UNIL | Université de Lausanne

Faculté de biologie et de médecine

Ecole de biologie

Can fungi diversity improve plant distribution models? An investigation in Swiss alpine grasslands

Travail de Maîtrise universitaire ès Sciences en comportement, évolution et conservation, spécialisation « géosciences, écologie et environnement »
Master Thesis of Science in Behaviour, Evolution and Conservation, specialisation « Geosciences, Ecology and Environment »

par

Clara SAND

Directeur·Directrice : Prof. Antoine Guisan
Superviseur·e·s : Ph.D. candidate Valentin Verdon
Expert·e·s : Titre, Anonyme
Département d'écologie et évolution (DEE)

January 2023

ABSTRACT

Species distribution models (SDM) use environmental factors to predict the distribution of species. Aboveground factors like climatic predictors are often used for plant SDMs, but do not always accurately capture the plants' geographical distribution. Since soil-related abiotic factors are considered as important drivers of plant distribution, we tested whether soil fungal diversity could be used to improve plant SDMs. Our expectation was that fungi diversity could capture some aspect of soil ecology, or a biotic interaction, that is important for plants, but not accounted for in SDMs today. We built SDMs of 85 alpine grassland plant species, first with only abiotic factors, and then adding the richness of various functional groups of fungi to the models. On average, the performance of the Random Forest models was clearly improved by the addition of fungal richness as a predictor, but the increase in predictive performance always stayed low. The highest increases of performance could be observed by including the richness of soil saprotrophs, lichenized fungi, or animal endosymbionts. This low increase in performance indicates that fungal richness does not significantly capture the variance of any soil-related factor that is not already included in the abiotic predictors.

RÉSUMÉ

Les modèles de distribution des espèces (SDM) utilisent des facteurs environnementaux pour prédire la distribution des espèces. Les facteurs environnementaux tels que les prédicteurs climatiques sont souvent utilisés pour modéliser la distribution des plantes, mais ne donnent pas toujours un aperçu complet de leur distribution géographique. Comme les facteurs abiotiques liés au sol sont considérés comme des éléments importants de la distribution des plantes, nous avons testé si la diversité de champignons du sol pouvait améliorer les SDM de plantes. Nous nous attendions à ce que la diversité des champignons puisse capturer un aspect de l'écologie du sol, ou une interaction biotique, qui est important pour les plantes, mais qui n'est pas pris en compte dans les SDM actuels. Nous avons construit des SDM de 85 espèces végétales de prairies alpines, d'abord en n'incluant que des facteurs abiotiques, puis en ajoutant aux modèles la richesse de divers groupes fonctionnels de champignons. En moyenne, les modèles Random Forest performaient mieux en incluant la richesse de champignons comme prédicteur, mais l'augmentation de la performance prédictive reste faible. Les augmentations de performance les plus élevées ont pu être observées en incluant la richesse en saprotrophes du sol, en champignons lichénisés ou en endosymbiontes animaux. Cette faible augmentation de la performance indique que la richesse de champignons ne capture pas de manière significative la variance d'un facteur lié au sol qui n'est pas déjà inclus dans les prédicteurs abiotiques.

INTRODUCTION

Fungi are a crucial element of soil ecology. Soil fungi are key to the carbon and nitrogen cycles through a variety of roles in organic matter decomposition (Frąc et al., 2018), and can thus shape the soil's conditions to their advantage. They also display a wide set of interactions with other living organisms, either as pathogens, parasites, or symbionts. One of the most well-known and investigated relationship between fungi and other organisms is the symbiotic mycorrhizal association with plants. Mycorrhizal fungi associate with plants' roots and facilitate access to certain nutrients, which can drive the distribution of those plants (Tedersoo, et al., 2020). This association also allows certain plant species to grow in environmental conditions that are otherwise not suitable for their survival (Püschel et al., 2020). Fungal pathogens can also restrict the habitat and spreading of plants by exploiting their resources. Even when excluding known interactions, indirect associations can be found between plants and fungi, if they share similar soil condition requirements and ecological needs, or if the fungi have a direct effect on another organism that affects some plant species' fitness. Plant distribution can thus be either extended or restricted by the fungal composition of soil.

Determining plant distribution can be done with species distribution models (SDM). SDMs use the relationship between one species and multiple environmental factors to predict the suitability of geographical areas for this species presence. With increasing availability of data on both the biological and abiotic factors needed for those models (Guisan et al., 2017), the importance of species distribution models in plant ecology has been increasing over the years (Araújo et al., 2019), for example for the assessment of species' status and threat level. The International Union for Conservation of Nature (IUCN), for example, benefits from species distribution modelling results for the threat assessment of their Red list of threatened species (Syfert et al., 2014). It plays an important part in the decision-making of numerous organizations involved in nature conservation and protection management (e.g. Wetlands International, see Breiner et al., 2022).

The reliance of those decisions on SDMs makes it important to utilize new methods to improve their performance, *i.e.* to make their predictions of the species' presence more accurate. One aspect of SDMs that can be improved is to get more relevant environmental predictors (Austin & Van Niel, 2011; Mod et al., 2020; Scherrer et al., 2017). For plant distribution models, the main environmental predictors used are usually topoclimatic data. This kind of data can be derived from satellite imagery, digital elevation models, and weather station records, and is generally measured at a large scale.

However, predictors measured at a large scale are not necessarily relevant for plant SDMs. Microclimatic changes, which are very localized, create heterogeneity in the environmental conditions that is not captured at a larger scale. Mountain ecosystems are a typical example of this heterogeneity, with steep changes in wind, temperature, soil moisture etc. that can be observed on relatively short distances (Oke & Thompson, 2015). Low-resolution/large-scale climatic predictors are thus less effective at predicting plant distribution in mountainous areas (Austin & Van Niel, 2011; Randin et al., 2009). Yet, increasing the resolution of those same topographical predictors does not considerably improve the models (Pradervand et al., 2014). In an attempt to understand how this habitat heterogeneity affects plant distributions, Buri et al. (2017) highlights the importance of including edaphic (i.e. soil-related) variables as predictors. For example pH, as well as other factors linked to the calcareous content of the soil, often comes up as an important variable to account for in order to increase plant SDM prediction capacities (Buri et al., 2020; Dubuis et al., 2013). The predicted suitability of habitat in SDMs often overlooks those relationships (Mod et al., 2020).

Another element that can be included in plant SDMs is the effect other organisms have on plant distribution, *i.e.* the biotic predictors. In theory, biotic interactions are already accounted for in species distribution modelling, as this method supposedly captured the realized niche of the species, but biotic interactions do happen to improve the prediction accuracy of plant SDMs (Zimmermann et al., 2010). Furthermore, the predominant type of plant-plant interactions has been shown to vary with elevation, with facilitation being more common at high elevation where the abiotic conditions are harsher, while competition was more common at lower elevation (Callaway et al., 2002). Some plants' distribution can be linked to those types of interactions (D'Amen et al., 2018), although their effect on SDM performance is quite low compared to the abiotic factors. Although many studies tackle the effect of plant-plant interactions on SDM results, we found little literature looking at what plant-fungi interactions bring to the table. In a pioneer study in this regard, Pellissier et al. (2013) investigated the potential of including the richness of two taxonomy classes of fungi to predict plant distribution in the Alps. They found that some models could be improved when including fungal richness, but they could not exclude that this improvement may be due to some other abiotic variable correlated with fungal richness. They also mention that a taxonomy-based approach might not be the most relevant one since fungi taxonomy does not distinguish clearly between the different ecological functions of fungi. Categorizing fungi sequences by ecological guilds (i.e. groups that harness the same resources using similar pathways) has become more popular in recent years (Nguyen et al., 2016; Zanne et al., 2020).

In this study we investigate whether some aspects of fungi diversity could improve the predictions on a wide set of alpine grassland plant species. Our hypothesis does not rely on a thoroughly investigated relationship between fungal communities and plant distribution. Instead, we focus on the information that fungal diversity could offer to explain plant distributions. We hypothesize that fungal diversity, used as a predictor for species distribution modelling, could capture some aspect of soil ecology that is important for plants, but not accounted for in SDMs today, whether it is linked to direct fungi-plant interactions or indirect fungi-soil relationships. We do not expect fungi that interact directly with plants to have a higher influence than other fungi on SDM performance. Indeed, species distribution models capture the realized niche, *i.e.* the environmental space that the species actually occupies, which accounts for both abiotic factors and biotic interactions (Guisan et al., 2017). Thus, SDMs include those biotic interactions by design.

Using 4 modelling methods, we look at the potential improvement of plant distribution models when including as predictors (a) the soil's total fungal richness, (b) the richness of different functional groups of fungi, and (c) at the relative abundance of fungal trophic groups. We first built models with only the abiotic factors, and then added the fungi predictors one at a time to see if the models were improved for most plant species. Investigating this will give us a better insight into the potential of including the soil's biotic aspect into species distribution modelling.

MATERIAL & METHODS

Area of study

The study area is located in the western Swiss Alps and covers an area of ca 700 km^2. It is a thoroughly studied region (von Däniken et al., 2014), especially in the fields of niche and distribution modelling of species and communities, e.g. species distribution models of various life forms, the structure of microbial communities in mountain grasslands (Pellissier et al., 2014; Pinto-Figueroa et al., 2019), the effect of climate change and microclimate on the suitability of habitats in a mountain environment (Giaccone et al., 2019), and the interactions of species driving the ecological range of organisms (Descombes et al., 2020).

Plant inventory & soil sampling

The data used were collected on sites ranging from 400 to 3000 meters above sea level (Figure 1B). The locations of those plots were selected via random-stratified sampling over all grasslands of the study area, covering all possible combinations of elevation, slope, and aspect

(Hirzel & Guisan, 2002). The 213 sites where both plant survey and microbial soil sampling occurred were kept for the analysis (Figure 1A).

As described in Pellissier et al (2014), exhaustive inventories of all plant species present in $4m^2$ plots were conducted between May and September, for the years 2002-2003 and 2009-2010. Only the most common species (present in at least 30 sites) were kept, *i.e.* 85 plant species from 20 families (Figure S1). The soil samples were collected between 2012 and 2013, as described in (Buri et al., 2017). For each plot, soil from the top 10cm of the ground at 5 points was sampled (4 at the corners of the square, and one in the middle). One small sample of soil per site was flash-frozen with liquid nitrogen while the rest was refrigerated and analysed within 36h after sampling.

A total of 75 environmental variables were used for the modelling (Table S1). The climatic layers were obtained from the CHclim25 dataset, and were calculated from MeteoSwiss maps generated from the weather stations close to the study area (Broennimann, 2018). The edaphic variables were collected in 2012 and 2013 (Buri et al., 2017).

