

# scientific data



**OPEN**  
**DATA DESCRIPTOR**

## Community composition of aquatic fungi across the thawing Arctic

Mariana Kluge<sup>1</sup>✉, Christian Wurzbacher<sup>2</sup>, Maxime Wauthy<sup>3,4</sup>, Karina Engelbrecht Clemmensen<sup>1</sup>, Jeffrey Alistair Hawkes<sup>5</sup>, Karolina Einarsdottir<sup>6</sup>, Jan Stenlid<sup>1</sup> & Sari Peura<sup>1</sup>

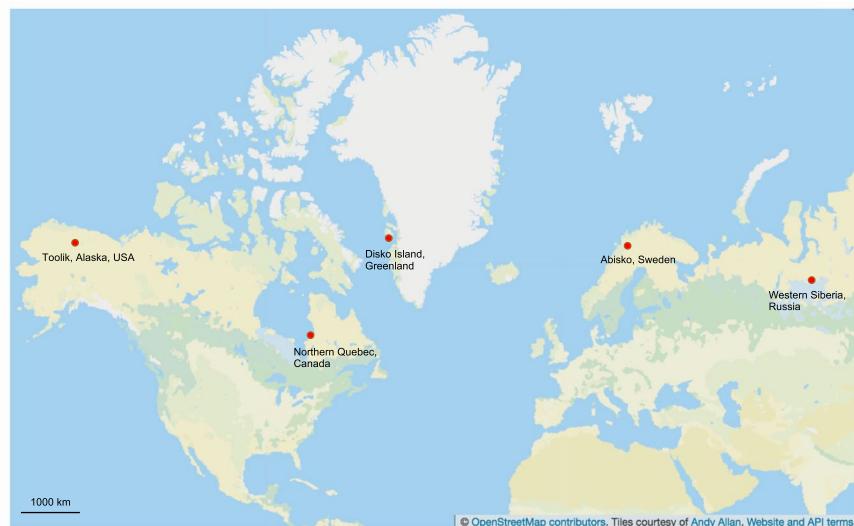
Thermokarst activity at permafrost sites releases considerable amounts of ancient carbon to the atmosphere. A large part of this carbon is released via thermokarst ponds, and fungi could be an important organismal group enabling its recycling. However, our knowledge about aquatic fungi in thermokarstic systems is extremely limited. In this study, we collected samples from five permafrost sites distributed across circumpolar Arctic and representing different stages of permafrost integrity. Surface water samples were taken from the ponds and, additionally, for most of the ponds also the detritus and sediment samples were taken. All the samples were extracted for total DNA, which was then amplified for the fungal ITS2 region of the ribosomal genes. These amplicons were sequenced using PacBio technology. Water samples were also collected to analyze the chemical conditions in the ponds, including nutrient status and the quality and quantity of dissolved organic carbon. This dataset gives a unique overview of the impact of the thawing permafrost on fungal communities and their potential role on carbon recycling.

### Background & Summary

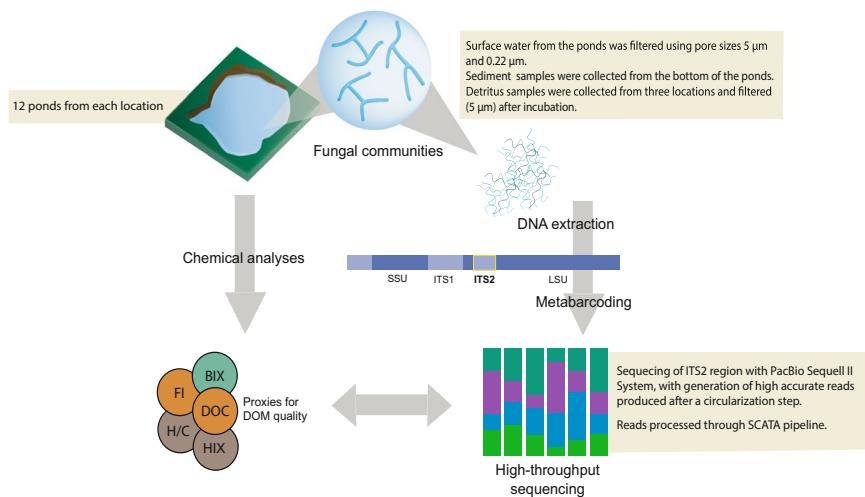
Frozen tundra soils hold one of the Earth's largest pools of organic carbon. With ongoing climate change, permafrost is thawing rapidly, especially in the Arctic and Subarctic regions, causing the release of a large fraction of this carbon<sup>1,2</sup>. The thawing of permafrost creates small and shallow waterbodies, hereafter referred to as thermokarst ponds<sup>3</sup>. The vast amount of organic matter released from the degrading permafrost ends up in these ponds<sup>4</sup>, where it can sink and be stored in the sediment, or be recycled in the microbial loop, generating greenhouse gases (GHG) as end products<sup>5,6</sup>. Most of the research on the microbial activity in the thermokarst ponds concentrates on prokaryotes<sup>7–10</sup> and, despite the central role of fungi as decomposers of the organic matter in terrestrial ecosystems<sup>11–13</sup>, very little is known about the fungal communities in aquatic environments, especially in the Arctic. To our knowledge, only one earlier study has specifically targeted the fungi in thermokarst ponds, highlighting that a major part of the aquatic fungal community in the thermokarst ponds belongs to unknown phyla<sup>14</sup>.

