Title: Mycorrhizal community dynamics and drivers after glacier retreat across the globe

Authors: Alexis Carteron, Alessia Guerrieri, Isabel Cantera, Silvio Marta, Aurélie Bonin, Simone Giachello, Roberto Ambrosini, Marco Caccianiga, Pierre Taberlet, Guglielmina Adele Diolaiuti, Jerome Poulenard, Diego Fontaneto, Wilfried Thuiller, Gentile Francesco Ficetola
Email: alexis.carteron@unimi.it

Aim: Mountain and high-latitude glaciers have been retreating around the world since the Little Ice Age. Large areas of deglaciated substrates are exposed to soil development and there is an urgent need to predict their evolution. Ecosystem development in these nutrient-poor and harsh environments is limited, with primary production heavily dependent on mutualisms such as arbuscular mycorrhiza and ectomycorrhiza, the two most widespread plant-fungal symbioses.

Method: To explore the dynamics and drivers of mycorrhizal fungi, we conducted a comprehensive inventory of 1251 plots in 265 forelands of 46 glaciers around the globe, with sites spanning from 1 to 483 years since glacier retreat. For each plot, we assessed fungal community using metabarcoding of soil environmental DNA. For a subset of 32 glaciers, we also estimated plant community, productivity and microhabitat conditions.

Results: Both types of mycorrhizal fungi colonize the substrate a few years after the retreat of the glaciers, although with a delay compared to the whole fungal community. Diversity of arbuscular mycorrhizal fungi is largely driven by time and plant community, while microclimate and productivity influence more strongly ectomycorrhizal fungi.

Conclusions: The establishment of mycorrhizal fungi is rapid with local dynamics driven by time after glacier retreat but also plant community, productivity and microhabitat conditions. Changes in the rate of ice melt and conditions such as microclimate could disrupt biotic colonization, potentially by causing a mismatch between mycorrhizal partners which would slow soil development and associated ecological processes. Further analyses using multi-trophic surveys are needed to predict ecosystem-level impacts.

METHODS

Sample collection

From 2014 to 2020, we collected soil samples from 1251 plots within 265 sites, located in the forelands of 46 mountain and high-latitude glaciers (Fig. 1) from five continents, including regions with different climatic conditions and rates of glacier retreat (Zemp *et al.*, 2019). In these forelands, information on deglaciation dates over the past centuries is available from Marta *et al.* (2021).

In each glacier foreland, we selected three to 17 suitable sites (mean = 5.8 sites per foreland, SD = 2.5), where glacier retreat occurred from 1 to 483 years before sampling. For each site, the age since glacier retreat was used as a proxy of the time available for ecosystem development, i.e. we used a chronosequence approach for the study of ecological successions (space-for-time substitution; Walker *et al.*, 2010). At each site, we established 2-10 plots (mean = 4.7 plots, SD = 0.8), evenly spaced at a distance of 20 meters, where possible. At each plot, we collected five soil subsamples at a distance of one meter. The soil was sampled to a depth of 0-20 cm, litter was excluded, as well as plant organs. The subsamples from the same plot were pooled, resulting in a composite soil sample of ~200 g per plot. After homogenization of the composite sample, 15 g of soil were taken and placed within 6 hours in a sterile box to be dried with 40 g of silica gel. This method allows reliable preservation of eDNA (Guerrieri *et al.*, 2021). An independent soil sample was taken for soil chemistry analyses.

Local conditions

Habitat characteristics were determined at the plot level by estimating primary productivity, plant diversity, soil temperature, topographic wetness index and, for a subset of 32 glacier forelands (out of a total of 46), measured soil chemistry. Total nitrogen (N) concentration was measured for each plot by elemental analysis (Flash2000 OEA analyzer, ThermoFisher). Soil pH was measured using a pH-meter, from a suspension composed by 4 g of soil and 10 ml of bi-distilled water. According to pH, two different methods were used for the measure of assimilable phosphorus (P) through inductively coupled plasma mass spectrometry (iCAP RQ ICP-MS, ThermoFisher): the Bray and Kurtz method (Bray & Kurtz, 1945) for samples with pH < 6.5 and the Olsen method (Olsen, 1954) for samples with pH ≥ 6.5. As an indicator of primary productivity, we used the normalized difference vegetation index (NDVI), which is known to be positively related to annual aboveground net primary production (Paruelo *et al.*, 1997). Yearly maximum productivity was retrieved from the optical satellite data acquired by Sentinel-2 (ESA, COPERNICUS, S2) at 10 m resolution and averaged over the 2016-2019 period using Google Earth Engine and the *rgee* R package (Aybar *et al.*, 2022). Because proglacial areas tend to have complex topography and prolonged snow cover, yearly maxima were preferred over standard masking algorithms in order to remove the noise caused by cloudiness (Lillesand *et al.*, 2015). Plant diversity was estimated based on the plant MOTU data (see next section for details). Fine-scale subsurface soil temperature (5 cm below surface) was estimated using a global microclimatic model approach, calibrated using data-loggers placed in 175 stations from polar, equatorial and alpine glacier forelands, developed in Marta *et al.* (2022). As a proxy of potential soil moisture, we used the topographic wetness index (TWI) calculated with the *dynatop* R package (Smith & Metcalfe, 2022), based on the ASTER Global Digital Elevation Map (version 3, Abrams *et al.*, 2020) with 1 arc-second resolution (~30 m at the equator). The TWI is based on the slope and the upstream contributing area; it has been found to correlate also with other

