



Ecole de biologie

Identifying reliable indicator species of climate change: a case study with bryophytes in the Vaud Alps

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, specialization « Geosciences, Ecology and Environment »

par

Tenzin JHELIL

Directeur : Prof. Antoine Guisan

Superviseur : Dr. Flavien Collart

Expert·e·s : Anonymous

Département d'écologie et d'évolution

(Janvier, 2024)

Table of content:

Abstract:	3
Keywords:	4
Introduction:	4
Material and methods:	7
1) Species distributions models of bryophytes across Switzerland.....	7
2) Independent evaluation set.....	8
3) Evaluation of the models.....	9
4) Linking species traits to model performance.....	10
Results:	13
Discussion:	15
Conclusion:	17
Acknowledgment and Contributions:	18
References:	19
Supplementary materials:	23

Table of figures:

Figure 1: Map of the plots on the study area.....	9
Table 1: Table of the used traits for the correlations.....	11
Figure 2: Performances of the models with internal and external evaluation according to the evaluation metrics.....	13
Table 2: Best performing plants according to model performance.....	13
Table 3: Table of the importance of the variables and their correlation with the evaluation metrics.....	14
Table S1: Table of the used environmental variables for the computation of the niche breadth and the niche centroids distance.....	23
Figure S1: Correlation plot between all the traits variable used.....	24
Table S2: Performances of the models on all the plants, using internal and external evaluation.....	25-26

Abstract:

Global temperature has increased by 1.5° compared to the preindustrial era, endangering numerous ecosystems such as the Alps by notably increasing biodiversity loss. To protect the ecosystems and make conservation planning, species distribution models (SDM) appear as a powerful tool. However, the performances of these models strongly depend on species characteristics, thus impacting the reliability of these models to make efficient conservation planning. In this context, this study aims to determine the combination of traits impacting model performances and determine which are the key species to model when studying the impact of climate change. To answer these questions, we externally evaluated Swiss-scale SDMs of bryophytes and linked model performances to species traits. We showed that model performances were negatively explained by species temperature preferences, the number of habitats where the species lived, and the niche optimum distance, while positively explained by the reactivity. These results implied that specialized alpine species were better modeled than the generalist lowland species, which are the ones of lower conservation relevance. In addition, SDM were more accurate when the local environment was close to the environment used to generate those models, showing the importance of adding local habitat information to improve local model performances.

La température globale sur Terre a augmenté de 1.5°C comparée à l'époque pré-industrielle, menaçant de nombreux écosystèmes tels que les Alpes, principalement en augmentant leur perte de biodiversité. Pour protéger ces écosystèmes et y faire des plans de conservation, les modèles de distributions d'espèces (SDM) sont un bon outil. Cependant, les performances de ces modèles dépendent fortement des caractéristiques des espèces, impactant donc la fiabilité de ces modèles pour faire des plans de conservation efficaces. Dans ce contexte, cette étude vise à déterminer la combinaison de traits impactant les performances des modèles, et déterminer quelles sont les espèces clés à modéliser pour étudier l'évolution du changement climatique. Pour répondre à ces questions, nous avons évalué de manière externe les SDMs à l'échelle de la Suisse et avons lié les performances de modèles aux traits de ces espèces. Nous montrons que les performances de modèles étaient négativement expliquées par la préférence en de l'espèce au niveau de la température, du nombre d'habitats où elle peut être trouvée, et la distance par rapport à l'optimum de sa niche, et positivement expliquées par la réactivité du sol. Ces résultats montrent que des espèces alpines spécialisées étaient mieux modélisées que des espèces de basse altitude plus généralistes, étant moins concernées par les mesures de conservation. De plus, les SDMs étaient plus précis lorsque l'environnement local était proche de l'environnement utilisé pour créer ces modèles, montrant l'importance d'ajouter des informations d'habitat locales, pour améliorer les performances des modèles locaux.

Keywords: Climate change, Bryophytes, Species distribution models, External model evaluation, Conservation planning, Performance impacting traits.

Introduction:

The last report of the intergovernmental panel on climate change (IPCC) showed that there already has been an increase in global temperature of approximately 1.5°C compared to the preindustrial era (1850-1900), which can reach around 2°C of increase in 2050 (Pörtner *et al.*, 2022). With this change in temperature also comes a change in the frequency and the intensity of the precipitations (Trenberth, 2011). This alteration of the environmental condition leads to changes in a lot of environments, such as forest, peatlands or even aquatic ecosystems (Gignac & Vitt, 1994; Kirschbaum & Fischlin, 1996; Grimm *et al.*, 2013). This climate change has been shown to impact the alpine cryosphere by accelerating the melting of the glaciers over the decades (Cannone *et al.*, 2008; Ernakovich *et al.*, 2014; Chersich *et al.*, 2015). This leads to an extinction risk of around 14% of the species in the south western Alps in negative climate change scenarios (Dagnino *et al.*, 2020) and more globally leads to more species being threatened due to a shrinkage of their habitat (Callaghan *et al.*, 2011; Shivanna, 2020). Also, as biodiversity loss is exponential, the protection of it all is important to avoid an even stronger loss (Danovaro *et al.*, 2008), resulting in a loss in ecosystem services useful for humans, thus, creating a need for conservation planning (IPBES, 2018).