A **B**

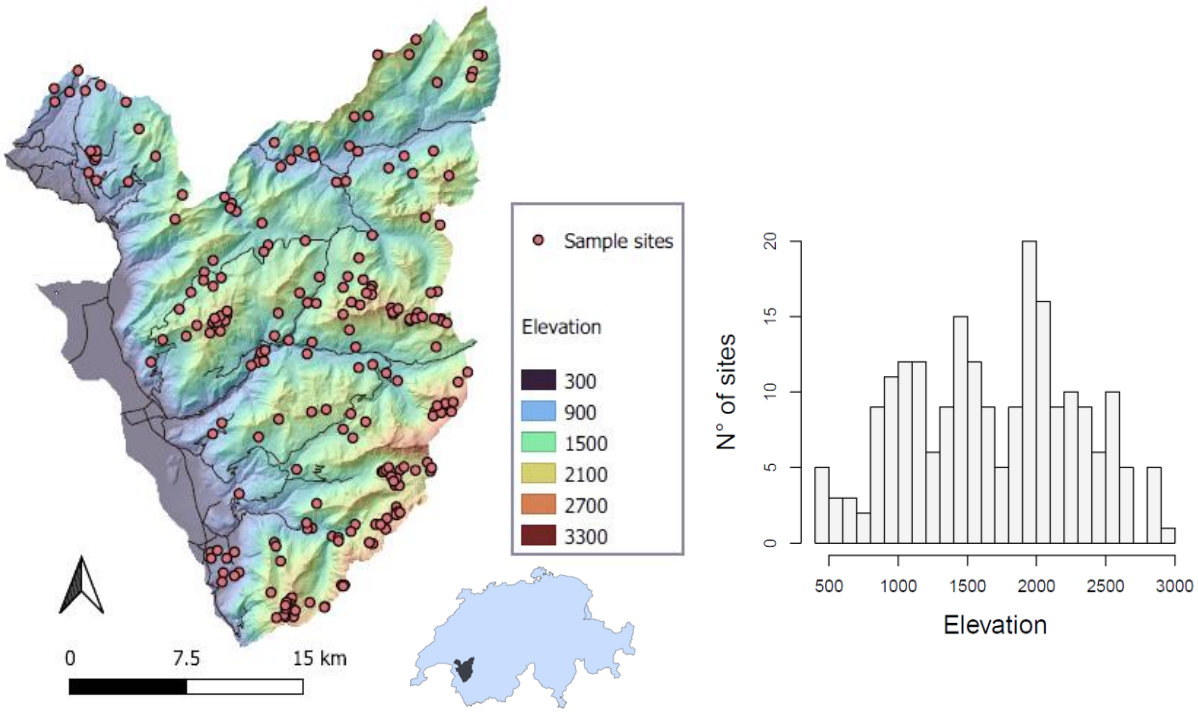


Figure 1**. Study region in the Western Swiss Alps**. *(A) Map of the area with all sample sites used for this study. The map is coloured according to elevation. (B) Histogram of distribution of sample sites across elevation.*

Soil microbiome DNA extraction, sequencing & processing

All soil samples were processed as described by Yashiro et al (2016). Briefly, 0.25 g of dry soil from each sample was used for DNA extraction using PowerSoil®-htp 96 Well Soil DNA Isolation Kit (MoBio Laboratory, Carlsbad, CA, USA). The fungi internal transcriber spacer 1 (ITS1) region was amplified using the primer ITS1F (5'-GAACCWGCGGARGGATC-3' and 5'-GCTGCGTTCTTCATCGATGC-3') and sequenced using HiSeq Illumina sequencing. Sequences with less than 100 reads in total were discarded. The raw reads were filtered to remove the short and anomalous ones, and the outer regions of the ITS1 sequences were used for the taxonomy. All reads were classified into amplicon sequence variants (ASV) using the DADA2 algorithm package in R (Callahan et al., 2016). Amplicon sequence variants, also referred to as zero-radius OTUs (zOTUs, *i.e.* OTUs with a 100% similarity threshold) (Edgar, 2018), have proven more useful than OTUs with a 97% similarity threshold to capture the actual richness of fungal communities (Pauvert et al., 2019), although they tend to overestimate the number of strains by splitting some into several groups. All ASVs found in less than 5% of the sites were discarded, leaving a total of 51 '863 ASVs.

Attribution of functional groups to the ASVs

The functions attributed to each ASV were determined using the FUNGuild algorithm, designed by Nguyen et al. (2016) and available on GitHub. FUNGuild allows to classify fungal OTUs into three major trophic groups (Saprotrophs, Symbiotrophs and Pathotrophs), and ecological guilds, which are sub-categories of the trophic groups. The categorization is based on a database of species- and genus-level annotations of fungal OTUs. FUNGuild will define the given sequence as part of one or more taxa, with levels of probability of belonging to this group, and attribute one or more trophic groups and/or guilds to it. Out of all ASVs kept for the analysis, about 52.52% had at least one function attributed to it by FUNGuild.

The following variables were then computed from the fungi data:

- Total fungal richness: the sum of fungi ASVs per site
- Fungal richness of functional groups: the sum of ASVs per functional groups (one of the major trophic groups, or one of the guilds) (Figure 2).

- Relative abundances of the three major trophic groups (Symbiotrophs, Saprotrophs & Pathotrophs). The relative abundances were computed using the counts of each ASVs from one of those categories (Figure S2). When FUNGuild attributed multiple functions to one ASV, the counts of this ASV were equally split between those functions (so for an ASV with 100 counts and attributed to both pathotroph and saprotroph, 50 counts are counted as pathotroph, and the other 50 as saprotroph).

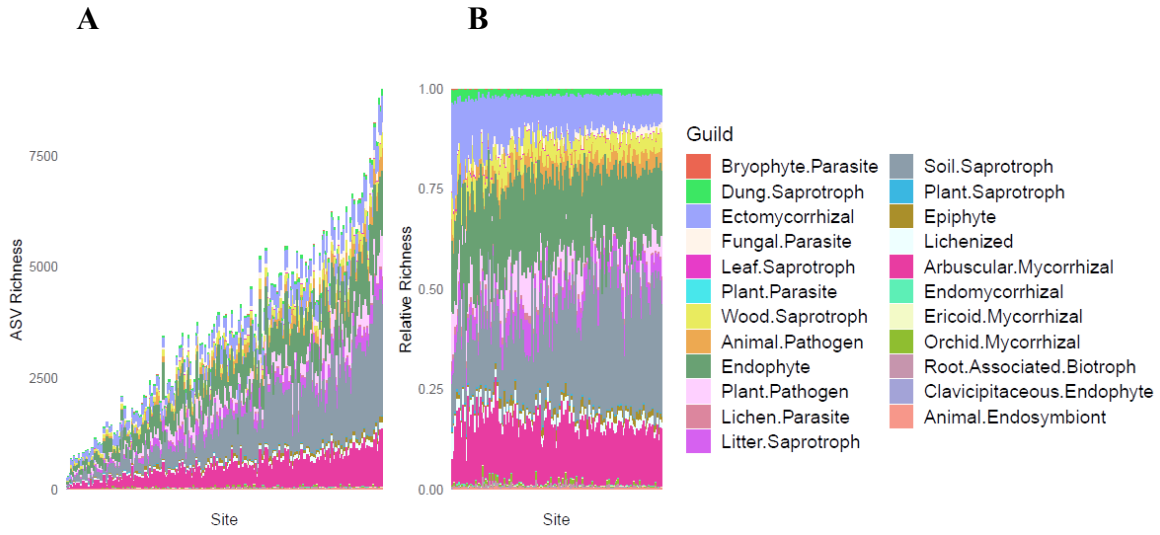


Figure 2. **Representation of fungi richness of all functional groups.** (A) *ASV richness of the ecological guilds as defined by FUNGuild, by site. Sites are arranged from left to right by increasing total richness. (B) relative richness, i.e. richness divided by the total sum of ASVs with a functional group attributed to it.*

Building and evaluating the SDMs

a) General framework

To fit model and assess the predictive power of adding a fungi predictor to plant models, we built 'base models' with only environmental predictors (reduced to synthetic axes) and then added different fungi predictors to it, one at a time (Figure 3). We repeated this for all species and with four different modelling techniques, two regression approaches (using two implementations of Generalized linear models) and two tree-based approaches (Random Forest and Gradient Boosting Machine). All model building processes and analyses were carried out with the R software (version 4.1.2).

b) *Reducing environmental dimensions*

Variables were standardized (mean=0, SD=1) and reduced to their principal component axes using dudi.pca() (from the ade4 package). Regression-based SDMs usually show the best

performances when the number of predictors is limited to a tenth of their occurrences in the data, following Harrell's rule of thumb (Harrell, 2001; Brun et al., 2020). Since the plant species considered have an average of ca. 50 occurrences in our data, and only a few are above 100 (Figure S1), keeping 10 principal component axes was a compromise to have enough variance from the abiotic predictors explained, while also saving on modelling time. This resulted in 73.3% of the total variance explained by the PC axes (Figure S3). The reduction of predictor variables into their principal components makes the interpretability of the models more difficult, since the effect size of each environmental variable cannot be clearly identified. However, it includes the variance of a high number of environmental factors without encountering any issues linked to too many predictors and multicollinearity.

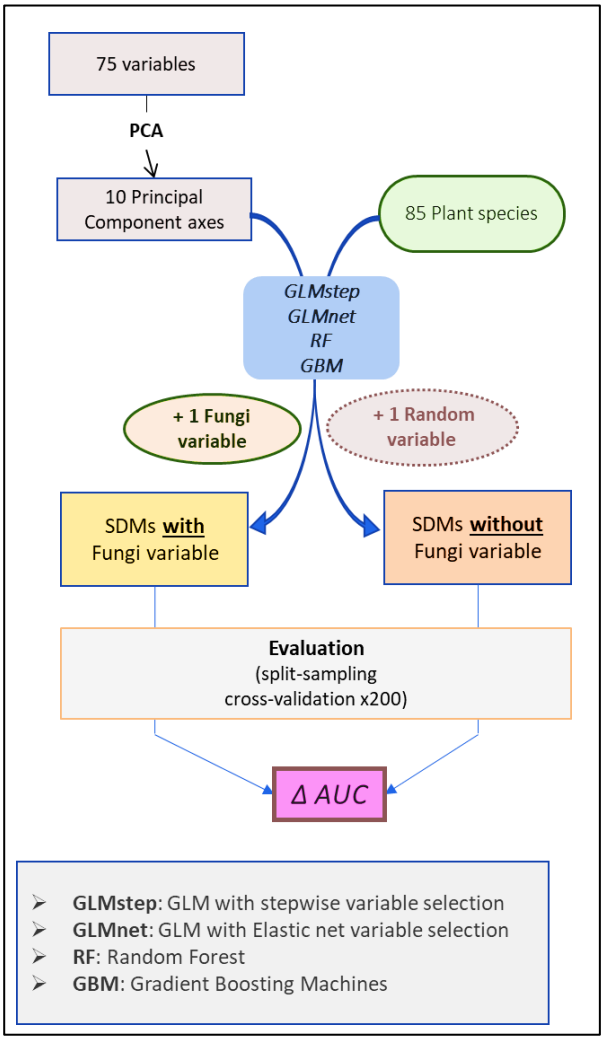


Figure 3. **Summary of the modelling steps, from preparing the predictors and response variable, to evaluating the performance of the models.**

c) *Building SDMs*

Four approaches were used to fit the models. Generalized linear models (GLM) use maximum-likelihood approach to determine the response variable, which is more powerful than a simple least-square linear regression. We used two variable selection procedures. The first one was conducted with a classical stepwise regression, which adds and removes variables from the model on·10-by-one, and always selects the option that offers the highest improvement in performance. The performance at each step was assessed with the Akaike's Information criteria (AIC). A limit of one variable per 10 presence data points was set for each plant species, so only the first variables selected were kept. The second approach used a GLM with the elastic net variable selection function from the *glmnet* package. This algorithm combines ridge regression and lasso to select coefficients for the variables such as to minimize the residual error of the predicted values, while also adding a penalty (lambda) to this error. We used elastic net with k-fold cross-validation and equal weighting to ridge regression and lasso (α=0.5).

Random forest (RF) and Gradient boosting machines (GBM) are decision tree-based algorithms. The main difference between them is that GBM builds its trees sequentially, by accounting for the error rate of the previous ones, instead of randomly for Random Forest. RF models were built with 1000 trees with the randomForest package, and the GBM models were built with the gbm package, using a slow learning rate (shrinkage = 0.01). No prior variable selection procedure is required for those two methods, as they proceed by recursive partitioning and thus already select the given variables at each node.

d) Addition of fungal predictors

For the base models, only the abiotic variables that were reduced to 10 principal component axes were used, along with an additional standardized random variable (mean=0, SD=1) since model building and variable selection outcome can be affected by the initial number of variables. One of the fungi-related variables was then added to the models, replacing the random variable. This way, the base models had the same initial number of variables as those with the fungi variables.