factors than soil moisture such as plant species composition or soil pH, and its ability to predict soil moisture varies as a function of the focus environment and the algorithm used (Kopecký *et al.*, 2021), hence analysis using the TWI should be interpreted with care.

DNA sequences acquisition

The molecular and bioinformatic workflows are detailed in Guerrieri *et al.* (2022). Briefly, sequences were obtained after: (i) mixing soil samples collected at each plot with phosphate buffer (Taberlet *et al.*, 2012); (ii) extraction of eDNA using the NucleoSpin® Soil Mini Kit; (iii) amplification in four replicates through PCR, targeting the ITS1 region for fungi (marker Fung02; Epp *et al.*, 2012) and the chloroplast trnL-P6 loop for vascular plants (marker Sper01; Taberlet *et al.*, 2007) including bioinformatic blanks, extraction and amplification of negative controls, and positive controls (see below); (iv) library preparation and sequencing of purified samples using the MiSeq (fungi) and HiSeq (plants) Illumina platforms. Positive controls were used to verify the performance of amplifications and consisted of 16 non-tropical plant species belonging to 15 families (Taxaceae, Lamiaceae, Salicaceae, Polygonaceae, Betulaceae, Oleaceae, Pinaceae, Caprifoliaceae, Pinaceae, Aceraceae, Poaceae, Rosaceae, Brassicaceae, Geraniaceae, Ericaceae) and two fungal strains (*Saccharomyces cerevisiae, Cryptococcus neoformans*) at known concentrations.

The bioinformatic workflow was realized using OBITools software (Boyer *et al.*, 2016). Paired-end reads were first assembled and only sequences with an alignment score > 40 were kept and assigned to the corresponding PCR replicate before dereplication. Singletons were discarded as well as artifacts that had lower or higher length than expected (68-919 bp for fungi and 10-220 bp for plants). The remaining high-quality sequences were clustered as molecular operational taxonomic units (MOTUs) considering optimal thresholds of intra- and inter-specific variations at 95% for fungi and at 97% for plants (Bonin *et al.*, 2023). For each MOTU, taxonomic assignment was performed using a reference database constructed from EMBL (version 140). In order to limit the presence of contaminants (Ficetola *et al.*, 2015; Boyer *et al.*, 2016; Zinger *et al.*, 2019), MOTUs were not included in the analyses if they had: i) a best identity score below 80% and total count below five (based on bioinformatic blanks) for fungi; or ii) a best identity score below 90% and total count below eight for plants. In addition, MOTUs were not included if they were detected in only one PCR replicate of the same sample or in more than one extraction or amplification of negative controls (potential false positives and contaminants; Ficetola *et al.*, 2015; Zinger *et al.*, 2019). Finally, the four PCR replicates were summed to obtain the final MOTU table following the relaxed stringency method (Mächler *et al.*, 2021).

Mycorrhizal type assignation

Mycorrhizal types were assigned using the FUNGuild database (Nguyen *et al.*, 2016). From the identified genera and families, the following ones were considered as EcM fungi: *Inocybe, Austropaxillus, Cantharellus, Cenococcum, Clavulina, Cortinariaceae, Gomphidiaceae, Helvella, Lactarius, Leucophleps, Rhizopogon, Russula, Sebacinaceae, Suillus* and *Tuberaceae* (Nguyen *et al.*, 2016). For AM fungi, the following families and orders were considered: *Acaulosporaceae, Archaeosporaceae, Archaeosporales, Diversisporaceae, Diversisporales, Glomeraceae, Glomerales* and *Paraglomeraceae* (Nguyen *et al.*, 2016). Both the database and the primers used are not free of biases toward specific taxa, still the functional assignment of ITS

fungal sequences offers some of the greatest potentials in the field of fungal mycorrhizal research (Fei *et al.*, 2022; Tedersoo *et al.*, 2022; Baldrian *et al.*, 2022).