In this dataset, we collected surface water, detritus and sediment from thermokarst ponds in five different permafrost areas in the Arctic. These areas represent different stages of permafrost degradation from unaffected permafrost sites (represented by Alaska and Greenland, pristine sites) to sites affected by increasing severity of thermokarst activity (represented by Canada, Sweden and Russia, degraded sites) (Fig. 1). For each site, 12 ponds were sampled. Moreover, at the Canadian site, the ponds represented three different stages of permafrost thaw, including emerging, developing and mature thermokarst ponds; four ponds were sampled for each of these three stages<sup>8,14</sup>. This allowed us to investigate whether there is a succession of the community over pond development, when the quality and availability of carbon sources gradually changes. All the samples were extracted for the total metagenomic DNA, which was then amplified for fungal ITS2 region of the ribosomal genes and sequenced using PacBio (Fig. 2). We also collected water samples for chemical and optical analyses, in order to investigate nutrients and GHG concentrations as well as the quantity and quality of the dissolved organic matter (DOM). This included nutrients (dissolved nitrogen (DN), nitrate ( $\text{NO}_3^-$ ), nitrite ( $\text{NO}_2^-$ ), ammonium ( $\text{NH}_4^+$ ), sulfate ( $\text{SO}_4^{2-}$ )), total phosphorous (total P), total iron (Fe), GHG (carbon dioxide ( $\text{CO}_2$ ), methane ( $\text{CH}_4$ )) and dissolved organic

<sup>1</sup>Department of Forest Mycology and Plant Pathology, Science for Life laboratory, Swedish University of Agricultural Sciences, Uppsala, Sweden. <sup>2</sup>Department of Civil, Geo and Environmental Engineering, Technische Universität München, Munich, Germany. <sup>3</sup>Département des Sciences Fondamentales, Université du Québec à Chicoutimi, Saguenay, Québec, Canada. <sup>4</sup>Centre for Northern Studies (CEN), Université Laval, Québec, Québec, Canada. <sup>5</sup>Department of Chemistry, BMC, Uppsala University, Uppsala, Sweden. <sup>6</sup>Limnology, Department of Ecology and Genetics, EBC, Uppsala University, Uppsala, Sweden. ✉e-mail: mariana.kluge@slu.se



**Fig. 1** Map showing the five locations of the sampling sites: Alaska, Canada, Greenland, Sweden and Siberia.



**Fig. 2** Workflow illustrating the experimental design for sampling and the generation of the amplicon and chemical data.

carbon (DOC) concentrations, as well as various proxies of DOM such as fluorescence index (FI), freshness index (BIX), humification index (HIX), specific ultraviolet absorbance ( $SUVA_{254}$ ), spectral slope for the intervals 279–299 nm ( $S_{289}$ ) and average H/C and O/C.

The aim of this data collection was to study how the fungal diversity is affected by permafrost thaw and the resulting inputs of organic matter to the thermokarst ponds. Further, the impact of general chemical conditions in the ponds and their relationship to fungal community composition was addressed. Generally, this dataset gives unique insights into the composition of fungal communities in aquatic habitats in the Arctic. Thus, the data can be used to study the general composition of arctic fungal communities and how the community changes together with their environment, such as the availability of the carbon substrates. Importantly, it also expands the database for fungal ITS sequences with a large number of previously unencountered sequences, widening the knowledge and database available for studying fungal diversity in undersampled biomes. Additionally, this dataset can be useful for studies that explore the Arctic fungal taxonomy and their geographic distribution.

## Methods

**Study sites.** We sampled ponds in the following five sites representing different regional-scale permafrost integrity: Toolik, Alaska, USA; Qeqertarsuaq, Disko Island, Greenland, Denmark; Whapmagoostui-Kuujjuarapik, Nunavik, Quebec, Canada; Abisko, Sweden and Khanymey, Western Siberia, Russia (Online-only Table 1). The aim was to include representatives of different stages of permafrost thaw in order to understand whether responses can be generalized across different geographic and environmental conditions.

The sampling site in Alaska is located in a continuous permafrost area, mostly dominated by moss-tundra characterized by tussock-sedge *Eriophorum vaginatum* and *Carex bigelowii*, and dwarf-shrub *Betula nana* and

*Salix pulchra*<sup>15</sup>. The average depth of the active layer in 2017 was ~50 cm<sup>16</sup>. Records of surface air temperature from 1989 to 2014 showed no significant warming trend, and there was no significant increase in the mean maximum thickness of the active layer or maximum thaw depth<sup>17</sup>.

The sampling site in Greenland is located in the Blæsedalen Valley, south of Disko Island, and is characterized as a discontinuous permafrost area. From 1991 to 2011, Hollensen *et al.*<sup>18</sup> observed an increase of the mean annual air temperatures of 0.2 °C per year in the area, while Hansen *et al.*<sup>19</sup> highlighted that sea ice cover reduced 50% from 1991 to 2004. Soil temperatures recorded by the Arctic Station from the active layer of the coarse marine stratified sediments also showed an increase over the years<sup>18</sup>. The sampling site is comprised of wet sedge tundra, and the dominating species are *Carex rariflora*, *Carex aquatilis*, *Eriophorum angustifolium*, *Equisetum arvense*, *Salix arctophila*, *Tomentypnum nitens* and *Aulacomnium turgidum*<sup>20</sup>.