The GLMs variable selection step included the fungi variables as options for the model to choose from. We assume that if the variable is not kept after variable selection, then its effect must be null and the model with the fungi-related variable will not be better than the model without it.

e) Evaluation of the SDMs

Each model (one per combination of model type, plant species and fungi variable) was evaluated with 200 80-to-20 split-sampling cross-validations (80% of the data for calibration, and the 20% left for evaluation). The variable selection processes for the GLMs (AIC stepwise selection and elastic net) were included in the evaluation, as recommended by Hornung et al. (2014). Performances of models with and without fungi richness values are compared with three performance assessments: the area under the ROC curve (AUC), maximum True Skill Statistics (maxTSS) and maximum kappa (maxKappa). To see if adding the fungal predictor improved the model performance, the difference of performance (dAUC, for the AUC metric) between the models with and without fungi diversity was computed (Figure 3).

RESULTS

Performance of the base models

Across the four model types with only the abiotic variables as predictors, RF performed the best, followed by GLM (first with elastic net, then with stepwise variable selection) and GBM. The GLM with stepwise variable selection performs slightly better on average than the one with elastic net when considering the maxTSS and maxKappa values. Otherwise, the differences

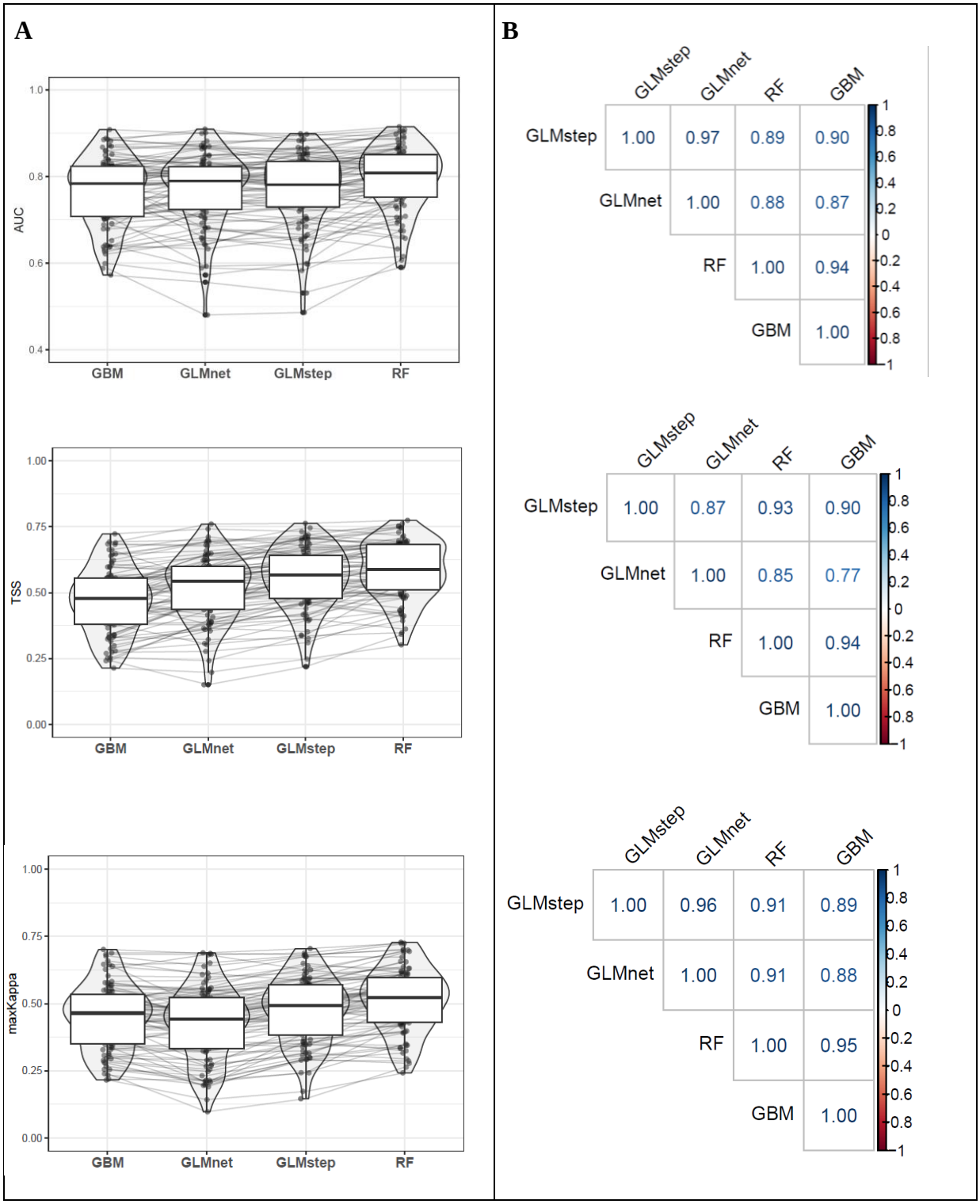


Figure 4. **Model performances of the base models.** (A) *Boxplots of performance values (AUC, maxTSS and maxKappa) for all plant species by models. Each point represents the mean performance value obtained for one plant species model. Points that represent the same plant species are connected horizontally. Horizontal black lines of the boxplots indicate the median values, and the box delimits the 25th and 75th percentile of the distribution. Whiskers' range is at 1.5 times the IQR (Interquartile Range) from the box. Violins plots in the background show the distribution density of the performance values. (B) Correlation (Spearman's ρ) between the*

model results. Models are: GBM=Gradient Boosting Machines; GLMnet=Generalized linear model with elastic net variable selection; GLMstep=Generalized linear model with stepwise variable selection; RF=Random Forest.

between the models' performances follow the same trend for AUC, maxTSS and maxKappa (Figure 4A). Some species' GLM models performed poorly, (*e.g. Briza media*, *Campanula rotundifolia*, *Festuca rubra, Helianthemum nummularium s.l.* and *Thymus pulegioides)* with AUC values is inferior to 0.6 (AUC = 0.5 being a model with random predictions). All those species have a low number of occurrences across the >200 sites, ranging from 33 to 42, except for Festuca rubra which has 86 (Figure S1). The performance values across models are highly correlated (Figure 4B), so no model type can be said to provide a different performance trend from the others. We thus assume the results will be similar regardless of what model is used. Since Random Forest models are those with the highest performance in average, the results from the other models will not be displayed in the following part but will be present as supplementary material.

Addition of total fungal richness

The addition of fungal total ASV richness only poorly increases the AUC of the Random Forest models, with an average AUC increase of 0.00614 ($V = 714$, $P = 1.45 \cdot 10^{-5}$) across all plant species. The other modelling techniques do not show any significant increase of the median AUC (Tables S4, S5, S6).

Addition of fungal richness of trophic groups, ecological guilds, and relative abundance of trophic groups

The steps done for total fungal ASV richness are applied to richness values of the three major trophic groups of fungi (saprotrophs, symbiotrophs and pathotrophs), with similar results (Figure S4A). The average change in AUC stays positive, though still low (Figure 5).

Out of all 23 fungal guild richness variables added to the RF models, 8 of them (4 from saprotrophs and 4 symbiotrophs) resulted in a higher mean AUC increase than total fungal richness (Figure 6). However, only soil saprotroph and animal endosymbiont richness had a higher median value (Table S2), indicating that what causes the higher improvement means are usually a few outliers with higher ΔAUC values. When also considering the changes in maxTSS and maxKappa, the guilds that improved the random forest models the most were animal endosymbionts, lichenized fungi, and soil saprotrophs (Table S2).

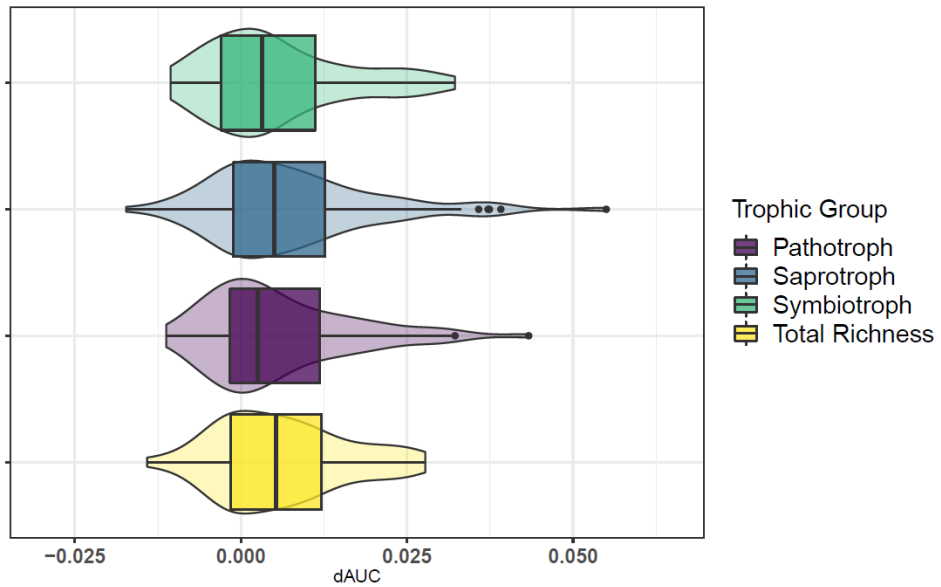


Figure 5. **Change in AUC of Random Forest models with the addition of total fungal richness and richness from the main trophic groups.** *Positive dAUC indicates that the models with fungal richness perform better than those without. Bold black lines of the boxplots indicate the median values, and the box delimits the 25th and 75th percentile of the distribution. Whiskers' range is at 1.5 times the IQR (Interquartile Range) from the box. Violins plots in the background show the distribution density of the performance values. The other performance measures (maxTSS & maxKappa) are found in Figure S3.*

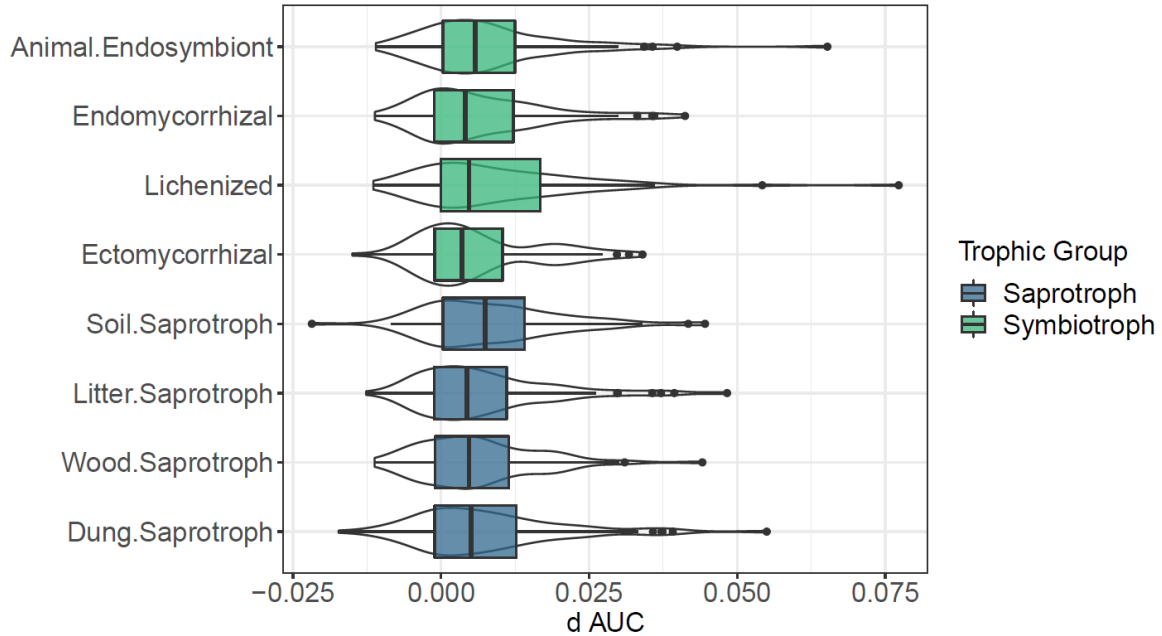


Figure 6**. Change in AUC according to the fungi guild richness added to the Random Forest models**. *Positive dAUC indicates that the models with fungal richness perform better than those without. Only the guilds with a higher mean increase than with total ASV richness are displayed. Each boxplot contains the average changes in AUC (d AUC) for each species and is coloured by trophic group. Bold black lines of the boxplots indicate the median values, and the box delimits the 25th and 75th percentile of the distribution. Whiskers' range is at 1.5 times the IQR (Interquartile Range) from the box. Violins plots in the background show the distribution density of the performance values.*

Adding relative abundance of the major trophic groups to the Random Forest models also increases the performance on average, but the mean and median increases are almost constantly lower than for the models with the addition of total fungal richness (Table S3).

