also revealed that 15 OTUs shared the same identity with several species. For

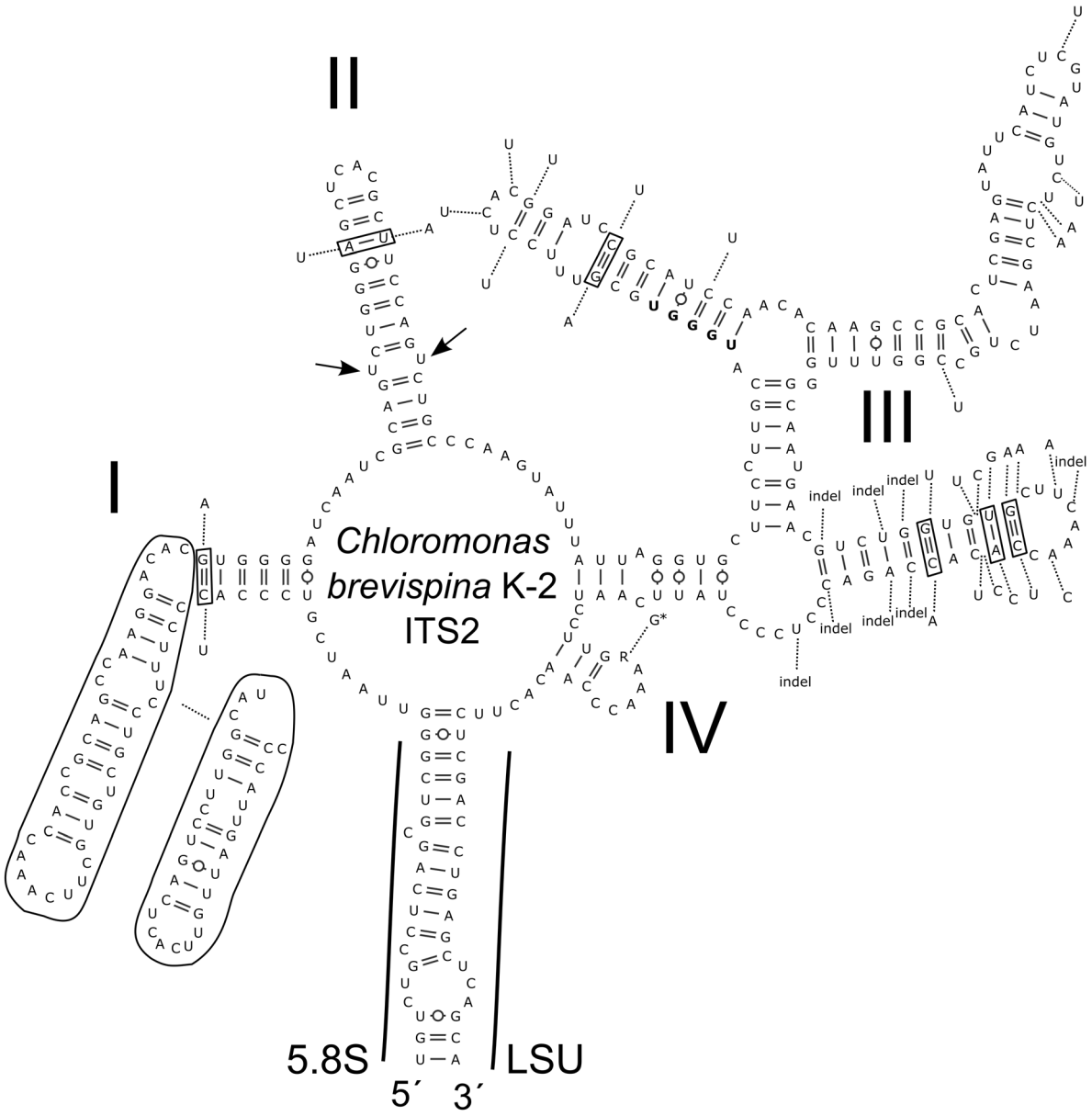

Fig. 4. Secondary structure of the ITS2 rDNA transcript *Chloromonas brevispina* K–2 (accession number MG791868). Differences between this species and the closely related OTU 'denovo99'are shown by nucleotides outside the structure linked by dotted lines. The U–U mismatch in helix II is indicated by arrows and the YGGY motif on the 5' side near the apex of helix III is in bold. CBCs in conserved parts of the structure are indicated by rectangles. The most significant CBC is located near the 5´apex of III helix. In addition, *Chloromonas brevispina* K–2 was identical with OTU 'denovo107' except for one ambiguous base (marked by an asterisk in helix IV).

instance, a vast number of *Chloromonas* species shared more than 99% identity in the hypervariable V3–V4 region of the 18S rDNA sequences (Table 3 and Table S4). This was the case for *Chloromonas brevispina* K–2, *Chloromonas* sp. TA 8 (AB902996.1), *Chloromonas* sp. Gassan–A (LC012714.1), *Chloromonas* sp. Gassan–B (LC012714.1), *Chloromonas polyptera* (JQ790556) and *Chloromonas* sp. Hakkoda–1 (LC012710.1), as well as *Scotiella cryophila* K–1 (MG253843; considering two ambiguous positions in the reference which can code for the same nucleotides). The same situation applied to *Raphidonema sempervirens* (AF514410.2), *Raphidonema nivale* (AB488604.1) and *Stichococcus* sp. (KP081395.1), which also shared more than 99% identity. Thus, those OTUs were not assigned unambiguously at one species level (see ambiguous hits in Tables 3 and Table S4). The most abundant OTU, denovo14334, showed a difference of 1 bp to *Chloromonas brevispina* K–2, *Chloromonas* sp. TA 8 (AB902996.1), *Chloromonas* sp. B (LC012714.1) and *Scotiella cryophila* K–1 (MG253843). Thus, several species are likely conflated in this OTU. In addition, several OTUs showed differences of more than 2 bp to the closest reference and their assignment was therefore discarded in this step and recorded as "no blast hit"

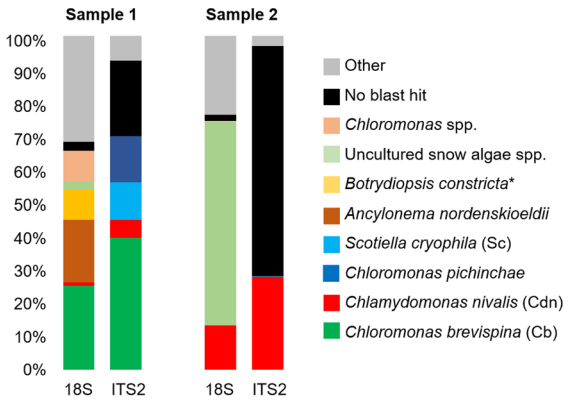

Fig. 5. Snow algal community composition after all optimization steps comprising all taxa with a minimum relative abundance of 5%. Full details can be found in Tables S4 and S6. Some discrepancies still prevailed between the 18S and ITS2 rDNA data sets, in particular in terms of the percentage of sequences with an unambiguous species assignment. Soil algae, which were only present in sample 1, are highlighted with an asterisk. Less abundant taxa are summarised in 'Other'. Species abbreviations in brackets are referring to cells identified by light microscopy. Sequences below the similarity threshold of 94.5% and of 89.0% for 18S rDNA and ITS2, respectively, were recorded as "no blast hit".