The Canadian site is located within a sporadic permafrost zone, in a palsa bog, in the valley of Great Whale river, close to the river mouth to Hudson Bay. The vegetation consists of a coastal forest tundra, dominated by the species *Carex* sp. and *Sphagnum* sp.<sup>21</sup> Since the mid-1990s, there has been a significant increase in the surface air temperature of the region for spring and fall, which has been correlated to a decline of sea ice coverage in Hudson Bay<sup>22</sup>. This area has experienced an accelerated thawing of the permafrost over the past decades, resulting in the collapse of palsas and the emergence of thermokarst ponds as well as significant peat accumulation<sup>21,23</sup>. In this specific site, thermokarst ponds at different development stage can be found, from recently emerging to older, mature thermokarstic waterbodies. The stage of the ponds was estimated based on the distance between the pond and the edge of the closest palsa, as well as based on satellite images<sup>14</sup>. The edges of the emerging ponds reached a maximum of 1 m from the closest palsa and were less than 0.5 m deep, whereas the edges of the developing ponds had a maximum distance of 2–3 m to the closest palsa and were ~1 m deep. Mature ponds were identified based on satellite images and were up to 60 years old.

The Swedish site is located in a discontinuous permafrost zone at the Stordalen palsa mire, on an area of collapsed peatland affected by active thermokarst. The region has experienced an increase in mean annual air temperature and active layer thickness since the 1980s, which has been followed by a shift to wetter conditions<sup>24</sup>. The vegetation found on the surface of the palsa depressions of Stordalen mire is dominated by sedges (*Eriophorum vaginatum*, *Carex* sp.) and mosses (*Sphagnum* sp.)<sup>24,25</sup>.

The Russian site is located in a discontinuous permafrost area in Western Siberia Lowland, near Khanymey village. The sampling site is a flat frozen palsa bog with a peat depth no more than 2 m, and is affected by active thermokarst, resulting in the emergence of thermokarst ponds<sup>26,27</sup>. The vegetation is dominated by lichens (*Cladonia* sp.), shrubs (*Ledum palustre*, *Betula nana*, *Vaccinium vitis-idaea*, *Andromeda polifolia*, *Rubus chamaemorus*) and mosses (*Sphagnum* sp.)<sup>28</sup>.

**Sample collection.** At all sites, water from the depth of 10 cm was collected from 12 ponds, totaling 60 ponds for the full dataset. Unfiltered water samples were collected for total P analysis. For analyzing Fe, various dissolved anions and cations, DOC concentrations, and perform optical and mass spectrometry analyses on DOM, water was filtered through GF/F glass fiber filters (0.7 µm, 47 mm, Whatman plc, Maidstone, United Kingdom). Moreover, water samples were collected in order to measure GHG (CO<sub>2</sub> and CH<sub>4</sub>) concentrations. Water, detritus and sediment samples were also collected from ponds for fungal community analyses. Water samples were collected and filtered sequentially first through 5 µm Durapore membrane filter (Millipore, Burlington, Massachusetts, USA) and then through a 0.22 µm Sterivex filter (Millipore) to capture fungal cells of different sizes. The samples were filtered until clogging or up to a maximum of 3.5 liters (filtered volume ranging from 0.11 to 3.5 l). Surface sediments were sampled from each of the ponds, with the exception of the Canadian site, where only one emerging and three developing ponds were sampled for sediments. From the sites in Alaska, Greenland, and Sweden, also detritus samples (dead plant material) were collected. The detritus was washed in the lab using tap water, followed by overnight incubation in 50 ml tap water to induce sporulation. The use of tap water may have added fungal spores to the samples, which should be kept in mind when using the detritus data. After the incubation, the water was filtered through a 5 µm pore size filter and the filter was stored at –20 °C.

All the samples for DNA extraction were transported to the laboratory frozen, with the exception of the Alaskan samples, which were freeze dried prior to transportation. The samples transported frozen were freeze dried prior to DNA extraction to ensure similar treatment of all samples. The samples for nutrient and carbon measurements were transported frozen with the exception of samples for DOC and fluorescence analyses, which were transported cooled.

**Chemical analyses.** All chemical, optical and mass spectrometry results are provided in OSF<sup>29</sup>. DOC quantification was carried out using a carbon analyzer (TOC-L + TNM-L, Shimadzu, Kyoto, Japan). Accuracy was assessed using EDTA at 11.6 mg C/l as a quality control (results were within + – 5%) and the standard calibration range was of 2–50 mg C/l. Fe(II) and Fe(III) were determined by using the ferrozine method<sup>30</sup>, but instead of reducing Fe(III) with hydroxylamine hydrochloride, ascorbic acid was used<sup>31</sup>. Absorbance was measured at 562 nm on a spectrophotometer (UV/Vis Spectrometer Lambda 40, Perkin Elmer, Waltham, Massachusetts, USA). The samples were diluted with milli-Q water if needed. The concentration of total P was determined using persulfate digestion<sup>32</sup>. The anion NO<sub>3</sub><sup>–</sup> was measured on a Metrohm IC system (883 Basic IC Plus and 919 Autosampler Plus; Riverview, Florida, USA). NO<sub>3</sub><sup>–</sup> were separated with a Metrosep A Supp 5 analytical column (250 × 4.0 mm) which was fit with a Metrosep A Supp 4/5 guard column at a flow rate of 0.7 ml/min, using a carbonate eluent (3.2 mM Na<sub>2</sub>CO<sub>3</sub> + 1.0 mM NaHCO<sub>3</sub>). SO<sub>4</sub><sup>2–</sup> was analyzed using Metrohm IC system (883 Basic IC Plus and 919 Autosampler Plus, Riverview), NH<sub>4</sub><sup>+</sup> spectrophotometrically as described by Solórzano<sup>33</sup>, and NO<sub>2</sub><sup>–</sup> and DN as in Greenberg *et al.*<sup>34</sup>.