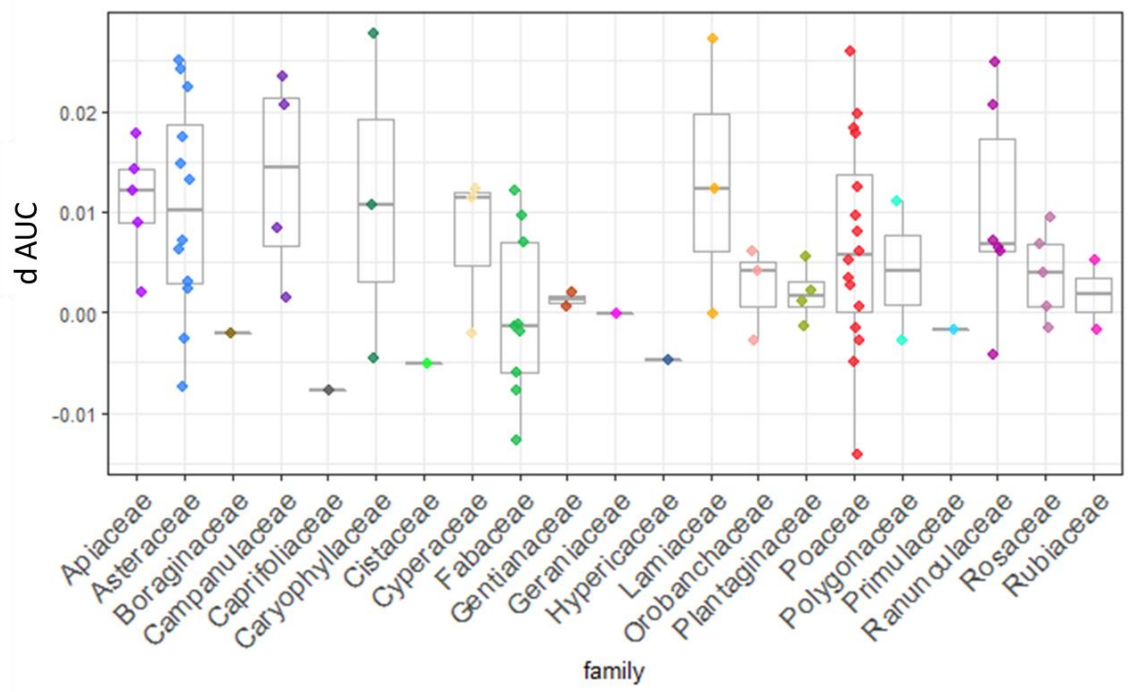


Figure 7. C**hange in AUC of random Forest models with the addition of total fungal richness, sorted by plant families**. *Each point represents the mean AUC change for one plant species, and is coloured by family. Bold black lines of the boxplots indicate the median values, and the box delimits the 25th and 75th percentile of the distribution. Whiskers' range is at 1.5 times the IQR (Interquartile Range) from the box.*

Species-specific response to the addition of fungal richness to SDMs

We can visually assess different responses of plant models to fungal richness based on the plant species' family (Figure 7), but no relevant analysis of variance could be performed to determine

if any family had a significantly higher change in AUC than the others. This was due to the low sample size, the fact that several families only had one plant species, and the uneven distribution of the variance.

DISCUSSION

The modelling techniques respond differently to the addition of fungal predictors

The four modelling techniques responded differently to the addition of the fungal predictors, with the tree-based techniques (RF and GBM) showing at least sometimes a significant increase in performance, while the regression-based ones showed none (Tables S4, S5). The variable selection procedures for the GLMs are probably the cause. Indeed, even if the fungal predictor could have a positive effect on the quality of the prediction, it might rarely be selected during the selection procedures, resulting in models that do not use the fungal predictor at all for its predictions.

Fungal richness improves the Random Forest models, but only on a small scale

Fungal total ASV richness did not greatly improve the Random Forest predictions, which was somewhat expected. Fungal richness is usually not specific to one or more plant species, but linked to a set of other factors like plant biomass (Cline et al., 2018) or soil pH. There is still a clear positive effect of integrating total ASV richness to the Random Forest models, as was the case for almost all ASV richness of fungal functional groups (Table S4). Fungal communities are driven by edaphic factors like pH and soil moisture (Kaisermann et al., 2015), but can also be linked to wider climatic variables like mean annual temperature (Shen et al., 2020). Some variance of the environmental predictors is lost with the dimensionality reduction of the PCA, and this lost variance might be compensated by adding fungal richness, which correlates to some predictors that are also important drivers of plant species distribution, like pH (Buri et al., 2017) and soil moisture-related factors (Fu et al., 2022), like NDMI and precipitation (Table S3).

Using the three main trophic groups' richness as predictors yield equivalent results, probably because they are highly correlated with total richness, as well as with each other. Only saprotroph richness improved the models better than total richness on average, suggesting that saprotroph richness might be linked to some aspect of soil that drives plant distribution. Saprotrophic fungi are a main component of organic matter decomposition processes, and their

diversity and community structure can be linked to some plant species presence and plant functional groups (Francioli et al., 2021; Semchenko et al., 2018).

Some symbiotic fungi richness also improves the predictions, in particular animal endosymbionts, which shows higher performance values than adding total richness in every metric (Table S3). This could be linked to the co-occurrence of some plant species with certain animal species, like arthropods that carry those endosymbiotic fungi. But it could also be linked to a completely different ecological process that happens to drive both animal endosymbiotic fungi richness and plant species distribution. Confirming either of those hypotheses would require a physiological study between the plants and fungi that goes beyond the analysis of co-occurrence patterns done with species distribution modelling.

Limitations

This study presents some limitations regarding the modelling process and the links made between plants and fungi data. First, we do not account for temporal disconnect between the vegetation survey and soil sampling for the fungi data, which were not always sampled on the same years. Secondly, we cannot exclude that some variation in fungi richness and abundances of can be influenced by seasonal variations (Kaisermann et al., 2015) that might not be the same across all sites, such as local droughts (Francioli et al., 2021). Also, the data FUNGuild bases itself on to attribute functional groups to fungi sequences is still somewhat incomplete, and might be biased towards some taxonomical groups while disregarding others (George et al., 2019). In our case, about half of the fungi sequence variants were not identified or lacked information about their ecological function. This kind of problem has already been mentioned in other studies (Cevallos et al., 2022; Monteiro et al., 2023), and highlights the importance of having more complete open access data on fungi sequences.

There is also an issue regarding the usage of fungal richness as a measure that reflects ecological factors. Since richness is only the sum of ASVs present after amplifying the wanted sequence, this measure is extremely sensitive to the sequencing depth, *i.e.* the sampling effort (Kleine Bardenhorst et al., 2022). With higher sequencing depth comes a higher probability of capturing higher richness due a better representation of the rarer sequences. Fungal richness might thus not be representative of the actual ecological properties of each site's soil. Further research on this subject should rather focus on diversity measures that accounts for the evenness of the microbial community, like the adjusted Shannon-Wiener index.

Conclusion

As we demonstrated here, the improvement of plant species distribution models with fungi-related variables depends on the modelling method used, and is most often low. The increase in performance that we observe here could be due to the dimensionality reduction of the abiotic predictors, which results in the loss of variance of some important ones. We hypothesized that fungi richness or relative abundance could be used as a proxy for some soil-related process that would improve plant SDMs, but our results were not conclusive in this regard. It would thus be relevant to test whether other diversity measures of fungi could have a greater effect on the prediction of plant distribution.

Fungi diversity improving plant SDMs because of a biotic interaction with plants is not shown here. This might be because biotic interactions are difficult to disentangle from variation in the abiotic predictors (D'Amen et al., 2018). For example, although competition between two species clearly drives their distribution, the effect of the competition can also be explained by the abiotic environment (Godsoe et al., 2017). This means that, in theory, including the biotic interaction in the modelling will not considerably improve the predictions, since the realized niche of the species is already captured by the abiotic predictors.

One missing aspect from this study is the effect of fungal richness on plant SDMs depending on the elevational range of their distribution. We know that plant-plant interactions vary across elevation (Chamberlain et al., 2014), and fungal diversity is also structured differently across an elevational gradient (Pellissier et al., 2014). Plant-fungi associations might be more crucial to the survival of the plant at higher elevation, where the environmental conditions are harsher, as was already suggested by Pellissier et al. (2013). Fungi diversity could thus be more important for shaping some specific plant species' distribution.

ACKNOWLEDGEMENTS

I would like to thank Valentin Verdon for supervising this project, providing me with both the needed data, advice on how to approach the subject, and for the revision of my draft. I thank Antoine Guisan for his advice and correction of my draft as well.

REFERENCES

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Austin, M. P., & Van Niel, K. P. (2011). Impact of landscape predictors on climate change modelling of species distributions: A case study with Eucalyptus fastigata in southern New South Wales, Australia: Impact of landscape predictors on climate change modelling. *Journal of Biogeography*, *38*(1), 9–19. https://doi.org/10.1111/j.1365-2699.2010.02415.x

Breiner, F. T., Anand, M., Butchart, S. H. M., Flörke, M., Fluet-Chouinard, E., Guisan, A., Hilarides, L., Jones, V. R., Kalyakin, M., Lehner, B., Leeuwen, M., Pearce-Higgins, J. W., Voltzit, O., & Nagy, S. (2022). Setting priorities for climate change adaptation of Critical Sites in the Africa-Eurasian waterbird flyways. *Global Change Biology*, *28*(3), 739–752. https://doi.org/10.1111/gcb.15961

Broennimann, O. (2018). CHclim25: A high spatial and temporal resolution climate dataset for Switzerland. Technical report. Ecospat laboratory, University of Lausanne, Switzerland. Available at www.unil.ch/ecospat/files/live/sites/ecospat/files/shared/PDF_site/chclim25.pdf).