(2 OTUs in Table 3 – column (3), and 15 OTUs in Table S4 – column (3)). The percentage of sequence cover in the pairwise comparison with the reference sequence was also considered. For example, the OTU denovo30674 was initially assigned to *Prototheca cutis* (AB470468.1; < 2 bp difference). Yet, manual post–processing revealed that the sequence cover was only 18%. Thus, this assignment was discarded.

Furthermore, based on light microscopic observations and known species distribution patterns for this European Alpine region, some taxa assignments had to be scrutinized despite a very high similarity of the amplicon sequences with the suggested affiliations. These included *Ancylonema nordenskioeldii,* which mainly thrives on glacial surfaces of Polar Regions (LUTZ et al. 2018; REMIAS et al. 2012), and *Chloromonas polyptera,* which can be found in Maritime Antarctica near penguin rockeries (REMIAS et al. 2013).

Since several OTUs seemed to contain multiple taxa, oligotyping was carried out for further refinement of the taxonomic assignments. OTU 'denovo14334' consisted of four oligotypes, of which the most abundant shared 100% similarity with *Chloromonas brevispina* K–2 in sample 1 and several uncultured snow algae species in sample 2 (Table 4, Table S5). Oligotyping of the sequences grouped in the OTUs 'denovo45654' and 'denovo36485' also resulted in the resolution of another two oligotypes in each OTU.

Output of the ITS2 rDNA HTS data

A total of 273,404 ITS2 rDNA sequences passed the quality control and 189,922 sequences could be assigned to Chlorophyta (Table 2). The remainder of the sequences was assigned to mostly Fungi and Alveolata (Table S6). An overview of the 10 most abundant OTUs (>88% of the total community) can be found in Table 5. 38 OTUs made up >98% of the total community composition (Table S7) and were selected for the data evaluation and workflow optimization (Fig. 1, Fig. S3). All sequence–structure alignments of ITS2 transcripts and ITS2 secondary structures of these most abundant OTUs, together with their reference sequences, can be found in the Supplementary Material (Figs S5–S24).

Evaluation of the different strategies for taxonomic assignments in the ITS2

Strategy A (basic version): The initial assignments based on the limited number of sequences available at NCBI resulted in species assignments for 17 out of the 38 OTUs (Table S7 – column (1)). Strategy B (extended version): After including the Sanger derived reference sequences of the three locally abundant taxa into custom reference sequence database, eight assignments for denovo OTUs were improved (Table S7 – column (2)) and one OTU previously without blast hit could be newly assigned (i.e. denovo63 to *Scotiella cryophila* K–1). *Chloromonas brevispina* K–2 and *Scotiella cryophila* K–1 contributed significantly to the pool of detected sequences. Strategy C (further extended version): After conducting ITS2 rDNA transcript secondary structure analyses and CBC detection in the most conserved part close to 5′–apex of III helix, the haplotype diversity of locally abundant *Chloromonas brevispina* revealed to be lower than expected from the HTS output (Fig. 3 – column ′automatic′), whereas it did not change for *Scotiella cryophila* and *Chlamydomonas nivalis*. In the analysis of the 38 most abundant OTUs, which comprised >98% of the community (Table S7), one dominant and up to three rare haplotypes for each of three locally abundant species was recovered (Fig. 3 – column ′manual′, Table S7). The dominant one was shown to be 100% identical with the reference sequence of a locally abundant taxon (Fig. 4, Figs S11, S16). For instance, the dominant haplotype of *Scotiella cryophila* K–1 in sample 1 accounted for 8,853 reads, whereas the second haplotype comprised 194 reads. The major haplotype of *Chlamydomonas nivalis* DL07 in sample 2 included 30,734 reads and the second haplotype only 61 reads. The major haplotype of *Chloromonas brevispina* K–2 in sample 1 was represented by 32,258 reads and the second haplotype by 155 reads only. The most abundant OTU 'denovo99' (59% of sequences in sample 2), which had previously been assigned to *Chloromonas brevispina* K–2, had six CBCs in the conservative regions (including one CBC in the most conserved apex of helix III close to the 5′end) of the ITS2 structure in comparison with the reference sequence (Fig. 4). Considering that the presence of at least one CBC between two organisms in the most conserved regions of ITS2 is predicting a failure to cross sexually (COLEMAN 2009), we infer that OTU 'denovo99' might represent an independent though undescribed species. Furthermore, several OTU taxonomic assignments were below the chosen similarity threshold

of 89% with respect to the reference sequences. The suitability of the used ITS2 identity threshold was verified by checking the presence of CBCs in all representative OTUs, also those with a lower similarity threshold. These included the three OTUs (denovo142, denovo254 and denovo127) initially assigned to *Chloromonas* cf. *rostafinskii* (HQ404863.1). Indeed, they had several CBCs in the most conserved 5′end of helix III (Figs S12–S14). Consequently, these unknown species contributed to 22.5% of the unidentified ITS2 diversity in sample 1. Another OTU with an identity above the threshold of 89% was 'denovo100', which shared 92% identity with *Chloromonas pichinchae* CCCryo 261–06. A single CBC outside the most conserved part in helix III was detected in this case when the secondary structures of these sequences were compared (Figs S6, S7). Therefore, OTU 'denovo100' was also assigned to *Chloromonas pichinchae*. Similarly, four CBCs were found between OTU 'denovo44' and *Scotiella cryophila.* They were located in helix III but outside its most conserved part (Figs. S15, S16), therefore these CBCs were treated as intraspecific variability (and maybe as an intragenomic variability) and OTU 'denovo44' was assigned to *Scotiella cryophila*. In addition, OTU 'denovo85' was 95% identical with the reference species and no CBC was found. Thus, 'denovo85' was assigned to *Chloromonas pichinchae* (Fig. S6, S7).