For the gas analyses, samples from Alaska and Canada were taken as previously described in Kankaala *et al.*<sup>35</sup>, except that room air was used instead of N<sub>2</sub> for extracting the gas from the water. Shortly, 30 ml of water was taken

into 50 ml syringes, which were warmed to room temperature prior to extraction of the gas. To each syringes 0.5 ml of  $\text{HNO}_3$  and 10 ml of room air was added and the syringes were shaken for 1 min. Finally, the volumes of liquid and gas phases were recorded and the gas was transferred into glass vials that had been flushed with  $\text{N}_2$  and vacuumed. For Greenland, Sweden and Russia 5 ml of water was taken for the gas samples with a syringe and immediately transferred to 20 ml glass vials filled with N and with 150  $\mu\text{L}$   $\text{H}_2\text{PO}_4$  to preserve the sample. All gas samples were measured using gas chromatography (Clarus 500, Perkin Elmer, Polyimide Uncoated capillary column 5 m x 0.32 mm, TCD and FID detector respectively).

**Optical analyses.** In order to characterize DOM, we recorded the absorbance of DOM using a UV-visible Cary 100 (Agilent Technologies, Santa Clara, California, USA) or a LAMBDA 40 UV/VIS (PerkinElmer) spectrophotometer, depending on sample origin. SUVA<sub>254</sub> is a proxy of aromaticity and the relative proportion of terrestrial versus algal carbon sources in DOM<sup>36</sup> and was determined from DOC normalized absorbance at 254 nm after applying a corrective factor based on iron concentration<sup>37</sup>. S<sub>289</sub> enlightens the importance of fulvic and humic acids related to algal production<sup>38</sup> and were determined for the intervals 279–299 nm by performing regression calculations using SciLab v 5.5.2.<sup>39</sup>

We also recorded fluorescence intensity on a Cary Eclipse spectrofluorometer (Agilent Technologies), across the excitation waveband from 250–450 nm (10 nm increments) and emission waveband of 300–560 nm (2 nm increments), or on a SPEX FluoroMax-2 spectrofluorometer (HORIBA, Kyoto, Japan), across the excitation waveband from 250–445 nm (5 nm increments) and emission waveband of 300–600 nm (4 nm increments), depending on sample origin. Based on the fluorometric scans, we constructed excitation–emission matrices (EEMs) after correction for Raman and Raleigh scattering and inner filter effect<sup>40</sup>. We calculated the FI as the ratio of fluorescence emission intensities at 450 nm and 500 nm at the excitation wavelength of 370 nm to investigate the origin of fulvic acids<sup>41</sup>. Higher values (~1.8) indicate microbial derived DOM (autochthonous), whereas lower values (~1.2) indicate terrestrial derived DOM (allochthonous), from plant or soil<sup>42</sup>. HIX is a proxy of the humic content of DOM and was calculated as the sum of intensity under the emission spectra 435–480 nm divided by the peak intensity under the emission spectra 300–445 nm, at an excitation of 250 nm. Higher values of HIX indicate more complex, higher molecular weight, condensed aromatic compounds<sup>43,44</sup>. BIX emphasizes the relative freshness of the bulk DOM and was calculated as the ratio of emission at 380 nm divided by the emission intensity maximum observed between 420 and 436 nm at an excitation wavelength of 310 nm<sup>45</sup>. High values (>1) are related to higher proportion of more recently derived DOM, predominantly originated from autochthonous production, while lower values (0.6–0.7) indicate lower production and older DOM<sup>42,44</sup>.

**High resolution mass spectrometry.** 50 ml water samples were collected from each of the ponds and were filtered with a Whatman GF/F filter for mass spectrometry analyses. For each sample, 1.5 ml of water was dried completely with a vacuum drier, and was then re-dissolved in 100  $\mu\text{L}$  20% acetonitrile, 80% water with three added compounds as internal standards (Hippuric acid, glycyrrhetic acid and capsaicin, all at 400 ppb v/v). Samples were filtered to an autosampler vials and injected at 50  $\mu\text{L}$  onto the column. In order not to overload the detectors, some of the higher concentration samples were injected at a lower volume, to give a maximum of 20  $\mu\text{g}$  carbon loaded.

High-performance liquid chromatography – high resolution mass spectrometry (ESI-HRMS) was conducted as described in Patriarca *et al.*<sup>46</sup> using a C18-Evo column (100 × 2.1 mm, 2.6  $\mu\text{m}$ ; Phenomenex, Torrance, California, USA). The ESI-HRMS data was averaged from 2–17 min to allow formula assignment to a single mass list. Formulas considered had masses 150–800 m/z, 4–50 carbon (C) atoms, 4–100 hydrogen (H) atoms, 1–40 oxygen (O) atoms, 0–1 nitrogen (N) atoms and 0–1 13 C atoms. Formulas were only considered if they had an even number of electrons, H/C 0.3–2.2 and O/C ≤ 1. The data are presented as a number of assigned formulas and weighted average O/C ratio, H/C ratio and m/z.

The analysis was run in two batches (36 and 24 samples per run, respectively) and to the latter run, three samples of Suwannee River fulvic acid (SRFA, reference material) were added. At the moment of the run, the DOC concentration of these samples was unknown, so 50  $\mu\text{L}$  was injected. From high resolution mass spectrometry, average H/C and a number of assigned formulas were obtained. The H/C can be used as a proxy of DOM aliphatic content; higher H/C values (> 1) indicate more saturated (aliphatic) compounds, whereas values lower than 1 indicate more unsaturated, aromatic molecules<sup>47</sup>.