Buri, A., Cianfrani, C., Pinto-Figueroa, E., Yashiro, E., Spangenberg, J. E., Adatte, T., Verrecchia, E., Guisan, A., & Pradervand, J.-N. (2017). Soil factors improve predictions of plant species distribution in a mountain environment. *Progress in Physical Geography: Earth and Environment*, *41*(6), 703–722. https://doi.org/10.1177/0309133317738162

Buri, A., Grand, S., Yashiro, E., Adatte, T., Spangenberg, J. E., Pinto-Figueroa, E., Verrecchia, E., & Guisan, A. (2020). What are the most crucial soil variables for predicting the distribution of mountain plant species? A comprehensive study in the

Swiss Alps. *Journal of Biogeography*, *47*(5), 1143–1153. https://doi.org/10.1111/jbi.13803

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Callaway, R. M., Brooker, R. W., Choler, P., Kikvidze, Z., Lortie, C. J., Michalet, R., Paolini, L., Pugnaire, F. I., Newingham, B., Aschehoug, E. T., Armas, C., Kikodze, D., & Cook, B. J. (2002). Positive interactions among alpine plants increase with stress. *Nature*, *417*(6891), 844–848. https://doi.org/10.1038/nature00812

Cevallos, S., Herrera, P., Vélez, J., & Suárez, J. P. (2022). Root-Associated Endophytic and Mycorrhizal Fungi from the Epiphytic Orchid Maxillaria acuminata in a Tropical Montane Forest in Southern Ecuador. *Diversity*, *14*(6), 478. https://doi.org/10.3390/d14060478

Chamberlain, S. A., Bronstein, J. L., & Rudgers, J. A. (2014). How context dependent are species interactions? *Ecology Letters*, *17*(7), 881–890. https://doi.org/10.1111/ele.12279

Cline, L. C., Hobbie, S. E., Madritch, M. D., Buyarski, C. R., Tilman, D., & Cavender-Bares, J. M. (2018). Resource availability underlies the plant-fungal diversity relationship in a grassland ecosystem. *Ecology*, *99*(1), 204–216. https://doi.org/10.1002/ecy.2075

D'Amen, M., Mod, H. K., Gotelli, N. J., & Guisan, A. (2018). Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. *Ecography*, *41*(8), 1233–1244. https://doi.org/10.1111/ecog.03148

Descombes, P., Pitteloud, C., Glauser, G., Defossez, E., Kergunteuil, A., Allard, P.-M., Rasmann, S., & Pellissier, L. (2020). Novel trophic interactions under climate change promote alpine plant coexistence. *Science*, *370*(6523), 1469–1473. https://doi.org/10.1126/science.abd7015

Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P., & Guisan, A. (2013). Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science*, *24*(4), 593–606. https://doi.org/10.1111/jvs.12002

Edgar, R. C. (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*, *34*(14), 2371–2375. https://doi.org/10.1093/bioinformatics/bty113

Frąc, M., Hannula, S. E., Bełka, M., & Jędryczka, M. (2018). Fungal Biodiversity and Their Role in Soil Health. *Frontiers in Microbiology*, *9*, 707. https://doi.org/10.3389/fmicb.2018.00707

Francioli, D., van Rijssel, S. Q., van Ruijven, J., Termorshuizen, A. J., Cotton, T. E. A., Dumbrell, A. J., Raaijmakers, J. M., Weigelt, A., & Mommer, L. (2021). Plant functional group drives the community structure of saprophytic fungi in a grassland biodiversity experiment. *Plant and Soil*, *461*(1–2), 91–105. https://doi.org/10.1007/s11104-020-04454-y

Fu, W., Chen, B., Rillig, M. C., Jansa, J., Ma, W., Xu, C., Luo, W., Wu, H., Hao, Z., Wu, H., Zhao, A., Yu, Q., & Han, X. (2022). Community response of arbuscular mycorrhizal fungi to extreme drought in a cold-temperate grassland. *New Phytologist*, *234*(6), 2003–2017. https://doi.org/10.1111/nph.17692

George, P. B. L., Creer, S., Griffiths, R. I., Emmett, B. A., Robinson, D. A., & Jones, D. L. (2019). Primer and Database Choice Affect Fungal Functional but Not Biological Diversity Findings in a National Soil Survey. *Frontiers in Environmental Science*, *7*, 173. https://doi.org/10.3389/fenvs.2019.00173

Giaccone, E., Luoto, M., Vittoz, P., Guisan, A., Mariéthoz, G., & Lambiel, C. (2019). Influence of microclimate and geomorphological factors on alpine vegetation in the Western Swiss Alps. *Earth Surface Processes and Landforms*, *44*(15), 3093–3107. https://doi.org/10.1002/esp.4715

Godsoe, W., Franklin, J., & Blanchet, F. G. (2017). Effects of biotic interactions on modeled species' distribution can be masked by environmental gradients. *Ecology and Evolution*, *7*(2), 654–664. https://doi.org/10.1002/ece3.2657

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat Suitability and Distribution Models: With Applications in R* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781139028271

Hirzel, A., & Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, *157*(2–3), 331–341. https://doi.org/10.1016/S0304-3800(02)00203-X

Kaisermann, A., Maron, P. A., Beaumelle, L., & Lata, J. C. (2015). Fungal communities are more sensitive indicators to non-extreme soil moisture variations than bacterial communities. *Applied Soil Ecology*, *86*, 158–164. https://doi.org/10.1016/j.apsoil.2014.10.009

Kleine Bardenhorst, S., Vital, M., Karch, A., & Rübsamen, N. (2022). Richness estimation in microbiome data obtained from denoising pipelines. *Computational and Structural Biotechnology Journal*, *20*, 508–520. https://doi.org/10.1016/j.csbj.2021.12.036

Mod, H. K., Scherrer, D., Di Cola, V., Broennimann, O., Blandenier, Q., Breiner, F. T., Buri, A., Goudet, J., Guex, N., Lara, E., Mitchell, E. A. D., Niculita-Hirzel, H., Pagni, M., Pellissier, L., Pinto-Figueroa, E., Sanders, I. R., Schmidt, B. R., Seppey, C. V. W., Singer, D., … Guisan, A. (2020). Greater topoclimatic control of above- versus below-ground communities. *Global Change Biology*, *26*(12), 6715–6728. https://doi.org/10.1111/gcb.15330

Monteiro, J. S., Almeida, M. S., Medeiros-Sarmento, P. S., Caldeira, C. F., Ramos, S. J., Oliveira, G., Valadares, R. B. S., & Gastauer, M. (2023). DNA metabarcoding reveals compositional and functional differences in fungal communities among Amazonian

canga formations. *Fungal Ecology*, *61*, 101209. https://doi.org/10.1016/j.funeco.2022.101209

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S., & Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006

Oke, O. A., & Thompson, K. A. (2015). Distribution models for mountain plant species: The value of elevation. *Ecological Modelling*, *301*, 72–77. https://doi.org/10.1016/j.ecolmodel.2015.01.019

Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., Lesur, I., Vallance, J., & Vacher, C. (2019). Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecology*, *41*, 23–33. https://doi.org/10.1016/j.funeco.2019.03.005

Pellissier, L., Niculita-Hirzel, H., Dubuis, A., Pagni, M., Guex, N., Ndiribe, C., Salamin, N., Xenarios, I., Goudet, J., Sanders, I. R., & Guisan, A. (2014). Soil fungal communities of grasslands are environmentally structured at a regional scale in the Alps. *Molecular Ecology*, *23*(17), 4274–4290. https://doi.org/10.1111/mec.12854

Pellissier, L., Pinto-Figueroa, E., Niculita-Hirzel, H., Moora, M., Villard, L., Goudet, J., Guex, N., Pagni, M., Xenarios, I., Sanders, I., & Guisan, A. (2013). Plant species distributions along environmental gradients: Do belowground interactions with fungi matter? *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00500

Pinto-Figueroa, E. A., Seddon, E., Yashiro, E., Buri, A., Niculita-Hirzel, H., van der Meer, J. R., & Guisan, A. (2019). Archaeorhizomycetes Spatial Distribution in Soils Along Wide Elevational and Environmental Gradients Reveal Co-abundance Patterns With Other Fungal Saprobes and Potential Weathering Capacities. *Frontiers in Microbiology*, *10*, 656. https://doi.org/10.3389/fmicb.2019.00656

Pradervand, J.-N., Dubuis, A., Pellissier, L., Guisan, A., & Randin, C. (2014). Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Progress in Physical Geography: Earth and Environment*, *38*(1), 79–96. https://doi.org/10.1177/0309133313512667

Püschel, D., Bitterlich, M., Rydlová, J., & Jansa, J. (2020). Facilitation of plant water uptake by an arbuscular mycorrhizal fungus: A Gordian knot of roots and hyphae. *Mycorrhiza*, *30*(2–3), 299–313. https://doi.org/10.1007/s00572-020-00949-9

Randin, C. F., Engler, R., Normand, S., Zappa, M., Zimmermann, N. E., Pearman, P. B., Vittoz, P., Thuiller, W., & Guisan, A. (2009). Climate change and plant distribution: Local models predict high-elevation persistence. *Global Change Biology*, *15*(6), 1557–1569. https://doi.org/10.1111/j.1365-2486.2008.01766.x

Scherrer, D., Massy, S., Meier, S., Vittoz, P., & Guisan, A. (2017). Assessing and predicting shifts in mountain forest composition across 25 years of climate change. *Diversity and Distributions*, *23*(5), 517–528. https://doi.org/10.1111/ddi.12548

Semchenko, M., Leff, J. W., Lozano, Y. M., Saar, S., Davison, J., Wilkinson, A., Jackson, B. G., Pritchard, W. J., De Long, J. R., Oakley, S., Mason, K. E., Ostle, N. J., Baggs, E. M., Johnson, D., Fierer, N., & Bardgett, R. D. (2018). Fungal diversity regulates plant-soil feedbacks in temperate grassland. *Science Advances*, *4*(11), eaau4578. https://doi.org/10.1126/sciadv.aau4578

Shen, C., Gunina, A., Luo, Y., Wang, J., He, J., Kuzyakov, Y., Hemp, A., Classen, A. T., & Ge, Y. (2020). Contrasting patterns and drivers of soil bacterial and fungal diversity across a mountain gradient. *Environmental Microbiology*, *22*(8), 3287–3301. https://doi.org/10.1111/1462-2920.15090

Syfert, M. M., Joppa, L., Smith, M. J., Coomes, D. A., Bachman, S. P., & Brummitt, N. A. (2014). Using species distribution models to inform IUCN Red List assessments. *Biological Conservation*, *177*, 174–184. https://doi.org/10.1016/j.biocon.2014.06.012

Tedersoo, L., Bahram, M., & Zobel, M. (2020). How mycorrhizal associations drive plant population and community biology. *Science*, *367*(6480), eaba1223. https://doi.org/10.1126/science.aba1223

von Däniken, I., Guisan, A., & Lane, S. (2014). *RechAlp.vd: Une nouvelle plateforme UNIL de support pour la recherche transdisciplinaire dans les Alpes vaudoises* [Text/html,application/pdf,text/html]. https://doi.org/10.5169/SEALS-513645

Yashiro, E., Pinto-Figueroa, E., Buri, A., Spangenberg, J. E., Adatte, T., Niculita-Hirzel, H., Guisan, A., & van der Meer, J. R. (2016). Local Environmental Factors Drive Divergent Grassland Soil Bacterial Communities in the Western Swiss Alps. *Applied and Environmental Microbiology*, *82*(21), 6303–6316. https://doi.org/10.1128/AEM.01170-16

Zanne, A. E., Abarenkov, K., Afkhami, M. E., Aguilar-Trigueros, C. A., Bates, S., Bhatnagar, J. M., Busby, P. E., Christian, N., Cornwell, W. K., Crowther, T. W., Flores-Moreno, H., Floudas, D., Gazis, R., Hibbett, D., Kennedy, P., Lindner, D. L., Maynard, D. S., Milo, A. M., Nilsson, R. H., … Treseder, K. K. (2020). Fungal functional ecology: Bringing a trait-based approach to plant-associated fungi. *Biological Reviews*, *95*(2), 409–433. https://doi.org/10.1111/brv.12570

Zimmermann, N. E., Edwards, T. C., Graham, C. H., Pearman, P. B., & Svenning, J.-C. (2010). New trends in species distribution modelling. *Ecography*, *33*(6), 985–989. https://doi.org/10.1111/j.1600-0587.2010.06953.x