Comparison of community compositions obtained with HTS using 18S and ITS2

The differences in the algal community structure obtained using the two markers is summarised in the NMDS ordination graphs (Figs. S25–S26). Whereas the 18S rDNA dataset resulted in a very similar taxonomic composition of both samples, ITS2 based analysis clearly separated the samples. The discrepant outcomes of the different assignment strategies are clearly visible for both markers. The presence of *Chloromonas brevispina* K–2 and *Chlamydomonas nivalis* DL07 were confirmed by both the 18S and the ITS2 rDNA data. Whereas *Scotiella cryophila* K–1 was one of the more abundant species in the ITS2 data in sample 1 (11.1%), it could not be unambiguously assigned in the 18S data (Table 3 and Table S4, denovo14334). Several other, less frequent species, (e.g, *Chloromonas* sp. Hakkoda–1 (LC012710.1), *Chloromonas platystigma* (AF514401.1), *Chloroidium saccharophilum (*KX024691.1), *Chloromonas* cf. *rostafinskii* (AF514402.1) could be detected in the 18S rDNA data, yet, were absent in the ITS2 data.

In contrast, the 18S rDNA data of sample 2 was dominated by a taxon sharing 100% similarity with several uncultured snow algal species, and *Chlamydomonas nivalis* (18S: 13.6%, ITS2: 28.3%; Tables 3 and 5, Figure 5). A considerably higher abundance of sequences with no species assignment was present in the ITS2 data sets (sample 1: 22.5%, sample 2: 68.6%) in comparison to the 18S rDNA data sets (sample 1: 3.6%, sample 2: 3.0%; Fig. 5). However, the vast majority of unassigned sequences in sample 2 was represented by a single dominant OTU (denovo99; 86% of all sequences without species assignment) closely related to *Chloromonas brevispina* K–2 (Fig. 4), and the abundance of *Chloromonas brevispina* K–2 was negligible (<10 reads). In contrast, in sample 1 *Chloromonas brevispina* K–2 represented the dominant abundant OTU and denovo99 was much less abundant. In sample 1, allochtonous soil algae like *Botrydiopsis constricta* (AJ579339.1)*, Botrydiopsis callosa* (AJ579340.1), *Heterococcus pleurococcoides/fuornensis/chodatii* (Xanthophyceae; Broady 1976; Negrisolo et al. 2004), *Chloroidium saccharophilum* (KX024691.1; Darienko et al. 2010), *Lobosphaera* sp. (KT119889.1), *Lobosphaera incisa* (KM020046.1) and *Lobosphaera tirolensis* (Chlorophyta; AB006051.1; Karsten et al. 2005) were present (Tables S4 and A7). In contrast, these species were absent in sample 2 (when surface and soil–near snow were avoided during sample collection). This highlights the importance of a consistent sampling strategy when the aim is to compare species composition and abundances between different sites (Fig. S2).

Discussion

Approaches to create a custom reference database of locally abundant taxa

Here, we show that by generating reference sequences of the locally abundant taxa and including them into the custom reference databases, the number of the identified OTUs increased. The use of monospecific snow algae blooms is advisable for obtaining long reference sequences of multiple DNA regions by Sanger sequencing. However, environmental sequences can be tricky and must be of high quality when used as references for HTS data (Rimet et al. 2018). A polyphasic approach (i.e., collectively using genetic, chemotaxonomic and phenotypic methods) is required to determine accurately the taxonomic identity of species found in field samples (Matsuzaki et al. 2015). Alternatively, single cells with identifiable morphologies, can be picked out of mixed samples for single–cell sequencing to link morphology to genotype (Bock et al. 2014). Light microscopy evaluations of cell morphologies should be conducted for each sample (directly after collection) to investigate which species might be present in the sequencing results. Qualitative light microscopic observation and identification may help link a phenotype of the most dominant morphotype with a genotype of the dominant OTU (e.g., OTU 'denovo99' prevailing in sample 2 is closely related to *Chloromonas brevispina* K–2 and they most likely share similar morphologies). The number of haplotypes of dominant species within a bloom might be revealed and compared with reports from elsewhere. For instance, a low haplotype diversity of *Chlamydomonas nivalis* DL07 is in par with previous findings on red snow from North America (Brown et al. 2016).

Table 5. Algal community structure based on the ITS2 data set, comprising the ten most abundant OTUs that made up >88% of the community. The table shows discrepancies between OTU assignments using three strategies: (A) basic version using Qiime and a custom database downloaded from NCBI, (B) extended version using Qiime and additional reference sequences of the locally abundant taxa (underlined), and (C) further extended version using final manual verification of taxa assignments which required ≥89.0% similarity (sequences below this threshold were recorded as "no blast hit") and the absence of compensatory base changes (CBC) in homology positions near the 5′– apex of helix III. A comprehensive list comprising the 38 most abundant taxa with their corresponding OTU identification numbers can be found in the Table S7. Our results reveal that the manual verification including secondary structure prediction and CBC search is essential, and thus, highly recommended.

OTU ID	Sample 1 (%)	Sample 2 (%)	(A) Qiime + NCBI database	(B) Qiime + NCBI database + local references	(C) Qiime + NCBI database + local references + manual verification (sequence similarity (%), sequence cover (%))
denovo99	1.8	59.1	*Chloromonas* sp. CCCryo289–06 HQ404893.1	*Chloromonas brevispina* K–2	No blast hit (88%, 83%, 6 CBC when compared denovo99 and *Chloromonas brevispina* K–2 – one CBC out of it is located in the most conserved part of structure, i.e. in top close to the 5′end of III helix, see Fig. 4)
denovo20	5.5	28.2	*Chlamydomonas nivalis* GU117577.1	*Chlamydomonas nivalis* DL07	*Chlamydomonas nivalis* DL07 (100%, 100%)
denovo107	39.8	<0.1	*Chloromonas* sp. CCCryo289–06 HQ404893.1	*Chloromonas brevispina* K–2	*Chloromonas brevispina* K–2 (100% identical except for one nucleotide – instead of 'R' in reference sequence, there was 'G', 100%)
denovo100	13.7	<0.1	*Chloromonas pichinchae* HQ404889.1	*Chloromonas pichinchae* HQ404889.1	*Chloromonas pichinchae* HQ404889.1 (92%, 100%, 1 CBC in helix III [outside the most conserved part] when compared denovo100 and *Chloromonas pichinchae*, see Figs S6, S7)
denovo63	10.9	0.2	No blast hit	*Scotiella cryophila* K–1	*Scotiella cryophila* K–1 (100%, 100%)
denovo142	8.2	1.7	*Chloromonas rostafinskii* HQ404863.1	*Chloromonas rostafinskii* HQ404863.1	No blast hit (79%, 60% – denovo142 vs. *Chloromonas rostafinskii*: 86%, 77% – denovo142 vs. *Chloromonas miwae* LC012762.1, four CBCs [one CBC out of it is located in the most conserved part of the structure, i.e., in the top close to the 5′end of III helix] when compared denovo142 and *Chloromonas miwa*e, sequence–structure alignment in Fig. S12)
denovo130	1.4	3.5	No blast hit	No blast hit	No blast hit (no significant similarity found)
denovo181	4.1	0	No blast hit	No blast hit	No blast hit (82%, 87%, *Desmococcus endolithicus* KX094830.1; five CBCs – one in helix II and four CBCs in helix III, see Figs S19, S20)