**DNA extraction, ITS2 amplification and sequencing.** All samples for molecular analyses (water and detritus filters and sediments) were extracted using DNeasy PowerSoil® kit (Qiagen, Hilden, Germany), following the manufacturer's recommendations for low input DNA. Extracts were eluted in 100  $\mu\text{l}$  of Milli-Q water and DNA concentrations were measured with Qubit dsDNA HS kit. The fungal ribosomal internal transcribed spacer 2 (ITS2) sequences were amplified using a modified ITS3 Mix2 forward primer from Tedesco<sup>48</sup>, named ITS3-mkmix2 CAWCGATGAAGAACGCGAG, and a reverse primer ITS4 (equimolar mix of cwmix1 TCCTCCGCTTAyTgATAtGc and cwmix2 TCCTCCGCTTAtTrATAtGc)<sup>14</sup>. Each sample received a unique combination of primers containing identification tags generated by Barcraw<sup>19</sup>. All tags had a minimum base difference of 3 and a length of 8 nucleotides. Both forward and reverse primer tags were extended by two terminal bases (CA) at the ligation site to avoid bias during ligation of sequencing adaptors, and the forward primer tag also had a linker base (T) added to it<sup>50</sup>. The list of primers and tags is found in Supplementary Table S1. PCR reactions were performed on a final volume of 50  $\mu\text{l}$ , with an input amount of DNA ranging from 0.07 ng to 10 ng, 0.25  $\mu\text{M}$  of each primer, 200  $\mu\text{M}$  of dNTPs, 1U of Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, Massachusetts, USA), 1X Phusion™ HF Buffer (1X buffer provides 1.5 mM MgCl<sub>2</sub>, Thermo Fisher Scientifics) and 0.015 mg of BSA. PCR conditions consisted of an initial denaturation cycle at 95 °C for 3 min, followed by 21–35 cycles for amplification (95 °C for 30 sec, 57 °C for 30 sec and 72 °C for 30 sec), and final extension

at 72 °C for 10 min. In order to reduce PCR bias, all samples (in duplicates) were first submitted to 21 amplification cycles. In case of insufficient yield, the number of cycles was increased up to 35 cycles (see the records on the number of cycles for each of the samples in Supplementary Table S2).

The PCR products were purified with Sera-Mag™ beads (GE Healthcare Life Sciences, Marlborough, Massachusetts, USA), visualized on a 1.5% agarose gel and quantified using Qubit dsDNA HS kit. The purified PCR products were randomly allocated into three DNA pools (20 ng of each sample), which were purified with E.Z.N.A.® Cycle-Pure kit (Omega Bio-Tek, Norcross, Georgia, USA). Nine of the samples (4 water, 1 sediment and 4 detritus) were left out of the pools because of too little PCR product, giving a total of 203 samples for sequencing (Online-only Table 1). Negative PCR controls were added to each pool, as well as a mock community sample containing 10 different fragment sizes from the ITS2 region of a chimera of *Heterobasidium irregular* and *Lophium mytilinum*, ranging from 142 to 591 bases, as described by Castaño *et al.*<sup>51</sup>. The size distribution and quality of all the pools were verified with BioAnalyzer DNA 7500 (Agilent Technologies), and purity was assessed by spectrophotometry (OD 260:280 and 260:230 ratios) using NanoDrop (Thermo Fisher Scientific). The libraries were sequenced at Science for Life Laboratory (Uppsala University, Sweden), on a Pacific Biosciences Sequel instrument II, using 1 SMRT cell per pool. This PacBio technology allows the generation of highly accurate reads (>99% accuracy) which are produced based on a consensus sequence after a circularization step.

**Quality filtering of reads, clustering and taxonomy identification of clusters.** The sequencing resulted in a total of 1071489 sequences, ranging from 397 to 9184 sequences per sample (average on 2551 sequences per sample). The raw sequences were filtered for quality and clustered using the SCATA pipeline (<https://scata.mykopat.slu.se/>), accessed on May 19<sup>th</sup>, 2020). For quality filtering, sequences from each pool were screened for the primers and tags, requiring a minimum of 90% match for the primers and a 100% match for the tags. Reads shorter than 100 bp were removed, as well as reads with a mean quality lower than 20, or containing any bases with a quality lower than 7. After this filtering, 582234 sequences were retained in the data. The sequences were clustered at the species level by single-linkage clustering at a clustering distance of 1.5%, with penalties of 1 for mismatch, 0 for gap open, 1 for gap extension, and 0 for end gaps. Homopolymers were collapsed to 3 and unique genotypes across all pools were removed. For a preliminary taxonomy affiliation of the clusters, hereafter called OTUs (Operational Taxonomic Units), sequences from the UNITE + INSD dataset for Fungi<sup>52</sup> database were included in the clustering process. After the clustering, the data included 518128 sequences, divided among 8218 OTUs. For taxonomical annotation, all OTUs with a minimum of ten total reads in the full dataset were included, retaining 3108 OTUs and 498414 sequences in the taxonomical analysis.

## Data Records

The raw sequences are deposited in the NCBI SRA database under accession number PRJNA701021 (Biosample accession numbers SAMN17843604-SAMN17843806)<sup>53</sup>. All the raw mass spectrometry data is available at the Mass Spectrometry Interactive Virtual Environment (MassIVE) under the accession number MSV000086952<sup>54</sup>. The fungal OTU sequences and their taxonomic classification, as well as all the environmental data related to the samples are deposited in Open Science Framework (OSF)<sup>29</sup>.

## Technical Validation

The DNA extractions were done in a laminar flow hood with a UV-C lamp and handled separately for each of the locations to avoid any possible cross-contamination between the samples or sites. For the PCRs, the samples were first randomized into three groups including samples from all locations to minimize the risk of batch effect at the sequencing step. Negative controls were included in the PCR step and added to the pools. The negative controls created 4, 12 and 4 sequences for pools 1, 2 and 3, respectively. For each pool, 100 ng of a positive control containing mock communities (as described in the methods section) was added. The mock communities captured all different fragment sizes. The sequences were controlled rigorously by removing singletons and rare reads, as described in the methods section, to remove any sequences that might be chimeric reads or a result of erroneous amplification. Finally, the sequences were compared against multiple databases and their taxonomy was verified manually to ensure that only fungal sequences were included to the tree shown in Fig. 2. For the resulting data, correlations between the number of OTUs and filtered water volume were checked to verify that the differential filtering volumes did not introduce a bias to the richness of the communities (Supplementary Figure S1).