SUPPLEMENTARY MATERIAL

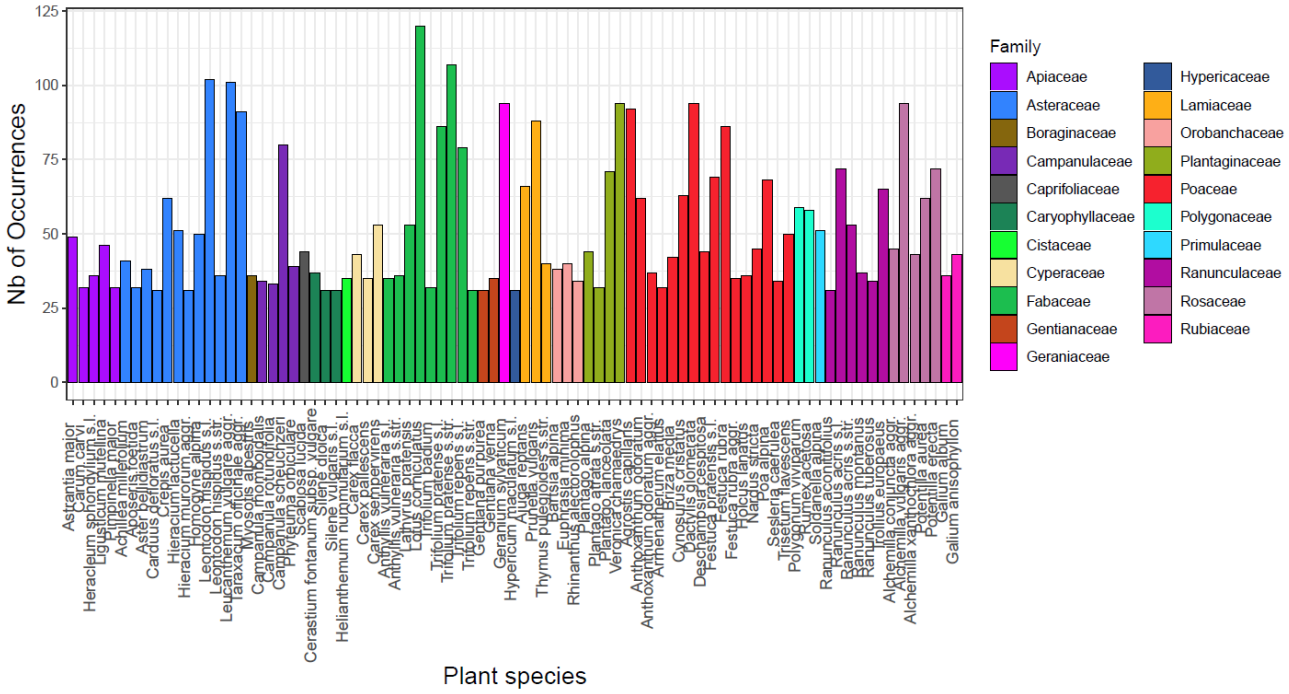


Figure S1. **Barplot of occurrences of plant species across all sites.** *Colour-coded by taxonomical family.*

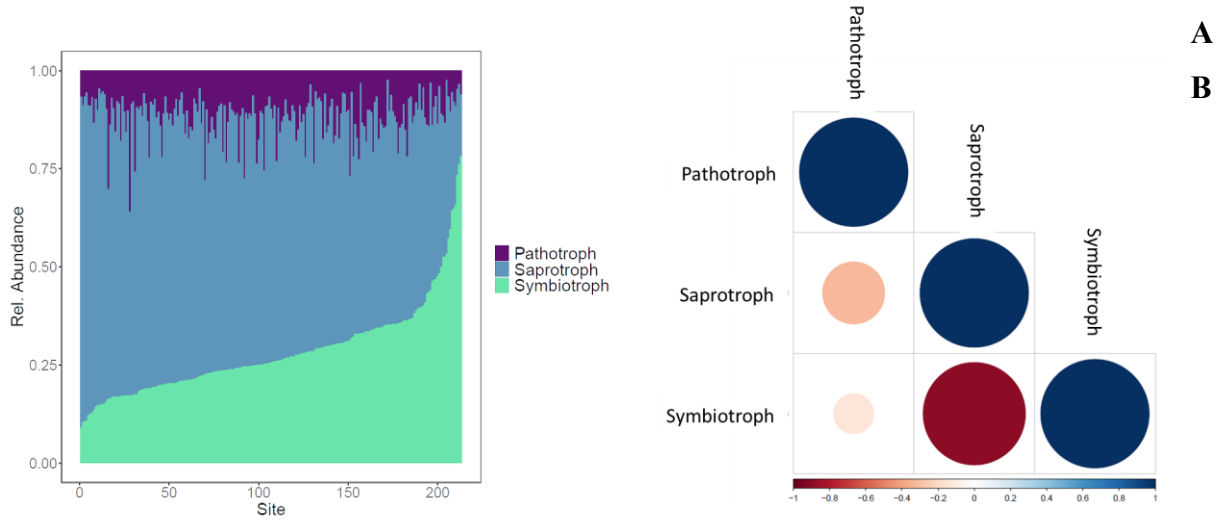


Figure S2**. Relative abundance of the 3 major fungal trophic groups in the sampling sites' soil.** *(A) Relative abundances of fungi from the 3 major trophic groups, as defined by FUNGuild.*

Sites (the x-axis) are ordered by increasing relative abundance of symbiotrophs. (B) Correlation of the relative abundances, with colour indicating the direction (blue = positive, red=negative), and size the intensity of the correlation.

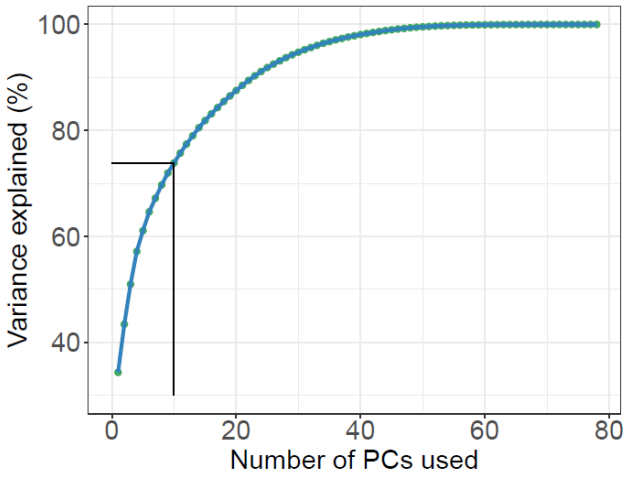


Figure S3**. Variance explained of all environmental factors, according to the number of principal component (PC) axes used**. *Black line indicates the chosen number of PCs kept for the analysis (here 10).*

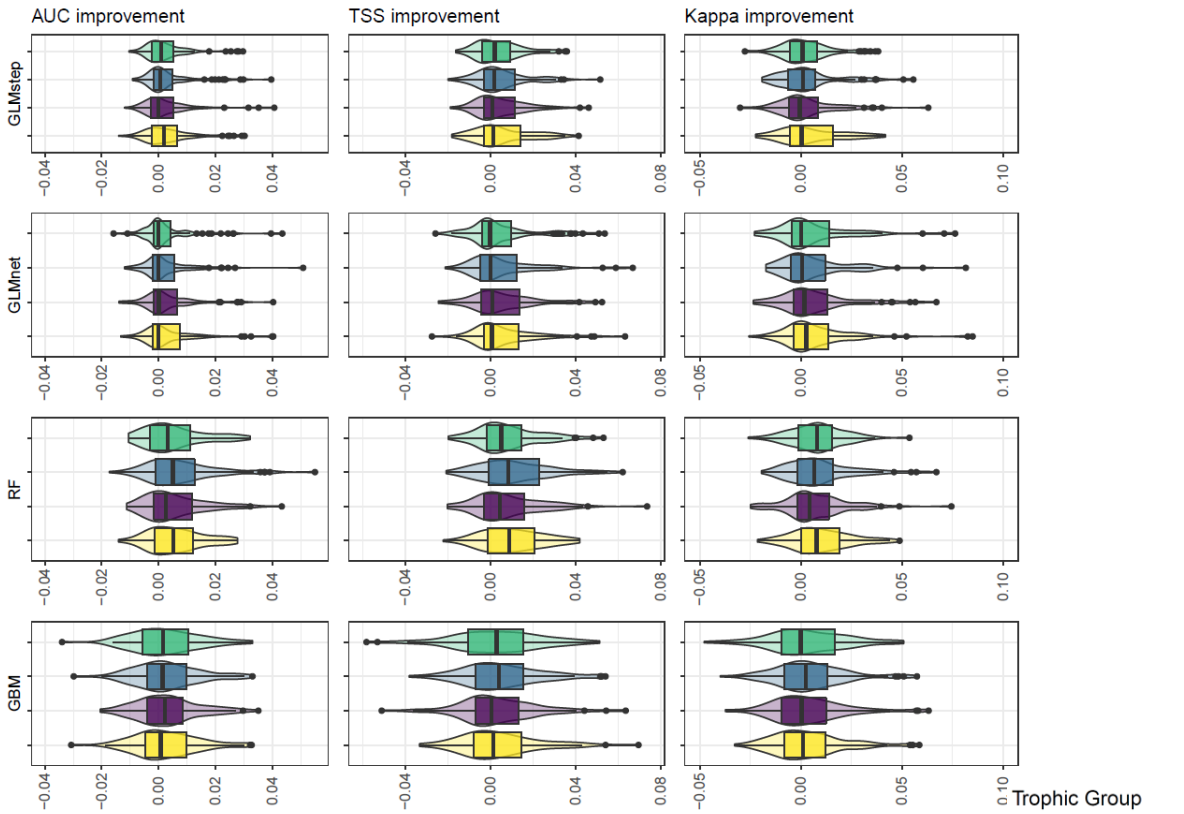
Environmental variables	
bio1_t	Annual Mean Temperature (°C)
bio10_twarmq	Mean Temperature of Warmest Quarter (°C)
bio11_tcoldq	Mean Temperature of Coldest Quarter (°C)
bio12_p	Annual Precipitation (mm)
bio13_pwet	Precipitation of wettest month (mm)
bio14_pdry	Precipitation of driest month (mm)
bio15_ps	Precipitation seasonality
bio16_pwetq	Precipitation of wet est quarter (mm)
bio17_pdryq	Precipitation of driest quarter (mm)
bio18_pwarmq	Precipitation of warmest quarter (mm)
bio19_pcoldq	Precipitation of colder quarter (mm)
bio2_tdr	Mean Diurnal Range (°C)
bio3_tiso	Isothermality
bio4_ts	Temperature seasonality (°C)
bio5_tmaxw	Max temperature of warmest month (°C)
bio6_tminc	Min Temperature of Coldest Month (°C)
bio7_tar	Temperature Annual Range (°C)
bio8_twetq	Mean Temperature of Wettest Quarter (°C)
bio9_tdryq	Mean Temperature of Driest Quarter (°C)