Table 5 Cont.

denovo254	3.1	0.1	*Chloromonas rostafinskii* HQ404863.1	*Chloromonas rostafinskii* HQ404863.1	No blast hit (82%, 48% – denovo254 vs. *Chloromonas rostafinskii*: 88%, 84% – denovo254 vs. *Chloromonas miwae* LC012762.1, three CBCs [one out of in the most conserved part of the structure] when compared denovo254 and *Chloromonas miwae*, see Figs S12, S14)
denovo23	0.1	2.2	No blast hit	No blast hit	No blast hit (no significant similarity found)
	11.4	5.0	Other	Other	Other

The 18S rDNA marker and its limitations

Currently, a considerably higher number of 18S rDNA reference sequences exists in public databases compared to ITS2 rDNA. Therefore, 18S rDNA seems to be the obvious marker of choice for high–throughput studies of eukaryotic communities. Several other gene loci, e.g. tufA (Vieira et al. 2016) and rbcL (Hall et al. 2010; Zou et al. 2016), have been recommended as the promising DNA barcode for some green algae (Hall et al. 2010). However, tufA records for the genus *Chloromonas*, which is frequently found in snow, are currently scarce in the NCBI database. Currently, the HTS technology limits the chosen marker to only a fraction of its actual length. The chosen V4–V5 region is the most variable region of the 18S rDNA gene for snow algal taxa, yet, the variability was not sufficient and unambiguous species assignments were often not feasible (Table 3). The most abundant 18S rDNA OTU was assigned to *Chloromonas brevispina* K–2. However, there was likely a 'hidden' diversity to a certain extent, since many different *Chloromonas* species can share up to 100% identity of this marker. An oligotyping approach (manual taxonomic assignment strategy C, further extended version), could resolve some of the "hidden" diversity for species with a difference of 1 bp in the sequenced amplicon. Therefore, this approach is highly recommended for further refinements of data sets that are characterized by very low variability, which is not detectable by conventional clustering of operational taxonomic units where 97% or 98% clustering levels are applied. Furthermore, due to the possibility of identical reference sequences of species that are not present in the habitat, it is essential to check alignments with the reference sequences manually. For instance, *Chloromonas polyptera* has been reported in HTS studies from several places in the northern hemisphere (Lutz et al. 2015; Terashima et al. 2017), yet, these results seem to be caused by ambiguous species assignments. For instance, Terashima et al. (2017) reported 100% identity in the 18S rDNA of their OTU44 with *Chloromonas polyptera* (JQ790556). However, the sequenced gene fragment also shares 100% identity with *Chloromonas* sp. Gassan B (LC012714.1), *Chloromonas* sp. ANT1 (AB903007.1), *Chloromonas* sp. TA8 (AB902996.1) and an uncultured 'Viridiplantae' clone (HQ188979.1). In summary, 18S rDNA amplicons do not adequately identify taxa on the species level in several taxonomic groups (Xiao et al. 2014).

Advantages of the ITS2 marker

In contrast to the 18S rDNA marker, Illumina reads can span the entire region of ITS2 rDNA. This hypervariable molecular marker provides a much higher resolution than 18S rDNA. The prediction of the ITS2 rDNA transcript secondary structures allowed a thorough identification of the haplotype diversity (manual taxonomic assignment strategy C, further extended version). It is therefore a powerful tool to delete wrong OTU assignments and represents an appropriate way of describing the true biodiversity (e.g., in sample 2 the most abundant OTU 'denovo99' is not *Chloromonas brevispina* K–2). Nevertheless, the methodological approach of ITS2 rRNA secondary structure prediction of each OTU is currently immensely time–consuming. The process is partly automatized (e.g., using Mfold, 4SALE), yet, significant input of manual validation and correction is still required. Wolf et al. (2013) reported a probability of ~0.99 that no intragenomic CBC took place, based on the comparison of ITS2 of 178 species of land plants. However, some *Chloromonas* species in snow possess intragenomic CBCs (Matsuzaki et al. 2015). As a consequence, evaluation of CBCs detection should be carried out carefully; i.e., not only the pure presence or absence of CBCs, but also the exact position in the ITS2 molecule (i.e., whether it is in or outside the most conserved part) and the overall level of genetic difference of ITS2 between OTUs should be taken into account. For instance, among three OTUs from aplanozygotes and strains of *C. miwae*, one CBC was observed between specimen Gassan–C/strain NIES–2379 and strain NIES–2380 (Matsuzaki et al. 2015). This CBC was located outside the most conserved branch of helix III and the genetic differences in the nuclear rDNA ITS2 region between

the aplanozygote specimen and the *C. miwae* strain were only 0.0 to 0.4% (see Fig. S15 in MATSUZAKI et al. 2015). Thus, both field specimens and the strain are regarded as one species. As a consequence, species boundaries in our proposed HTS approach rely on CBCs in the most conserved part of helix III only (i.e., close to its 5′apex).

A two–marker approach is mandatory

None of the two markers adequately described the community composition, either due to their low resolution (18S) or due to the lack of reference sequences (ITS2). Moreover, the output can also be partly influenced by library production biases and primer inefficiencies. Combining the strengths of both markers is thus recommended, particular in less–well studied environments. In addition, one marker can guide the data optimization of the other marker. Despite its low resolution, the 18S marker can provide guidance, which ITS2 rDNA reference sequences need to be generated and vice versa. Depending on the used markers, the estimated community composition may differ, mainly as a result of different primer specificity (VĚTROVSKÝ et al. 2016). The use of different markers can also address different aspects of community composition analyses: (a) the comparison of single–copy versus multi–copy markers provides better relative abundance approximations, (b) in contrast to non–coding markers, coding genes can be used to identify pseudogenes and construct phylogenetic (TONK et al. 2013; VĚTROVSKÝ et al. 2016). The potential of multi–marker approaches has been highlighted for species discovery in metabarcoding studies (MARCELINO & VERBRUGGEN 2016), as well as for assessing the effectiveness to distinguish cryptic species in a model morphospecies (EVANS et al. 2007).

