To investigate the PCR bias of the ITS2 primers, we performed shotgun metagenomic sequencing on the Svalbard sample (Austre Brøggerbreen) and compared taxonomic compositions between the ITS2 amplicon sequencing data and the shotgun metagenomic sequencing data. Both of the taxonomic compositions were inferred using ITS2 sequences.

DNA concentration was determined with the Quant-iT dsDNA HS assay kit using a Qubit fluorometer (Life Technologies). For construction of the genome library for Illumina sequencing, an aliquot of DNA (60 ng) was sheared to a target peak size of 450 bp using the Covaris S220 Focused-Ultrasonicator system (Covaris) according to the manufacturer's recommendations. To generate PCR-free DNA sequencing libraries, the KAPA Hyper Prep kit (Kapa Biosystems) was used according to the manufacturer's instructions. The library products were isolated via agarose gel electrophoresis (size range, 450–800 bp) and purified by use of a NucleoSpin Gel and PCR Clean-up (Macherey Nagel). The sequencing library was used as a template for paired-end sequencing using the MiSeq reagent kit v3 and the MiSeq sequencer (Illumina). Read files (fastq.gz) were generated using MiSeq Reporter software version 2.3.32 (Illumina).

For shotgun metagenomic sequencing, *in silico* extractions of ITS2 sequences and taxonomic assignments were performed as follows. We discarded: (i) the R2 reads of the MiSeq paired-end reads, (ii) the R1 reads that contained ambiguous nucleotides, (iii) reads that contained <50 nt, and (iv) reads that mapped to the PhiX genome sequence based on a search using Bowtie 2 (version 2.2.4) with default parameters. Then, 14,308,717 high-quality metagenomic reads remained. These high-quality metagenomic reads were then used to compare taxonomic composition against the ITS2 amplicon sequencing results. For this purpose, we carried out a BLASTN-based taxonomic assignment against 348 representative sequences of 98% OTUs that were identified in our present study; the taxonomic assignment was performed with a top-hit E-value <1e–8, identity >90%, and alignment length >150 bp.

The taxonomic assignment results (Fig. 1) revealed that both the amplicon sequencing data and shotgun metagenomic sequencing data for the sample exhibited a similar taxonomic composition (Pearson correlation coefficient = 0.99). This result indicates that the taxonomic composition based on amplicon sequencing of ITS2 presented in this study is not biased.

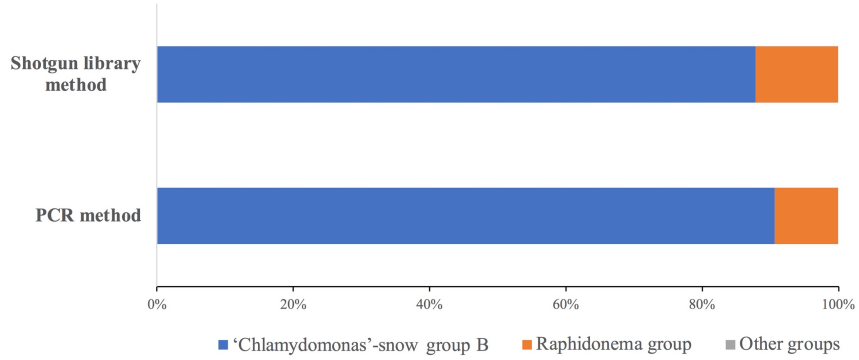

Fig. 1 Taxonomic classification of snow algae using shotgun reads data and PCR amplicon sequencing in red snow of the Svalbard sample (Austre Brøggerbreen) based on the ITS2 region sequences.