GDD0	Sum of growing degree days above 0°C (°C)
ETP	Annual potential evapotranspiration (mm)
cumday_nofrost	Cum. days of the year without frost (T°<0°C)
sRadY	Annual solar radiation (kJ * day-1)
aspect	Aspect
slope	slope (°)
MarlShale_MarlDeposits	Presence of marl deposits (0-1)
Marlyshale_limestonePhyllite_sandstone	Presence of limestone/phyllite/sandstone (0-1)
MassiveLimestones	Presence of massive limestone (0-1)
Silt_clay_loess_groundmoraine_surfacemoraine	Presence of loess (0-1)
forest_aggr	Landuse type: forest (0-1)
hydro_aggr	Landuse type: water bodies (0-1)
lowVeg_aggr	Landuse type: low vegetation (0-1)
anthropos_aggr	Landuse type: urban area (0-1)
noise	Noise
deciduous	Deciduous trees cover (%)
Altitude	Elevation (masl)
ndmi	Normalized Difference Moisutre Index
ndvi	Normalized Difference Vegetation Index
soilTemp	Soil Temperature (°C)
bulkSoilWaterContent	Bulk soil water content (wt %)
pH	Soil pH
EC_1_5	Electrical conductivity ($S \cdot m^{-1}$)
TotalP	Total phosphorus content (mg/g)
Nitrogen	Bulk nitrogen content (wt %)
Carbon	Bulk Carbon content (wt %)
Hydrogen	Bulk Hydrogen content (wt %)
Phyllosilicates	Phyllosilicates (%)
Quartz	Quartz (%)
Feldspath_K	Feldspath-K (%)
Plagioclase_Na	Plagioclas·10-Na (%)
Calcite	Calcite (%)
Dolomite	Dolomite (%)
Goethite	Goethite (%)
Ankerite	Ankerite (%)
Indoses	Indoses (%)
SiO2	Silicon oxide (wt %)
TiO2	Titanium oxide (wt %)
Al2O3	Aluminium oxide (wt %)
Fe2O3	Iron oxide (wt %)
MnO	Manganese oxide (wt %)
MgO	Magnesium oxide (wt %)
CaO	Calcium oxide (wt %)
Na2O	Sodium oxide (wt %)
K2O	Potassium oxide (wt %)
P2O5	Phosphorus pentoxide (wt %)
OM	Organic matter content (%)
Cr2O3	Chromium oxide (wt %)

NiO	Nickel oxide (wt %)
d15N	Nitrogen stable isotope ratio (per ml vs air-N2)
d13C	Carbon stable isotope (per ml vs VPDB)
clay	Clay
ThinSilt	Thin silt
ThickSilt	Thick silt
ThinSand	Thin sand
ThickSand	Thick sand

Table S1**. List of environmental variables used for modelling**. *Some abbreviations: wt % = percentage of weight; (0-1) = binary data.*

A



B

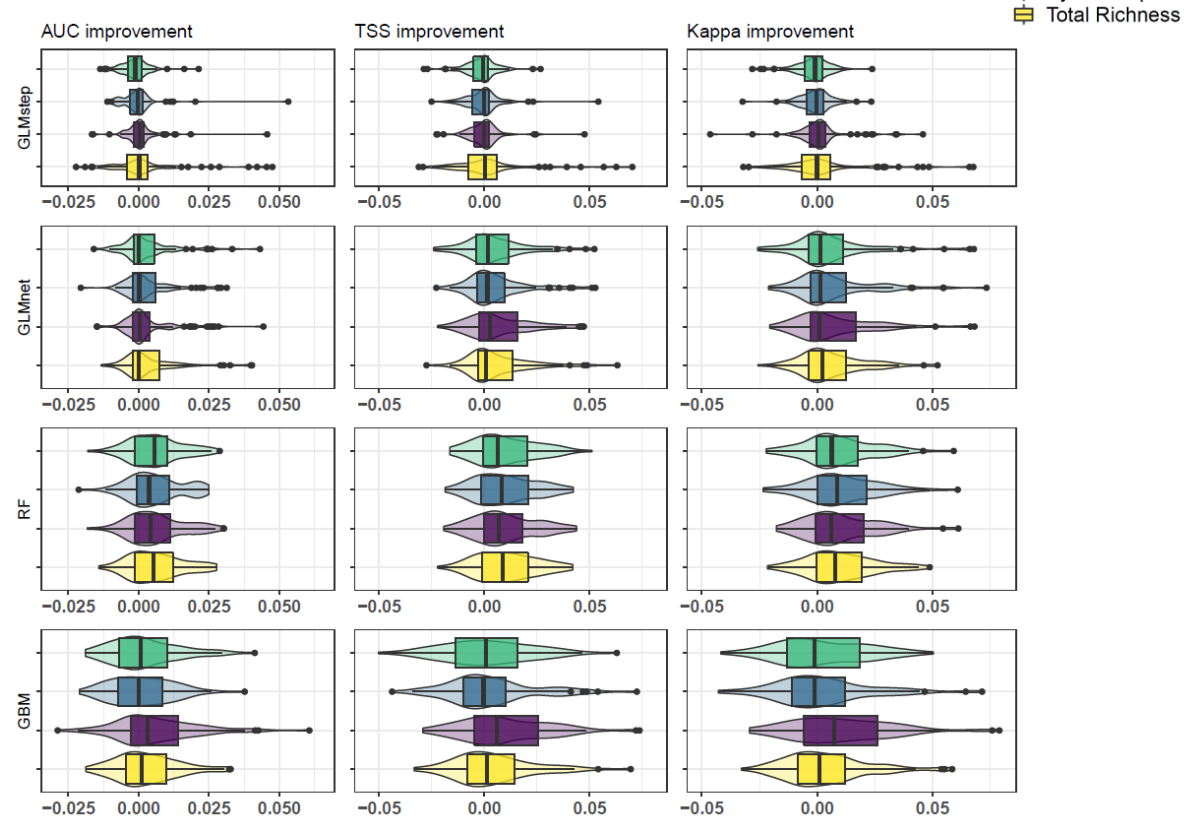


Figure S4. **Change in model performance with the addition of fungal richness from the main trophic groups.** *Boxplots are made from the mean change in SDM performance per species, when*

adding fungal richness, and are coloured by trophic group. The boxplots are categorized by modelling technique and by performance metric. Horizontal black lines of the boxplots indicate the median values, and the box delimits the 25th and 75th percentile of the distribution. Whiskers' range is at 1.5 times the IQR (Interquartile Range) from the box. Violins plots in the background show the distribution density of the performance values. (A) Change in model performance when adding fungal richness of total richness and trophic groups; (B) Change in model performance when adding fungal richness of total richness and relative abundances of trophic groups.

Variable		***|Pearson's R| correlation with total fungal richness***
pH	*pH*	0.5661
NDVI	*Normalized Difference Vegetation Index*	0.5468
CaO	*Calcium oxide*	0.4464
NDMI	*Normalized Difference Moisture Index*	0.4257
bio3_tiso	*Isothermality*	0.4185
bio13_pwet	*Precipitation of the wettest month*	0.3991
Na2O	*Sodium oxide*	0.3905
bio16_pwetq	*Precipitation of wettest quarter*	0.3879
bio18_pwarmq	*Precipitation of warmest quarter*	0.3878
SiO2	*Silicon dioxide*	0.3783

Table S2. **List of the 10 environmental variables that were the most correlated with total fungal richness**. *The correlation metric used is Pearson's R.*

RF	*Trophic group/guild*	*Median ΔAUC*	*Mean ΔAUC*	*V-statistics*	*Adj. p.-value*	*Median ΔTSS*	*Mean ΔTSS*	*Median ΔKappa*	*Mean ΔKappa*
	Total Richness	0.005230	0.006137	714	**1.45E-05**	0.008660	0.010256	0.007742	0.010001
Richness	**Pathotroph**	0.002535	0.005693	929	**0.00112398**	0.004500	0.008209	0.004220	0.007545
	Bryophyte Parasite	0.002512	0.003288	1240	0.13646501	0.002969	0.004656	0.002373	0.005009
	Fungal Parasite	0.003547	0.005715	764	**4.31E-05**	0.004028	0.007178	0.005821	0.006742
	Plant Parasite	0.002115	0.003434	1152	**0.04184375**	0.001970	0.005006	0.003158	0.004983
	Animal Pathogen	0.002676	0.005222	879	**0.00044123**	0.004390	0.007179	0.004210	0.006206
	Plant Pathogen	0.003888	0.005317	879	**0.00044123**	0.004697	0.006679	0.003222	0.005316
	Lichen Parasite	0.002153	0.004776	1015	**0.00504497**	0.002292	0.004790	0.003734	0.005631
	Saprotroph	0.005069	**0.007814**	683	**7.24E-06**	0.008353	**0.011125**	0.006607	0.009445
	Litter Saprotroph	0.004388	**0.007293**	644	**2.94E-06**	0.006667	0.009945	0.006991	0.009122
	Soil Saprotroph	**0.007454**	**0.008395**	512	**1.12E-07**	**0.010139**	**0.012143**	0.012166	**0.011351**
	Plant Saprotroph	0.00317	0.004565	979	**0.00273529**	0.005571	0.007334	0.004970	0.006062
	Dung Saprotroph	0.005069	**0.007814**	683	**7.24E-06**	0.008353	**0.011125**	0.006607	0.009445
	Leaf Saprotroph	0.002794	0.004714	966	**0.00218014**	0.005417	0.006266	0.003812	0.006471
	Wood Saprotroph	0.004735	**0.006255**	684.5	**7.50E-06**	0.005625	0.007726	0.004023	0.006698
	Symbiotroph	0.003137	0.005452	985	**0.00303405**	0.005000	0.008450	**0.007999**	0.008159
	Epiphyte	0.003788	0.005931	750	**3.20E-05**	0.006581	0.008719	0.006415	0.007894
	Lichenized	0.004714	**0.00924**	558	**3.63E-07**	**0.011389**	**0.014448**	0.012568	**0.015699**
	Arbuscular Mycorrhizal	0.002089	0.004165	989	**0.00324996**	0.005972	0.007113	**0.008182**	0.006967
	Endomycorrhizal	0.004051	**0.006682**	733	**2.21E-05**	**0.009000**	**0.011991**	0.004177	**0.010700**
	Ericoid Mycorrhizal	0.003009	0.005140	925	**0.00104471**	0.004899	0.007783	0.003694	0.007012
	Orchid Mycorrhizal	0.003143	0.004474	857	**0.00028819**	0.004259	0.008050	0.004928	0.007647
	Root-Associated Biotroph	0.001608	0.002732	1056	**0.00984459**	0.002583	0.004817	0.003405	0.005830
	Clavicipitaceous Endophyte	0.001400	0.002203	1364	0.57352218	0.001528	0.002672	0.001307	0.001754
	Ectomycorrhizal	0.003542	**0.006138**	785	**6.72E-05**	0.005375	0.009735	**0.008807**	**0.010704**
	Animal Endosymbiont	**0.005729**	**0.008147**	517	**1.28E-07**	**0.008833**	**0.011950**	**0.009187**	**0.010477**
	Endophyte	0.001400	0.002203	1364	0.57352218	0.001528	0.002672	0.001307	0.001754
Rel. Abundance	**Pathotroph**	0.00408	**0.00616**	714	**1.62E-06**	0.00680555	0.00972786	0.006268	0.00962275
	Saprotroph	0.0037	0.00579	744	**3.12E-06**	0.00819444	0.00971178	**0.00854753**	**0.01025072**
	Symbiotroph	0.0057	0.00586	712	**1.55E-06**	0.00633333	0.00969627	0.00632547	0.00951332

Table S3**. Average increase in AUC, maxTSS and maxKappa with the addition of fungal richness, or relative abundance, to the Random Forest models**. *P-values from the Wilcoxon test testing whether the median ΔAUC is superior to 0 are adjusted with a Bonferroni correction (p-value multiplied by 27), and are in bold when significant. Mean and median performance values superior to those of the random Forest models with total fungal richness are also in bold.*