Advantages, limitations and perspectives of HTS for community composition analyses

HTS allowed a more comprehensive assessment of the prevailing biodiversity than traditional Sanger sequencing and light microscopic observations. In addition to the detection of low–abundant taxa, a multitude of sequences, which did not match any references in the databases, were generated. Some of these likely represent new species (e.g., OTU 'denovo99'). Strain–based taxonomic studies or accurate species determination of monospecific field blooms by Sanger sequencing (to gain complete reads of the target marker) are required to increase the number of reference sequences. On the other hand, Sanger sequencing can be problematic for unrecognized mixed communities in terms of chromatogram corrections unless cloning is involved. Alternatively, putative monospecies snowfields can be sampled for HTS studies to evaluate the biodiversity and for the detection of any intragenomic variations of ITS2.

The quality of the reference databases is crucial for identification of various microorganisms including dinophytes (SOEHNER et al. 2012), diatoms (VISCO et al. 2015) and cryptophytes (HOEF–EMDEN 2012). New entries must be continuously updated. The Qiime–compatible Silva database delivers reference sequences in one batch; yet, an updated version is only released about once a year. In contrast, NCBI is under continuous revision and therefore we recommend that new potential reference sequences are added manually to the Qiime–compatible Silva database prior to use. Only such optimized data can then be used for further evaluations including phylogeography and phylogenetic studies based on the generation of multi–locus sequence data in a fast and cost–effective way (MCCORMACK et al. 2013). Alternatively, the use of the PR2 database (GUILLOU et al. 2013) may offer a better taxonomic assignment than Silva, especially for green algae.

In this methodological case study, we evaluated the application of high–throughput sequencing on an unconventional ecosystem of melting snowfields. Based on light microscopic observations, the investigated snowfields were dominated by three algal species, which were however not always reflected in the sequencing dataset. Consequently, HTS data need to be handled with care if applied on habitats or groups of organisms that are (highly) underrepresented in molecular databases. Currently, the need to generate appropriate reference sequences for the key taxa in the studied environment is an inevitable task for such studies. Furthermore, the two–marker approach, a consistent sampling strategy, light microscopy–based guidance and a final manual verification of all taxonomic assignments are strongly recommended.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Andreas Holzinger for providing access to his lab at the Institute of Botany, Dr. Birgit Sattler (Institute of Ecology) for providing access to the Limnological Station in Kühtai, both at the University of Innsbruck, Austria, and Dr. Thomas Leya (Fraunhofer IZI–BB, Potsdam–Golm, Germany) for providing access to his CCCryo 18S rDNA database. DR acknowledges funding from the Austrian Science Fund (FWF): P29959. SL and LGB acknowledge funding from the Helmholtz Recruiting Initiative (award number I–044–16–01). LP and LN acknowledge funding from the Czech Science Foundation (GACR) project 18–02634S and from the Institutional Research Concept RVO67985939 (LN).

REFERENCES

ALANAGREH, L.; PEGG, C.; HARIKUMAR, A. & BUCHHEIM, M. (2017): Assessing intragenomic variation of the internal transcribed spacer two: Adapting the Illumina metagenomics protocol. − PLOS one 12: e0181491.

ANESIO, A.M.; LUTZ, S.; CHRISMAS, N.A.M. & BENNING, L.G. (2017): The microbiome of glaciers and ice sheets. − Npj Biofilms and Microbiomes 3: 10.

BENGTSSON–PALME, J.; RYBERG, M.; HARTMANN, M.; BRANCO, S.; WANG, Z.; GODHE, A.; DE WIT, P.; GARCÍA-SÁNCHEZ, M.; EBERSBERGER, I.; DE SOUSA, F.; AMEND, A.; JUMPPONEN, A.; UNTERSEHER, M.; KRISTIANSSON, E.; ABARENKOV, K.; BERTRAND, Y.J.K.; SANLI, K.; ERIKSSON, K.M.; VIK, U.; VELDRE, V. & NILSSON, R.H. (2013): Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. − Methods

in Ecology and Evolution 4: 914–919.

Bock, C.; Medinger, R.; Jost, S.; Psenner, R. & Boenigk, J. (2014): Seasonal variation of planktonic chrysophytes with special focus on *Dinobryon*. − Fottea 14: 179–190.

Bradley, I.M.; Pinto, A.J. & Guest, J.S. (2016): Design and Evaluation of Illumina MiSeq–Compatible, 18S rRNA gene–specific primers for improved characterization of mixed phototrophic communities. − Applied and Environmental Microbiology 82: 5878–5891.

Brandariz–Fontes, C.; Camacho–Sanchez, M.; Vilà, C.; Vega–Pla, J.L.; Rico, C. & Leonard, J.A. (2015): Effect of the enzyme and PCR conditions on the quality of high–throughput DNA sequencing results. − Scientific Reports 5: 8056.

Broady, P.A. (1976): Six new species of terrestrial algae from Signy Island, South Orkney Islands, Antarctica. − British Phycological Journal 11: 387–405.

Brown, S.P.; Ungerer, M.C. & Jumpponen, A. (2016): A Community of Clones: Snow Algae Are Diverse Communities of Spatially Structured Clones. − International Journal of Plant Sciences 177: 432–439.

Buchheim, M.A.; Keller, A.; Koetschan, C.; Förster, F.; Merget, B.; Wolf, M. & M. (2011): Internal Transcribed Spacer 2 (nu ITS2 rRNA) Sequence–Structure Phylogenetics: Towards an Automated Reconstruction of the Green Algal Tree of Life. − PLoS ONE 6: e16931.

Caisová, L.; Marin, B. & Melkonian, M. (2013): A Consensus Secondary Structure of ITS2 in the Chlorophyta Identified by Phylogenetic Reconstruction. − Annals of Anatomy 164: 482–496.