A powerful tool to make conservation planning is the species distribution models (Guisan *et al.*, 2013). It allows making predictions on present or even future suitable environments for given species, taking into account the environmental conditions of their midst (Guisan *et al.*, 2017). They are based on the actual known presence of species, as well as environmental variables to project on another area the predicted presence of the studied species (Guisan *et al.*, 2017). They are widely used in ecology and conservation (Araújo *et al.*, 2019), certainly because of the increasing availability of worldwide environmental databases (such as Chelsa (Karger *et al.*, 2017) or Worldclim (Hijmans *et al.*, 2005)), species information (Global Biodiversity Information Facility (gbif.org) or Swissbryophytes (swissbryophytes.org) and ready-to-use software and R packages (e.g. Maxent (Phillips *et al.*, 2004), Wallace (Kass *et al.*, 2023), biomod2 (Thuiller *et al.*, 2023)).

A limitation of the use of these models is that their performances depend on the biology of the studied species and its different traits (Regos *et al.*, 2019). It has been shown to have an impact on a large range of living species, such as fishes (Luan *et al.*, 2020) or plants (Syphard & Franklin, 2010; McCune *et al.*, 2020). Species dispersal capacities can impact model performances, with highly dispersive species performing better than the others (McCune *et al.*, 2020). Indeed, species with a

better capacity to disperse through the study region, matching more with the environmental condition where they grow (McCune *et al.*, 2020). Species life duration can also play a role in making accurate in angiosperms, where a longer lifespan implies better model performances (McCune *et al.*, 2020), as their ability to survive for a long time makes them less sensible to an alteration of their environment (McCune *et al.*, 2020). Their size also seems to have an impact on the model performances, where the models performed better on bigger species (Syphard & Franklin, 2010) that may be easier to spot during the samplings or more resistant to the environment too. Even the soil preference of the plant has been shown to impact the performances of the models in bryophytes (Collart, Broennimann *et al.*, 2023) as it may impact on the niche of the species, impacting the models. Another limitation of these models is their evaluation method, which is mostly internal, leading to higher performances shown by the model that it actually is, compared to when using an external validation method, with an independent dataset (Consonni *et al.*, 2010). The use of this independent dataset on the model will be to fully inform population demography across space on the SDM, making them more reliable (A. Lee-Yaw *et al.*, 2022).

This study focuses on bryophytes, the second more diverse group of land plants (Vanderpoorten & Goffinet, 2010; Shaw *et al.*, 2011). This plant group is known for its use to the rest of the species of its environment, as it is a good carbon and nitrogen source (Turetsky, 2003). Bryophytes are also reliable to assess air purity and naturalness due to its sensibility to its surroundings (Puglisi *et al.*, 2012). They are, however, predicted to be highly sensitive to climate change due to their ecophysiological features (He *et al.*, 2016). For example, they are poikilohydric, where the individual cannot self regulate the water in its cells. They thus strongly depend on the air humidity and rainfalls (Gignac, 2001). They are also a group of plants sensitive to temperature as it has a direct impact on their CO₂ intake (Tuba *et al.*, 2011). They also have a lower temperature optimum than the angiosperms, making them even more sensible to its variation (Zanatta, Engler & Collart *et al.*, 2020). Even though they are known to have good dispersal capacities (Vanderpoorten *et al.*, 2019), they are unfortunately predicted to be highly impacted by climate change and more specifically arctic-alpine species with a predicted range loss of 39±15% (Zanatta, Engler & Collart *et al.*, 2020). In this context, we need to perform conservation planning on these species. Studying the relationship between SDM performances and biological traits is thus of utmost importance to make more accurate conservation planning.

Here, we took advantage of two Swiss projects : the Valpar project and the RechAlp project. The Valpar project generated distribution maps of most of Switzerland's species (Reynard *et al.*, 2021). As it has national implications, it is thus important to test independently their distribution maps, to ensure that the ecosystem services in the environmental fields, society and economy are preserved. Independent evaluation has become possible with the RechAlp project, which is a project

promoting disciplinary and interdisciplinary research by creating an open field laboratory. The RechAlp project is a project that heavily studied the Vaud Alps, creating a lot of fine scale dataset (Däniken *et al.*, 2014), with bryophytes dataset (Collart, 2018), both permitting the realization of this study.