## Usage Notes

To verify that the acquired OTUs were of fungal origin, we scrutinized all the OTUs with a minimum of ten total reads for their phylogenetic origin, retaining 3108 OTUs and 498414 sequences in the analysis. For taxonomic annotation, the Protax-Fungi<sup>55</sup> and massBLASTer analyses available through the Pluto-F platform of the UNITE database (<https://plutof.ut.ee/>), accessed on May 23<sup>rd</sup> 2020) were used. An OTU was assigned to a taxonomic level if the Protax identification probability was at least 95% and matched the taxonomy based on the massBLAST against the UNITE species hypothesis (SH) database. To support the taxonomic identification and to discard all non-fungal OTUs, a phylogenetic tree was built, which included all the OTUs with at least 50 reads in total. For building the tree, the ITS2 region was extracted with ITSx<sup>56</sup>, aligned with MUSCLE<sup>57</sup> and a Neighbor-Joining phylogenetic tree was constructed using MEGA7<sup>58</sup>, with a p-distance, bootstrap of 1000 replicates, gamma distributed rates and gaps treated as pairwise deletions. As reference sequences, 29 different eukaryotes were selected from the ITS2 database<sup>59</sup> (<http://its2.bioapps.biozentrum.uni-wuerzburg.de/>), accessed on August 18<sup>th</sup> 2020 – accession numbers: AB084092, AY752993, HQ219352, GQ402831, AF228083, AF053158, AF163102, JN113133, AF353997, GU001158, JQ340345, AM396560, FJ946912, EU812490, KJ925151, KF772413, JX988759, AF315074, AJ400496, AJ566147, AY458037, AY479922, EF060369, FN397599, JF750409, KF524372, U65485, Z48468, AY499004). Eight references were further derived from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), accessed on

September 29<sup>th</sup> 2020 - accession numbers: KC357673, AF508774, AY676020, MN158348, HM161704, AY264773, AB906385, AJ296818). These sequences cover different eukaryote lineages: Centroheliozoa, Choanoflagellida, Ichthyosporea, Oomycetes, Streptophyta, Chlorophyta, Rhodophyta, Cercozoa, Amoebozoa, Apusozoa, Cryptophyta, Haptophyceae, Heterolobosea, Katablepharidophyta, Arthropoda, Picozoa, Alveolata, Cnidaria, Stramenopiles, Protostomia and Porifera. Additionally, 530 fungal sequences (SHs from UNITE database) were selected to represent different fungal lineages and included the closest SH matches to each of the OTUs. Finally, to further establish higher level taxonomy, BLASTn searches (E-value cutoff of 1e-3, against the NCBI nucleotide database (<https://www.uppmx.uu.se/resources/databases/blast-databases/>) were run on September–October 2020, excluding uncultured organisms and environmental samples), for all the OTUs. The (at least) 10 best hits from the BLASTn search were evaluated as follows: an OTU was classified as a “likely fungi” if all resulting hits would match fungal sequences. The OTUs that had also matched other eukaryotic sequences were checked manually, and classified as a “likely fungi” if the best hits were fungal sequences, and also based on the quality of the alignments (higher query coverage and identity %). Any OTU that could not be classified as a “likely fungi” from the BLASTn searches and/or would cluster with any other eukaryote reference in the phylogenetic tree was discarded. The final taxonomic assignment of all fungal OTUs was based on matches to UNITE SHs and Protax probabilities. OTUs with undetermined phylum, class or order were assigned to higher taxonomic levels whenever supported by the phylogenetic tree (considering a minimal bootstrap value of 70%) or BLASTn results (min of 90% query coverage and 90% identity for class, and 100% query coverage and min 97% identity for order). The final data set with 1334 OTUs (178531 sequences) identified as fungi is presented in Supplementary Figure S2.

## Code availability

The code used for the extraction of ITS2 regions using ITSx and for the Blastn searches are provided in Supplementary Note 1.

Received: 9 March 2021; Accepted: 29 July 2021;