Campo, J. del; Kolisko, M.; Boscaro, V.; Santoferrara, L.F.; Nenarokov, S.; Massana, R.; Guillou, L.; Simpson, A.; Berney, C.; de Vargas, C.; Brown, M.W.; Keeling, P.J. & Parfrey, L.W. (2018): EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. − PLOS Biology 16: e2005849.

Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I.; Huttley, G.A.; Kelley, S.T.; Knights, D.; Koenig, J.E.; Ley, R.E.; Lozupone, C.A.; McDonald, D.; Muegge, B.D.; Pirrung, M.; Reeder, J.; Sevinsky, J.R.; Turnbaugh, P.J.; Walters, W.A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J. & Knight, R. (2010): QIIME allows analysis of high–throughput community sequencing data. − Nature Methods 7: 335–336.

Chase, M.W. & Fay, M.F. (2009): Ecology. Barcoding of plants and fungi. − Science 325: 682–683.

Cheung, M.K.; Au, C.H.; Chu, K.H.; Kwan, H.S. & Wong, C.K. (2010): Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. − The ISME Journal 4: 1053–1059.

Coleman, A. (2007): Pan–eukaryote ITS2 homologies revealed by RNA secondary structure. − Nucleic Acids Research 35: 3322–3329.

Coleman, A.W. (2000): The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. − Protist 151: 1–9.

Coleman, A.W. (2009): Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. − Molecular Phylogenetics and Evolution 50: 197–203.

Coleman, A.W.; Suarez, A. & Goff, L.J. (1993): Molecular delineation of species and syngens in Volvocacean green algae (Chlorophyta). − Journal of Phycology 30: 80–90.

Darienko, T.; Gustavs, L.; Mudimu, O.; Menendez, C.R.; Schumann, R.; Karsten, U.; Friedl, T. & Pröschold, T. (2010): *Chloroidium*, a common terrestrial coccoid green alga previously assigned to Chlorella (Trebouxiophyceae, Chlorophyta). − European Journal of Phycology 45: 79–95.

Darty, K.; Denise, A. & Ponty, Y. (2009): VARNA: Interactive drawing and editing of the RNA secondary structure. − Bioinformatics 25: 1974–1975.

Eddy, S.R. (1996): Hidden Markov models. − Current Opinion in Structural Biology 6: 361–365.

Eren, A.M.; Maignien, L.; Sul, W.J.; Murphy, L.G.; Grim, S.L.; Morrison, H.G. & Sogin, M.L. (2013): Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. − Methods in Ecology and Evolution 4: 1111–1119.

Evans, K. M.; Wortley, A. H. & Mann, D. G. (2007): An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). − Protist 158: 349−364.

Frey, B.; Rime, T.; Phillips, M.; Stierli, B.; Hajdas, I.; Widmer, F. & Hartmann, M. (2016): Microbial diversity in European alpine permafrost and active layers. − FEMS Microbiology Ecology 92: 1–17.

Grossmann, L.; Beisser, D.; Bock, C.; Chatzinotas, A.; Jensen, M.; Preisfeld, A.; Psenner, R.; Rahmann, S.; Wodniok, S. & Boenigk, J. (2016): Trade–off between taxon diversity and functional diversity in European lake ecosystems. − Molecular Ecology 25: 5876–5888.

Guillou, L.; Bachar, D.; Audic, S.; Bass, D.; Berney, C.; Bittner, L.; Boutte, C.; Burgaud, G.; de Vargas, C.; Decelle, J.; del Campo, J.; Dolan, J.R.; Dunthorn, M.; Edvardsen, B.; Holzmann, M.; Kooistra, W.H.C.F.; Lara, E.; Le Bescot, N.; Logares, R.; Massana, F.M.R.; Montresor, M.; Morard, R.; Not, F.; Pawlowski, J.; Probert, I.; Sauvadet, A-L.; Siano, R.; Stoeck, T.; Vaulot, D.; Zimmermann, P. & Christen, R. (2013): The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub–Unit rRNA sequences with curated taxonomy. − Nucleic Acids Research 41: D597–D604.

Hall, J.D.; Fučíková, K.; Lo, C.; Lewis, L.A. & Karol, K.G. (2010): An assessment of proposed DNA barcodes in freshwater green algae. − Cryptogamie, Algologie 31: 529–555.

Hoham, R.W. & Duval, B. (2001): Microbial ecology of snow and freshwater ice with emphasis on snow algae. – In: Jones, H.G.; Pomeroy, J.W.; Walker, D.A. & Hoham, R. (eds.): Snow Ecology. An Interdisciplinary Examination of Snow–covered Ecosystems. – pp. 168–228, Cambridge University Press, Cambridge.

Hoef–Emden, K. (2012): Pitfalls of establishing DNA barcoding systems in protists: the Cryptophyceae as a test case. – PloS One 7: e43652.

Hoham, R.W.; Roemer, S.C. & Mullet, J.E. (1979): The life history and ecology of the snow alga *Chloromonas brevispina* comb. nov. (Chlorophyta, Volvocales). − Phycologia 18: 55–70.

Illumina: 16S Metagenomic Sequencing Library Preparation Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System. 1–28. (support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). Accessed13August2014.

Karsten, U.; Friedl, T.; Schumann, R.; Hoyer, K. & Lembcke, S. (2005): Mycosporine–like amino acids and phylogenies in green algae: *Prasiola* and its relatives from the Trebouxiophyceae (Chlorophyta). − Journal of Phycology 41: 557 Stuttgart: 566.

Kent, W.J. (2002): BLAT—the BLAST–like alignment tool. − Genome Research 12: 656–664.

Kol, E. (1968): Kryobiologie; Biologie und Limnologie des Schnees und Eises. Die Binnengewasser (Vol. 24). – 216 pp., Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.

Komárek, J. & Nedbalová, L. (2007): Green Cryosestic Algae. – In: Seckbach, J. (ed.): Algae and Cyanobacteria in Extreme Environments. – pp. 321–342, Springer, Netherlands.

Leliaert, F.; Verbruggen, H.; Vanormelingen, P.; Steen, F.; López–Bautista, J.M.; Zuccarello, G.C. & De Clerck, O. (2014): DNA–based species delimitation in algae. − European Journal of Phycology 49: 179–196.

Leya, T. (2013): Snow algae: adaptation strategies to survive on snow and ice. – In: Seckbach, J.; Oren, A. & Stan–Lotter, H. (eds.): Polyextremophiles. – pp. 401–423, Springer, Cham.

Lindner, D.L. & Banik, M.T. (2011): Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. − Mycologia 103: 731–740.

Lindner, D.L.; Carlsen, T.; Henrik Nilsson, R.; Davey, M.; Schumacher, T. & Kauserud, H. (2013): Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. − Ecology and Evolution 3: 1751–1764.