GLMSTEP	*Trophic group/guild*	*Median ΔAUC*	*Mean ΔAUC*	*V-statistics*	*Adj. p.-value*	*Median ΔTSS*	*Mean ΔTSS*	*Median ΔKappa*	*Mean ΔKappa*
	Total Richness	0.000248	0.002743	1651.5	1	0.000278	0.00226	-6.80E-05	0.002658
Richness	**Pathotroph**	-0.000521	0.001711	1893	1	-0.000417	0.001259	0.000107	0.001554
	Bryophyte Parasite	-0.000898	-0.000522	2320.5	1	-0.001667	-0.001286	-0.000177	-0.00084
	Fungal Parasite	-0.000074	0.001432	1903	1	-0.000139	0.002193	0.000419	0.001511
	Plant Parasite	-0.000294	-0.00026	1982	1	-0.000278	-0.001322	-6.30E-05	0.00015
	Animal Pathogen	-0.000370	0.000791	1938.5	1	-0.000441	0.000203	-0.000708	-6.80E-05
	Plant Pathogen	0.000544	0.001577	1736	1	0.000417	0.001289	0.000555	0.000982
	Lichen Parasite	0.000030	0.000739	1821	1	0	9.50E-05	-0.000198	-0.000572
	Saprotroph	0.000041	0.002283	1676.5	1	0.000556	0.000552	0.000136	0.001251
	Litter Saprotroph	-0.000375	0.001003	1957	1	-0.000833	0.000671	-0.000860	0.000644
	Soil Saprotroph	-0.000202	0.002551	1789.5	1	0	0.002152	0.000107	0.002057
	Plant Saprotroph	0	0.001201	1833.5	1	-0.000417	0.000829	-0.000476	0.00128
	Dung Saprotroph	0.000041	0.002283	1676.5	1	0.000556	0.000552	0.000136	0.001251
	Leaf Saprotroph	-0.000857	0.000164	2181	1	-0.000571	-0.000627	6.40E-05	0.000149
	Wood Saprotroph	0.000221	0.000434	1855	1	0.000143	-0.00028	0.000750	0.000269
	Symbiotroph	-0.000023	0.001465	1717	1	-0.000441	0.001068	-0.000476	0.001193
	Epiphyte	0.000174	-0.000149	1901	1	-0.000139	-0.000336	3.40E-05	-5.00E-04
	Lichenized	0.000625	0.005303	1428	1	0.000625	0.005083	0.000669	0.005157
	Arbuscular Mycorrhizal	0.000208	0.002368	1569	1	0.000286	0.001967	-0.000176	0.002885
	Endomycorrhizal	0.000324	0.004569	1661.5	1	-0.000556	0.004746	-0.000232	0.004146
	Ericoid Mycorrhizal	0	0.000947	1841	1	0.000188	0.000843	-0.000415	0.00117
	Orchid Mycorrhizal	0.000609	0.002797	1629	1	0.000972	0.002042	0.001068	0.002264
	Root-Associated Biotroph	-0.000028	0.002316	1770	1	0.000412	0.001165	0.000774	0.002389
	Clavicipitaceous Endophyte	0.000055	-3.00E-04	1919	1	3.10E-05	-0.000579	0.000450	-0.000782
	Ectomycorrhizal	0.000208	0.000597	1960	1	-0.001	-0.000836	-0.001337	0.00018
	Animal Endosymbiont	0.000067	0.002074	1689	1	-0.000344	0.00203	0	0.00273
	Endophyte	0.000055	-3.00E-04	1919	1	3.10E-05	-0.000579	0.00045	-0.000782
Rel. Abundance	**Pathotroph**	0.000257	0.000796	1677	0.95	0	-0.000224	0.000639	0.001144
	Saprotroph	-0.000347	-0.000214	2163	1	0	-0.000436	-0.000422	-0.000608
	Symbiotroph	-0.001184	-0.001285	2458	1	-0.000694	-0.001749	-0.000923	-0.00207

Table S4**. Average increase in AUC, maxTSS and maxKappa with the addition of fungal richness, or relative abundance, to GLM models with stepwise variable selection**. *P-values from the Wilcoxon test testing whether the median ΔAUC is superior to 0 are adjusted with a Bonferroni correction (p-value multiplied by 27). The median and mean performance changes are written as 0 when the absolute value is inferior to* 10^{-5}.

GLMNET	*Trophic group/guild*	*Median ΔAUC*	*Mean ΔAUC*	*V-statistics*	*Adj. p.-value*	*Median ΔTSS*	*Mean ΔTSS*	*Median ΔKappa*	*Mean ΔKappa*
	Total Richness	0	0.003439	1337	1	0.000571	0.005941	0.00263	0.007632
Richness	**Pathotroph**	0.000129	0.002929	1320	0.74	0.000735	0.005454	0.001705	0.006357
	Bryophyte Parasite	2.00E-05	0.003012	1329	0.82	0.001671	0.005008	0.001136	0.006022
	Fungal Parasite	0.000357	0.00342	1304	0.43	0.002361	0.006154	0.002291	0.007025
	Plant Parasite	0.000408	0.003078	1376	0.92	0.001857	0.005881	0.002144	0.007147
	Animal Pathogen	0.000294	0.003266	1217	0.23	0.001736	0.006286	0.00172	0.00777
	Plant Pathogen	3.50E-05	0.002951	1377	1	0.00175	0.005299	0.001486	0.00597
	Lichen Parasite	0.000184	0.003022	1343	0.66	0.001393	0.005736	0.002756	0.006776
	Saprotroph	-4.60E-05	0.002796	1490	1	-0.000139	0.005031	0.000724	0.006419
	Litter Saprotroph	0	0.002983	1520	1	0.000385	0.004705	0.000776	0.006003
	Soil Saprotroph	0.000332	0.003605	1235.5	0.19	0.001389	0.006044	0.002188	0.007112
	Plant Saprotroph	6.90E-05	0.003058	1340	0.91	0.000972	0.004603	0.000788	0.005987
	Dung Saprotroph	-4.60E-05	0.002796	1490	1	-0.000139	0.005031	0.000724	0.006419
	Leaf Saprotroph	0.000197	0.003581	1351	0.72	0.000556	0.005333	0.001266	0.006439
	Wood Saprotroph	0	0.002894	1568	1	0.000714	0.004837	8.00E-04	0.005858
	Symbiotroph	0	0.002791	1475	1	-0.000199	0.004678	0	0.005957
	Epiphyte	0	0.002873	1428	1	0.001071	0.005779	0.001377	0.006498
	Lichenized	-3.50E-05	0.002933	1448.5	1	0.000286	0.004907	0.001093	0.00569
	Arbuscular Mycorrhizal	0.000423	0.0038	1289	0.25	0.001458	0.006426	0.001796	0.007235
	Endomycorrhizal	0	0.003208	1465	1	0.001525	0.006348	0.00083	0.007103
	Ericoid Mycorrhizal	0	0.002544	1484.5	1	0.00025	0.005228	0.000703	0.006115
	Orchid Mycorrhizal	0.000517	0.003013	1310	0.46	0.000216	0.005051	0.000784	0.006248
	Root-Associated Biotroph	0.000146	0.003473	1260	0.56	0.000597	0.00574	0.000735	0.006526
	Clavicipitaceous Endophyte	0.000491	0.002924	1258	0.25	0.002	0.005104	0.001934	0.005706
	Ectomycorrhizal	-0.000143	0.002804	1493	1	0.001389	0.00545	0.000752	0.006087
	Animal Endosymbiont	0.000463	0.002682	1361.5	0.56	0.001286	0.004916	0.001832	0.005615
	Endophyte	0.000491	0.002924	1258	0.25	0.002	0.005104	0.001934	0.005706
Rel. Abundance	**Pathotroph**	0.000255	0.003219	1344.5	0.10614854	0.002518	0.006857	0.001004	0.007335
	Saprotroph	0.000191	0.002939	1377.5	0.10425554	0.001429	0.005195	0.001574	0.00679
	Symbiotroph	0.000116	0.002912	1405	0.13583018	0.001875	0.004885	0.001274	0.00564

Table S5**. Average increase in AUC, maxTSS and maxKappa with the addition of fungal richness, or relative abundance, to GLM models with elastic net variable selection**. *P-values from the Wilcoxon test testing whether the median ΔAUC is superior to 0 are adjusted with a Bonferroni correction (p-value multiplied by 27). The median and mean performance changes are written as 0 when the absolute value is inferior to* 10^{-5}.

GBM	*Trophic group/guild*	*Median ΔAUC*	*Mean ΔAUC*	*V-statistics*	*Adj. p.-value*	*Median ΔTSS*	*Mean ΔTSS*	*Median ΔKappa*	*Mean ΔKappa*
	Total Richness	0.000703	0.002610	1448.5	1	0.001207	0.004097	0.001121	0.003693
Richness	**Pathotroph**	0.002315	0.002991	1365	0.58	0.000278	0.003202	0.000317	0.003467
	Bryophyte Parasite	0.001887	0.003164	1354	0.52	0.002778	0.005441	0.001548	0.004947
	Fungal Parasite	0.00149	0.003132	1354	0.52	0.00151	0.004507	-0.001137	0.004015
	Plant Parasite	0.001875	0.002671	1470.5	1	0.001806	0.00432	0.000172	0.003435
	Animal Pathogen	0.002755	0.002393	1375	0.64	0.000906	0.002346	0.00164	0.002272
	Plant Pathogen	0.000247	0.003429	1380	0.96	-0.000278	0.003534	0.002805	0.003421
	Lichen Parasite	0.000972	0.002528	1505	1	-0.001114	0.00297	-0.000728	0.004632
	Saprotroph	0.001364	0.002769	1382	0.98	0.003787	0.005215	0.002521	0.004137
	Litter Saprotroph	0.003289	0.004663	1264	0.18	0.003382	0.003884	0.003806	0.003809
	Soil Saprotroph	0.005245	0.006337	952	**0.0017**	0.006857	0.010176	0.007725	0.009255
	Plant Saprotroph	0.002972	0.003940	1300	0.28	0.006667	0.00592	0.005007	0.005689
	Dung Saprotroph	0.001364	0.002769	1382	0.98	0.003787	0.005215	0.002521	0.004137
	Leaf Saprotroph	0.001891	0.003089	1376	0.65	-0.000903	0.004334	-0.000125	0.003684
	Wood Saprotroph	0.003009	0.003689	1271	0.20	0.003024	0.004412	0.001894	0.004447
	Symbiotroph	0.001505	0.002381	1488	1	0.002857	0.002972	-5.10E-05	0.003105
	Epiphyte	0.005717	0.005635	983	**0.0029**	0.004857	0.007785	0.004287	0.007001
	Lichenized	0.003426	0.007787	959	**0.0019**	0.01	0.01366	0.008202	0.012225
	Arbuscular Mycorrhizal	0.000810	0.002296	1608	1	0.002941	0.002955	0.001821	0.001959
	Endomycorrhizal	0.003918	0.006046	996	**0.0036**	0.005295	0.009223	0.003621	0.006795
	Ericoid Mycorrhizal	0.004284	0.003948	1162	**0.0481**	0.004506	0.006973	0.00641	0.006631
	Orchid Mycorrhizal	0.001011	0.003416	1273.5	0.21	0.006875	0.008477	0.006257	0.00763
	Root-Associated Biotroph	0.000735	0.002372	1475	1	0.004306	0.008181	0.005521	0.007348
	Clavicipitaceous Endophyte	-0.001743	-0.000907	2204	1	-0.002794	-0.000761	-0.001531	-0.001646
	Ectomycorrhizal	0.002959	0.003256	1278.5	0.22	0.003472	0.00551	0.001563	0.004917
	Animal Endosymbiont	0.003184	0.005845	1029	**0.0063**	0.005429	0.006965	0.002073	0.005484
	Endophyte	-0.001743	-0.000907	2204	1	-0.002794	-0.000761	-0.001531	-0.001646
Rel. Abundance	**Pathotroph**	0.003033	0.006095	1098	**0.00210252**	0.006103	0.011414	0.007561	0.010961
	Saprotroph	6.80E-05	0.001172	1685	0.80070536	-0.000447	0.002689	-0.00095	0.001955
	Symbiotroph	0.000729	0.002228	1627	0.57125758	0.000656	0.001492	-0.001245	0.001954

Table S6**. Average increase in AUC, maxTSS and maxKappa with the addition of fungal richness, or relative abundance, to GLM models with elastic net variable selection**. *P-values from the Wilcoxon test testing whether the median ΔAUC is superior to 0 are adjusted with a Bonferroni correction (p-value multiplied by 27), and are in bold when significant (P<0.05).*