Published online: 19 August 2021

## References

- Schuur, E. A. G. *et al.* Climate change and the permafrost carbon feedback. *Nature*. **520**, 171–179 (2015).
- Teufel, B. & Sushama, L. Abrupt changes across the Arctic permafrost region endanger northern development. *Nature Climate Change*. **9**, 858–862 (2019).
- Vonk, J. E. *et al.* Reviews and syntheses: Effects of permafrost thaw on Arctic aquatic ecosystems. *Biogeosciences*. **12**, 7129–7167 (2015).
- Schuur, E. A. G. *et al.* Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle. *BioScience*. **58**, 701–714 (2008).
- McGuire, A. D. *et al.* Sensitivity of the carbon cycle in the Arctic to climate change. *Ecol. Monogr.* **79**, 523–555 (2009).
- Negandhi, K. *et al.* Small Thaw Ponds: An Unaccounted Source of Methane in the Canadian High Arctic. *PLoS One*. **8**, e78204 (2013).
- Negandhi, K., Laurion, I. & Lovejoy, C. Temperature effects on net greenhouse gas production and bacterial communities in arctic thaw ponds. *FEMS Microbiol. Ecol.* **92**, 117 (2016).
- Peura, S. *et al.* Ontogenetic succession of thermokarst thaw ponds is linked to dissolved organic matter quality and microbial degradation potential. *Limnol. Oceanogr.* **65**, S248–S263 (2020).
- Crevecoeur, S., Vincent, W. F., Comte, J. & Lovejoy, C. Bacterial community structure across environmental gradients in permafrost thaw ponds: Methanotroph-rich ecosystems. *Front. Microbiol.* **6**, 192 (2015).
- Vigneron, A., Cruaud, P., Bhiry, N., Lovejoy, C. & Vincent, W. F. Microbial community structure and methane cycling potential along a Thermokarst pond-peatland continuum. *Microorganisms*. **7**, 486 (2019).
- De Boer, W., Folman, L. B., Summerbell, R. C. & Boddy, L. Living in a fungal world: Impact of fungi on soil bacterial niche development. *FEMS Microbiology Reviews*. **29**, 795–811 (2005).
- Fabian, J., Zlatanovic, S., Mutz, M. & Premke, K. Fungal-bacterial dynamics and their contribution to terrigenous carbon turnover in relation to organic matter quality. *ISME J.* **11**, 415–425 (2017).
- Collado, S., Oulego, P., Suárez-Iglesias, O. & Díaz, M. Biodegradation of dissolved humic substances by fungi. *Applied Microbiology and Biotechnology*. **102**, 3497–3511 (2018).
- Wurzbacher, C., Nilsson, R. H., Rautio, M. & Peura, S. Poorly known microbial taxa dominate the microbiome of permafrost thaw ponds. *ISME J.* **11**, 1938–1941 (2017).
- Walker, D. A. & Maier, H. A. Vegetation in the vicinity of the Toolik Field Station, Alaska. Biological Papers of the University of Alaska. **28**, 12–14 (2020).
- Shaver, G. & Rastetter, E. Late season thaw depth measured in the Arctic Long Term Ecological Research (ARC LTER) moist acidic tussock experimental plots at Toolik Field station, Alaska Arctic 1993 to 2018. *Environmental Data Initiative* <https://doi.org/10.6073/pasta/e84452e07fe88c29e1ca4606adcd2702> (2019).
- Hobbie, J. E. *et al.* Ecosystem responses to climate change at a Low Arctic and a High Arctic long-term research site. *Ambio*. **46**, 160–173 (2017).
- Hollesøen, J. *et al.* Winter warming as an important co-driver for *Betula nana* growth in western Greenland during the past century. *Glob. Chang. Biol.* **21**, 2410–2423 (2015).
- Hansen, B. U., Elberling, B., Humlum, O. & Nielsen, N. Meteorological trends (1991–2004) at Arctic Station, Central West Greenland (69°15'N) in a 130 years perspective. *Geogr. Tidsskr.* **106**, 45–55 (2006).
- Christiansen, C. T. *et al.* Enhanced summer warming reduces fungal decomposer diversity and litter mass loss more strongly in dry than in wet tundra. *Glob. Chang. Biol.* **23**, 406–420 (2017).
- Bhiry, N. *et al.* Environmental change in the Great Whale River region: Five decades of multidisciplinary research by Centre d'études nordiques (CEN). *Écoscience*. **18**, 182–203 (2011).
- Hochheim, K. P. & Barber, D. G. An Update on the Ice Climatology of the Hudson Bay System. *Arctic, Antarct. Alp. Res.* **46**, 66–83 (2014).
- Payette, S., Delwaide, A., Caccianiga, M. & Beauchemin, M. Accelerated thawing of subarctic peatland permafrost over the last 50 years. *Geophys. Res. Lett.* **31**, 18 (2004).
- Swindles, G. T. *et al.* The long-term fate of permafrost peatlands under rapid climate warming. *Sci. Rep.* **5**, 1–6 (2015).
- Olefeldt, D. *et al.* Net carbon accumulation of a high-latitude permafrost palsa mire similar to permafrost-free peatlands. *Geophys. Res. Lett.* **39**, 3 (2012).
- Pokrovsky, O. S. *et al.* Thermokarst lakes of Western Siberia: a complex biogeochemical multidisciplinary approach. *Int. J. Environ. Stud.* **71**, 733–748 (2014).