Lutz, S.; Anesio, A.M.; Edwards, A. & Benning, L.G. (2015): Microbial diversity on icelandic glaciers and ice caps. − Frontiers in Microbiology 6: 307.

Lutz, S.; Anesio, A.M.; Edwards, A. & Benning, L.G. (2017): Linking microbial diversity and functionality of arctic glacial surface habitats. − Environmental Microbiology 19: 551–565.

Lutz, S.; Anesio, A.M.; Field, K. & Benning, L.G. (2015): Integrated "Omics", targeted metabolite and single–cell analyses of arctic snow algae functionality and adaptability. − Frontiers in Microbiology 6: 1323.

Lutz, S.; Anesio, A.M.; Raiswell, R.; Edwards, A.; Newton, R.J.; Gill, F. & Benning, L.G. (2016): The biogeography of red snow microbiomes and their role in melting arctic glaciers. − Nature Communications 7: 11968.

Lutz, S.; McCutcheon, J.; McQuaid, J.B. & Benning, L.G. (2018): The diversity of ice algal communities on the Greenland Ice Sheet as revealed by oligotyping. − Microbial Genomics 4: e000159.

Marcelino, V. R. & Verbruggen, H. (2016): Multi–marker metabarcoding of coral skeletons reveals a rich microbiome and diverse evolutionary origins of endolithic algae. − Scientific Reports 6: 31508.

Matsuzaki, R.; Hara, Y. & Nozaki, H. (2012): A taxonomic revision of *Chloromonas reticulata* (Volvocales, Chlorophyceae), the type species of the genus *Chloromonas*, based on multigene phylogeny and comparative light and electron microscopy. − Phycologia 51: 74–85.

Matsuzaki, R.; Kawai–Toyooka, H.; Hara, Y. & Nozaki, H. (2015): Revisiting the taxonomic significance of aplanozygote morphologies of two cosmopolitan snow species of the genus *Chloromonas* (Volvocales, Chlorophyceae). − Phycologia 54: 491–502.

McCormack, J.E.; Hird, S.M.; Zellmer, A.J.; Carstens, B.C. & Brumfield, R.T. (2013): Applications of next–generation sequencing to phylogeography and phylogenetics. − Molecular Phylogenetics and Evolution 66: 526–538.

Mikhailyuk, T.I.; Sluiman, H.J.; Massalski, A.; Mudimu, O.; Demchenko, E.M.; Kondratyuk, S.Y. & Friedl, T. (2008): New streptophyte green algae from terrestrial habitats and an assessment of the genus *Interfilum* (Klebsormidiophyceae, Streptophyta). − Journal of Phycology 44: 1586–1603.

Negrisolo, E.; Maistro, S.; Incarbone, M.; Moro, I.; Dalla Valle, L.; Broady, P.A. & Andreoli, C. (2004): Morphological convergence characterizes the evolution of Xanthophyceae (Heterokontophyta): evidence from nuclear SSU rDNA and plastidial rbcL genes. − Molecular Phylogenetics and Evolution 33: 156–170.

Novis, P.M. (2002): Ecology of the snow alga *Chlainomonas kolii* (Chlamydomonadales, Chlorophyta) in New Zealand. − Phycologia 41: 280–292.

Oyola, S.O.; Otto, T.D.; Gu, Y.; Maslen, G.; Manske, M.; Campino, S.; Turner, D.; MacInnis, B.; Kwiatowski, D.; Swerdlow, H. & Quail, M.A. (2012): Optimizing illumina next–generation sequencing library preparation for extremely AT–biased genomes. − BMC Genomics 13: 1.

Procházková, L.; Remias, D.; Holzinger, A.; Řezanka, T. & Nedbalová, L. (2018b): Ecophysiological and morphological comparison of two populations of *Chlainomonas* sp. (Chlorophyta) causing red snow on ice–covered lakes in the High Tatras and Austrian Alps. − European Journal of Phycology 53: 230–243.

Procházková, L.; Remias, D.; Řezanka, T. & Nedbalová, L. (2018a): *Chloromonas nivalis* subsp. *tatrae*, subsp. nov. (Chlamydomonadales, Chlorophyta): Re–examination of a snow alga from the High Tatra Mountains (Slovakia). − Fottea 18: 1–18.

Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J. & Glöckner, F.O. (2013): The SILVA ribosomal RNA gene database project: Improved data processing and web–based tools. − Nucleic Acids Research 41: D590–D596.

Remias, D. (2012): Cell structure and physiology of alpine snow and ice algae. – In: Lütz, C. (ed.): Plants in Alpine Regions. – pp. 175–185, Springer, Vienna.

Remias, D.; Holzinger, A.; Aigner, S. & Lütz, C. (2012): Ecophysiology and ultrastructure of *Ancylonema nordenskiöldii* (Zygnematales, Streptophyta), causing brown ice on glaciers in Svalbard (high arctic). − Polar Biology 35: 899–908.

Remias, D.; Karsten, U.; Lütz, C. & Leya, T. (2010): Physiological and morphological processes in the Alpine snow alga *Chloromonas nivalis* (Chlorophyceae) during cyst formation. − Protoplasma 243: 73–86.

Remias, D.; Lütz–Meindl, U. & Lütz, C. (2005): Photosynthesis, pigments and ultrastructure of the alpine snow alga *Chlamydomonas nivalis*. − European Journal of Phycology 40: 259–268.

Remias, D.; Pichrtová, M.; Pangratz, M.; Lütz, C. & Holzinger, A. (2016): Ecophysiology, secondary pigments and ultrastructure of *Chlainomonas* sp. (Chlorophyta) from the European Alps compared with *Chlamydomonas nivalis* forming red snow. − FEMS Microbiology Ecology 92: fiw030.

Remias, D.; Procházkova, L.; Holzinger, A. & Nedbalová, L. (2018): Ecology, cytology and phylogeny of the snow alga *Scotiella cryophila* K–1 (Chlorophyceae) from the Austrian Alps. − Phycologia. 57: 581–592.

Remias, D.; Wastian, H.; Lütz, C. & Leya, T. (2013): Insights into the biology and phylogeny of *Chloromonas polyptera* (Chlorophyta), an alga causing orange snow in Maritime Antarctica. − Antarctic Science 25: 648–656.