Data analyses

Alpha-diversity at the plot level was calculated as the number of MOTUs (i.e. richness) and as the Shannon diversity index (i.e. the exponential of the Shannon entropy, which also corresponds to diversity estimated using the Hill's number $q = 1$). $q = 1$ provides biodiversity estimates that are appropriate for the specificity of eDNA metabarcoding data (Calderón-Sanou *et al.*, 2020; Mächler *et al.*, 2021). We used linear mixed models to test the hypothesis that AM fungi would colonize first because their host plants are more tolerant to stressful environments. First, we quantified the difference in diversity (estimated with $q = 1$) between AM and EcM fungi. Positive values would indicate that AM communities are more diverse than EcM communities, and *vice versa*. In the mixed model, the difference in diversity was the independent variable, time was the fixed factor and glacier with site nested were random factors. Models were implemented in the *brms* package (Bürkner, 2017). The model ran on four parallel chains of length 10,000 with a burn-in of 1,000 iterations, a thinning rate of 10 and uninformative priors as provided in the *brms* package. Convergence was assessed by visually inspecting the Markov chains and considered sufficient when $\hat{R} = 1$. The absence of spatial autocorrelation was checked by visualizing spline correlograms using the *ncf* package (Bjornstad & Cai, 2022).

To assess the potential impacts of time, glacier identity and habitat (i.e. productivity, plant diversity, N, P, pH, temperature, TWI) on the patterns of AM and EcM fungal alpha-diversity, we used a random forest algorithm to fit nonlinear multiple regressions with the *randomForest* package (Cutler & Wiener, 2022). The number of bootstrap replicates (*ntree*) was set to 1,000, with convergence verified visually by assessing out-of-bag error. Variable importance was based on the increase in the mean squared error and their significance was estimated after 5000 repetitions. Plant alpha-diversity ($q = 1$) was calculated based on the plant MOTU data. For this analysis, we used data from 793 plots in 32 proglacial areas.

The potential drivers of AM and EcM fungal beta-diversity (i.e. changes in community composition between plots belonging to the same foreland, N = 2031) were assessed using the generalized dissimilarity modelling (GDM) approach with the *gdm* package (Fitzpatrick *et al.*, 2022). This approach is well suited to identify the drivers of community dissimilarity across plots and to analyse relationships potentially affected by non-linearity. Beta-diversity between the communities inhabiting different plots was related to differences in time and habitat variables, as well as geographic distances. Furthermore, as a measure of plant community changes, we computed a principal coordinates analysis (PCoA) from the plant dissimilarity matrix using the Jaccard index and used the scores of the first axis for each plot as an explanatory variable. We focused on dissimilarities between pairs of plots located in the same foreland (i.e. pairs of plots located in different forelands were removed from GDM), as the aim was to assess the factors determining community variation within each landscape. Variable significance was estimated after 1000 permutations. Plots with zero MOTUs of fungi or vascular plants were removed from GDM.

The following variables were log-transformed prior to modelling to reduce skewness: time since glacier retreat, vascular plant alpha-diversity N, P, TWI and NDVI. Additional R packages used for data wrangling and visualization included: *dplyr* (Wickham *et al.*, 2017), *ggplot2* (Wickham,

2016), *ggspatial* (Dunnington, 2018), *ggrepel* (Slowikowski *et al.*, 2021), *phyloseq* (McMurdie & Holmes, 2013), *rnaturalearth (South, 2017)*, *tidyr* (Wickham & Henry, 2019) and *vegan* (Oksanen *et al.*, 2017).

REFERENCES

Abrams M, Crippen R, Fujisada H. **2020**. ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD). *Remote Sensing* **12**: 1156.

Aybar C, Qiusheng W, Bautista L, Yali R, Barja A, Ushey K, Ooms J, Appelhans T, Allaire JJ, Tang Y, *et al.* 2022. rgee: R Bindings for Calling the 'Earth Engine' API.

Baldrian P, Bell-Dereske L, Lepinay C, Větrovský T, Kohout P. **2022**. Fungal communities in soils under global change. *Studies in Mycology*.

Bjornstad ON, Cai J. **2022**. ncf: Spatial Covariance Functions.