27. Pokrovsky, O. S., Shirokova, L. S., Kirpotin, S. N., Kulizhsky, S. P. & Vorobiev, S. N. Climate of the Past Geoscientific Instrumentation Methods and Data Systems Impact of western Siberia heat wave 2012 on greenhouse gases and trace metal concentration in thaw lakes of discontinuous permafrost zone. *Biogeosciences*. **10**, 5349–5365 (2013).
28. Loiko, S. *et al.* Microtopography Controls of Carbon and Related Elements Distribution in the West Siberian Frozen Bogs. *Geosciences*. **9**, 291 (2019).
29. Kluge, M. Arctic ponds. *Open Science Framework* <https://doi.org/10.17605/OSF.IO/HVWPYU> (2021).
30. Viollier, E., Inglett, P. W., Hunter, K., Roychoudhury, A. N. & Van Cappellen, P. The ferrozine method revisited: Fe(II)/Fe(III) determination in natural waters. *Appl. Geochemistry*. **15**, 785–790 (2000).
31. Verschoor, M. J. & Molot, L. A. A comparison of three colorimetric methods of ferrous and total reactive iron measurement in freshwaters. *Limnol. Oceanogr. Methods*. **11**, 113–125 (2013).
32. Menzel, D. H. & Corwin, N. The measurement of total phosphorus in seawater based on the liberation of organically bound fractions by persulfate oxidation. *Limnol. Oceanogr.* **10**, 280–282 (1965).
33. Solórzano, L. Determination of ammonia in natural waters by the phenolphthalein method. *Limnol. Oceanogr.* **14**, 799–801 (1969).
34. Greenberg, A. E., Clesceri, L. S. & Eaton, A. D. *Standard Methods for the Examination of Water and Wastewater*. (American Public Health Association, 1992).
35. Kankaala, P., Taipale, S., Nykänen, H. & Jones, R. I. Oxidation, efflux, and isotopic fractionation of methane during autumnal turnover in a polyhumic, boreal lake. *J. Geophys. Res.* **112**, G02033 (2007).
36. Weishaar, J. L. *et al.* Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environ. Sci. Technol.* **37**, 4702–4708 (2003).
37. Poulin, B. A., Ryan, J. N. & Aiken, G. R. Effects of iron on optical properties of dissolved organic matter. *Environ. Sci. Technol.* **48**, 10098–10106 (2014).
38. Loiselle, S. A. *et al.* Variability in photobleaching yields and their related impacts on optical conditions in subtropical lakes. *J. Photochem. Photobiol. B Biol.* **95**, 129–137 (2009).
39. Enterprises, S. Scilab: free and open source software for numerical computation (OS, Version 5.5.2.). (2020).
40. Murphy, K. R. *et al.* Measurement of dissolved organic matter fluorescence in aquatic environments: An interlaboratory comparison. *Environ. Sci. Technol.* **44**, 9405–9412 (2010).
41. McKnight, D. M. *et al.* Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* **46**, 38–48 (2001).
42. Fellman, J. B., Hood, E. & Spencer, R. G. M. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnol. Oceanogr.* **55**, 2452–2462 (2010).
43. Ohno, T., Chorover, J., Omoike, A. & Hunt, J. Molecular weight and humification index as predictors of adsorption for plant- and manure-derived dissolved organic matter to goethite. *Eur. J. Soil Sci.* **58**, 125–132 (2007).
44. Huguet, A. *et al.* Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Org. Geochem.* **40**, 706–719 (2009).
45. Parlanti, E., Wörz, K., Geoffroy, L. & Lamotte, M. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Org. Geochem.* **31**, 1765–1781 (2000).
46. Patriarca, C. *et al.* Investigating the Ionization of Dissolved Organic Matter by Electrospray. *Anal. Chem.* **92**, 14210–14218 (2020).
47. Riedel, T. *et al.* Molecular Signatures of Biogeochemical Transformations in Dissolved Organic Matter from Ten World Rivers. *Front. Earth Sci.* **4**, 85 (2016).
48. Tedersoo, L. *et al.* Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys*. **10**, 1–43 (2015).
49. Frank, D. N. BARCRAWL and BARTAB: Software tools for the design and implementation of barcoded primers for highly multiplexed DNA sequencing. *BMC Bioinformatics*. **10**, 362 (2009).
50. Clemmensen, K. E., Ihrmark, K., Durling, M. B. & Lindahl, B. D. Sample preparation for fungal community analysis by high-throughput sequencing of barcode amplicons. in *Methods in Molecular Biology* vol. 1399 61–88 (Humana Press Inc., 2016).
51. Castaño, C. *et al.* Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities. *New Phytol.* **228**, 1149–1158 (2020).
52. Abarenkov, K. *et al.* Full UNITE+INSD dataset for Fungi. *UNITE Community* <https://doi.org/10.15156/BIO/786372> (2020).
53. NCBI Sequence Read Archive <https://identifiers.org.ncbi/bioproject:PRJNA701021> (2021).
54. Kluge, M. MassIVE MSV000086952 - High resolution mass spectrometry of Arctic ponds. *MassIVE* <https://doi.org/10.25345/C5VF7D> (2021).
55. Abarenkov, K. *et al.* Protax-fungi: a web-based tool for probabilistic taxonomic placement of fungal internal transcribed spacer sequences. *New Phytol.* **220**, 517–525 (2018).
56. Bengtsson-Palme, J. *et al.* 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 Ecol. Evol.* **4**, 914–919 (2013).
57. Edgar, R. C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. **5**, 113 (2004).
58. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
59. Merget, B. *et al.* The ITS2 Database. *J. Vis. Exp.* 1–5 <https://doi.org/10.3791/3806> (2012).

## Acknowledgements

We thank the staff at the Toolik Field Station, Arctic Station, CEN W-K Station, Abisko Research Station and Khanymey Station, as well as Anelise Kluge, Gaëtan Martin and Martin Petersson for their assistance with the samplings. We offer our gratitude to the Cree and Inuit communities in Whapmagoostui-Kuujjuarapik for giving us access to their ancestral lands. Katarina Ihrmark, Maria Jonsson, Christoffer Bergvall and Anna Ferguson are acknowledged for their assistance with the laboratory analyses. We are grateful to the INTERACT and Science for Life Laboratory for funding. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2020-5-196.

## Author contributions

M.K. planned the samplings together with C.W. and S.P., did the molecular laboratory work and sequence analyses and wrote the first draft of the manuscript together with S.P. C.W. planned the study together with S.P., planned the sampling together with M.K. and S.P., and participated in the collection of the samples. M.W. participated in the collection of the samples, analyzed carbon samples and data. K.E.C. instructed the amplicon processing. J.H. analyzed carbon samples and data. K.E. analyzed carbon samples and data. J.S. helped with the study design and data analysis. S.P. planned the study together with C.W., planned the samplings together with C.W. and M.K., participated in the collection of the samples and wrote the first draft of the manuscript together with M.K. All authors participated in the revision of the manuscript.

## Funding

Open access funding provided by Swedish University of Agricultural Sciences.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-021-01005-7>.

**Correspondence** and requests for materials should be addressed to M.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021