Rimet, F.; Abarca, N.; Bouchez, A.; Kusber, W.H.; Jahn, R.; Kahlert, M.; Keck, F.; Kelly, M.G.; Mann, D.G.; Piuz, A.; Trobajo, R.; Tapolczai, K.; Vasselon, V. & Zimmermann, J. (2018): The potential of High–Throughput Sequencing (HTS) of natural samples as a source of primary taxonomic information for reference libraries

of diatom barcodes. − Fottea 18: 37–54.

Schirmer, M.; Ijaz, U.Z.; D'Amore, R.; Hall, N.; Sloan, W.T. & Quince, C. (2015): Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. − Nucleic Acids Research 43: e37.

Schloss, P.D.; Gevers, D. & Westcott, S.L. (2011): Reducing the effects of PCR amplification and sequencing Artifacts on 16S rRNA–based studies. − PLoS One 6: e27310.

Schmidt, P.A.; Bálint, M.; Greshake, B.; Bandow, C.; Römbke, J. & Schmitt, I. (2013): Illumina metabarcoding of a soil fungal community. − Soil Biology and Biochemistry 65: 128–132.

Schultz, J. & Wolf, M. (2009): ITS2 sequence–structure analysis in phylogenetics: A how–to manual for molecular systematics. − Molecular Phylogenetics and Evolution 52: 520–523.

Segawa, T.; Matsuzaki, R.; Takeuchi, N.; Akiyoshi, A.; Navarro, F.; Sugiyama, S.; Yonezawa, T. & Mori, H. (2018): Bipolar dispersal of red–snow algae. − Nature Communications 9: 3094.

Seibel, P.N.; Müller, T.; Dandekar, T.; Schultz, J. & Wolf, M. (2006): 4SALE − a tool for synchronous RNA sequence and secondary structure alignment and editing. − BMC Bioinformatics 7: 498.

Seibel, P.N.; Müller, T.; Dandekar, T. & Wolf, M. (2008): Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. − BMC Research Notes 1: 91.

Simon, U. & Weiss, M. (2008): Intragenomic variation of fungal ribosomal genes is higher than previously thought. − Molecular Biology and Evolution 25: 2251−2254.

Soehner, S.; Zinssmeister, C.; Kirsch, M. & Gottschling, M. (2012): Who am I—and if so, how many? Species diversity of calcareous dinophytes (Thoracosphaeraceae, Peridiniales) in the Mediterranean Sea. − Organisms Diversity & Evolution 12: 339−348.

Stibal, M. & Elster, J. (2005): Growth and morphology variation as a response to changing environmental factors in two Arctic species of *Raphidonema* (Trebouxiophyceae) from snow and soil. − Polar Biology 28: 558–567.

ter Braak, C.J.F. & Šmilauer, P. (2012): CANOCO reference manual and user's guide: Software for ordination (version 5.0). − 496 pp. Microcomputer Power, Ithaca, NY.

Terashima, M.; Umezawa, K.; Mori, S.; Kojima, H. & Fukui, M. (2017): Microbial community analysis of colored snow from an alpine snowfield in Northern Japan reveals the prevalence of Betaproteobacteria with snow algae. − Frontiers in Microbiology 8: 1–13.

Thornhill, D.J.; Lajeunesse, T.C. & Santos, S.R. (2007): Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. − Molecular Ecology 16: 5326–5340.

Tonk, L.; Bongaerts, P.; Sampayo, E. M. & Hoegh–Guldberg, O. (2013): SymbioGBR: a web–based database of *Symbiodinium* associated with cnidarian hosts on the Great Barrier Reef. − BMC Ecology 13: 7.

Tragin, M.; Zingone, A. & Vaulot, D. (2017): Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. − Environmental Microbiology 20: 506–520.

Větrovský, T.; Kolařík, M.; Žifčáková, L.; Zelenka, T. & Baldrian, P. (2016): The rpb2 gene represents a viable alternative molecular marker for the analysis of environmental fungal communities. – Molecular Ecology Resources 16: 388–401.

Vieira, H.H.; Bagatini, I.L.; Guinart, C.M. & Vieira, A.A.H. (2016): tufA gene as molecular marker for freshwater Chlorophyceae. – Algae 31: 155–165.

Visco, J. A.; Apothéloz–Perret–Gentil, L.; Cordonier, A.; Esling, P.; Pillet, L. & Pawlowski, J. (2015): Environmental monitoring: inferring the diatom index from next–generation sequencing data. Environmental science & technology 49: 7597–7605.

White, T.J.; Bruns, T.; Lee, S. & Taylor, J. (1990): Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. – In: Innis, M.A.; Gelfand, D. H.; Sninsky, J. J. & White, T. J. (eds.): PCR Protocols. – pp. 315–322, Elsevier.

Wolf, M.; Chen, S.; Song, J.; Ankenbrand, M. & Müller, T. (2013): Compensatory Base Changes in ITS2 Secondary Structures Correlate with the Biological Species Concept Despite Intragenomic Variability in ITS2 Sequences – A Proof of Concept. − PLoS One 8: e66726.

Xiao, X.; Sogge, H.; Lagesen, K.; Tooming–Klunderud, A.; Jakobsen, K.S. & Rohrlack, T. (2014): Use of High Throughput Sequencing and Light Microscopy Show Contrasting Results in a Study of Phytoplankton Occurrence in a Freshwater Environment. − PLoS One 9: e106510.

Yao, H.; Song, J.; Liu, C.; Luo, K.; Han, J.; Li, Y.; Pang, X.; Xu, H.; Zhu, Y.; Xiao, P. & Chen, S. (2010): Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. − PLoS One 5: e13102.

Zou, S.; Fei, C.; Wang, C.; Gao, Z.; Bao, Y.; He, M. & Wang, C. (2016): How DNA barcoding can be more effective in microalgae identification: a case of cryptic diversity revelation in *Scenedesmus* (Chlorophyceae). – Scientific reports 6: 36822.

Zuker, M. (2003): Mfold web server for nucleic acid folding and hybridization prediction. − Nucleic Acids Research 31: 3406–3415.

Supplementary material

the following supplementary material is available for this article:

Supplementary material is available for this article.

This material is available as part of the online article (http://fottea.czechphycology.cz/contents)

Data sharing and data accessibility

18S and ITS2 rDNA amplicon sequences have been deposited to the European Nucleotide Archive (ENA) under accession number PRJEB24479 (Please note that data has been deposited but not been released yet, and they will be made public upon manuscript acceptance). Sanger sequences of reference species were deposited in NCBI under accession numbers listed in supplementary Table S5. All other data are presented in this manuscript and in the further supplemental files.

Received October 25, 2018
Accepted February 6, 2019