Bonin A, Guerrieri A, Ficetola GF. **2023**. Optimal sequence similarity thresholds for clustering of molecular operational taxonomic units in DNA metabarcoding studies. *Molecular Ecology Resources* **23**: 368–381.

Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. **2016**. obitools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources* **16**: 176–182.

Bray RH, Kurtz LT. **1945**. Determination of Total Organic and Available Forms of Phosphorus in Soils. *Soil Science* **59**: 39–46.

Bürkner P-C. **2017**. brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software* **80**: 1–28.

Calderón-Sanou I, Münkemüller T, Boyer F, Zinger L, Thuiller W. **2020**. From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? *Journal of Biogeography* **47**: 193–206.

Cutler F original by LB and A, Wiener R port by AL and M. **2022**. randomForest: Breiman and Cutler's Random Forests for Classification and Regression.

Dunnington D. **2018**. *ggspatial: Spatial Data Framework for ggplot2*. https://CRAN.R-project.org/package=ggspatial.

Epp LS, Boessenkool S, Bellemain EP, Haile J, Esposito A, Riaz T, Erséus C, Gusarov VI, Edwards ME, Johnsen A, *et al.* 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology* **21**: 1821–1833.

Fei S, Kivlin SN, Domke GM, Jo I, LaRue EA, Phillips RP. **2022**. Coupling of plant and mycorrhizal fungal diversity: its occurrence, relevance, and possible implications under global change. *New Phytologist* **234**: 1960–1966.

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, *et al.* 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* **15**: 543–556.

Fitzpatrick M, Mokany K, Manion G, Nieto-Lugilde D, Ferrier S, Lisk M, Ware C, Woolley S, Harwood T. 2022. gdm: Generalized Dissimilarity Modeling.

Guerrieri A, Bonin A, Münkemüller T, Gielly L, Thuiller W, Ficetola GF. 2021. Effects of soil preservation for biodiversity monitoring using environmental DNA. *Molecular Ecology* **n/a**.

Kopecký M, Macek M, Wild J. 2021. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Science of The Total Environment* **757**: 143785.

Lillesand T, Kiefer RW, Chipman J. 2015. *Remote Sensing and Image Interpretation, 7th Edition*. Wiley.

Mächler E, Walser J-C, Altermatt F. 2021. Decision-making and best practices for taxonomy-free environmental DNA metabarcoding in biomonitoring using Hill numbers. *Molecular Ecology* **30**: 3326–3339.

McMurdie PJ, Holmes S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE* **8**: e61217.

Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG. 2016. FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology* **20**: 241–248.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, *et al.* 2017. vegan: Community Ecology Package.

Olsen SR (Sterling R. 1954. *Estimation of available phosphorus in soils by extraction with sodium bicarbonate*. Washington, D.C. : U.S. Dept. of Agriculture.

Paruelo JM, Epstein HE, Lauenroth WK, Burke IC. 1997. ANPP Estimates from NDVI for the Central Grassland Region of the United States. *Ecology* **78**: 953–958.

Slowikowski K, Schep A, Hughes S, Dang TK, Lukauskas S, Irisson J-O, Kamvar ZN, Ryan T, Christophe D, Hiroaki Y, *et al.* 2021. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'.

Smith P, Metcalfe P. 2022. dynatop: An Implementation of Dynamic TOPMODEL Hydrological Model in R.

South A. 2017. rnaturalearth: World Map Data from Natural Earth.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. **2012**. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**: 2045–2050.

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E. **2007**. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35**: e14–e14.

Tedersoo L, Bahram M, Zinger L, Nilsson RH, Kennedy PG, Yang T, Anslan S, Mikryukov V. **2022**. Best practices in metabarcoding of fungi: From experimental design to results. *Molecular Ecology* **31**: 2769–2795.

Walker LR, Wardle DA, Bardgett RD, Clarkson BD. **2010**. The use of chronosequences in studies of ecological succession and soil development. *Journal of Ecology* **98**: 725–736.

Wickham H. **2016**. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Wickham H, Francois R, Henry L, Müller K. **2017**. *dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham H, Henry L. **2019**. *tidyr: Tidy messy data*. https://CRAN.R-project.org/package=tidyr.

Zemp M, Huss M, Thibert E, Eckert N, McNabb R, Huber J, Barandun M, Machguth H, Nussbaumer SU, Gärtner-Roer I, *et al.* 2019. Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**: 382–386.

Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, *et al.* 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology* **28**: 1857–1862.