In this context, this study aims at finding indicator species to reliably predict global warming impact by externally evaluating the species distribution models. To do so, we questioned: (1) What is the drop of model performances when evaluating these models with an external evaluation; (2) Is the expected drop un model performances correlated with species traits; (3) Which trait combination are involved in model accuracy and which species can become indicators of climate change according to these traits ?

Material and methods:

1) Species distributions models of bryophytes across Switzerland

For the Valpar project, among all the modeled plants, 526 bryophyte species have been modeled following the Nested Species Distribution Model method (NSDM). An NSDM is a framework dealing with the niche truncation problem which can occur when the entire species niche is not captured with local data. This method thus uses two datasets, one at a global scale, allowing to capture the entire climatic niche of the studied species, as well as a regional one, where other fine-scale parameters such as land cover or topography can be used to refine the climatic predictions (Adde *et al.*, 2023). In the case of this study, a global distribution model has first been generated. To do so, Gbif occurrence data at the European scale and InfoSpecies data at the Swiss scales were combined with 19 bioclimatic variables to characterize the global niche of the species. Once this climatic niche has been generated, models are then projected onto the whole Switzerland at 25m resolution. This newly generated map of the niche at the Switzerland's scale is afterwards used as a variable for the local model.

In these models, the used covariates were selected by selecting 5000 candidate covariates (Külling *et al.*, 2024) that were then filtered using an embedded covariate selection (Adde *et al.*, 2023). This covariate selection is realized in two steps by using the covsel package (Adde *et al.*, 2023). The first step is the collinearity filtering, removing all the correlated covariates using univariate Generalized Linear Models (GLMs). The subset of covariates are then employed in multivariate models (GLM with elastic-net, (Zou & Hastie, 2005), the Generalized Additive Model (GAM) with null space addition (Marra & Wood, 2011), and the guided regularized random forest (Deng & Runger, 2013) to select the best combination of covariates. This embedded covariate selection permits to filter the initial covariate candidate to only the most informative one per species with the addition of the global climatic niche that was computed before, that is put as a forced variable (Adde *et al.*, 2023).

The final models were generated through a combination of 5 modeling methods: the GLM, GAM, the Maximum Entropy approach (Phillips & Dudík, 2008), the Random Forest (Breiman, 2001) and the Gradient Boosting Machine (Ke *et al.*, 2017).

To evaluate these models, an internal evaluation was performed by using a 100 fold repeated-split-sampling cross-validation, with 70% of the dataset to calibrate the models and 30% of the dataset for their evaluation. 4 evaluation metrics were computed to determine model performances: the Area under the curve (AUC) which varies between 0 and 1, a value of 0.5 meaning that the model explains the repartition the same as if it was random, a value under 0.5 is less effective than a random prediction and above 0.5 is more effective than a random prediction; the max True Skill Statistic (maxTSS); the max Symmetric Extremal Dependence Index

(max SEDI) and; the Boyce Index which only needs presence data. These three last metrics scale between -1 and 1, where 0 is as efficient as a random prediction, less than 0 is less effective as random prediction and above 0 is more efficient than a random prediction.

Models were then projected on a resolution grid of 25m covering the whole Switzerland to reflect the habitat suitability of each species in this country.

2) Independent evaluation set

Taking the advantage of the RechAlp project, the area of the Vaud's Pre-Alps region was selected as a study area. This study area is a mountainous area located in the canton de Vaud in Switzerland, ranging from 375 to 3210 meters of altitude with a soil mainly calcareous (Dubuis *et al.*, 2011). Between 2017 and 2021, bryophyte species composition data has been sampled in 575 plots of 2 by 2 meters, which were selected using a random stratified procedure (Collart, 2018; Kasprzyk, 2020; Collart *et al.*, under review). In March and August 2023, we increased the number of sampled plots to 650 (see Fig1). All these plots were selected using a random stratified method, taking the slope, the elevation and the aspect as variables, each of these variables separated into 3 categories (roughly, weak, medium and strong for the slope, low, medium and high for the altitude and finally, the aspect was angles comprised between [0-120°], [121-240°] and [241-360°]), following the sampling strategy of Dubuis *et al* (2011) and Hirzel & Guisan (2002).

Once on the field, to retrieve these plots, we used a Trimble R2 precision gps, which performs a real time correction to give accurate estimation of the position at the centimeter level. Presence/absence of bryophyte species was inventoried inside a 2*2 m quadrat. Individuals of each present species on the plot were sampled to be identified in a laboratory, using the floras from Ireland, United Kingdom (AJE Smith & Smith, 1990; A J E Smith & Smith, 2004), Scandinavia (Nyholm, 1986; Damsholt & Pugh, 2002) and Italy (Cortini-Pedrotti, 2001). These newly sampled plots will then be added to the previously sampled ones, creating a significant independent dataset for the evaluation of the NSDMs.

In total, 345 different species were sampled and identified among the 650 plots. This new dataset was then compared with the list of bryophytes that have been modeled using NSDM, reducing the total of available species to 210.

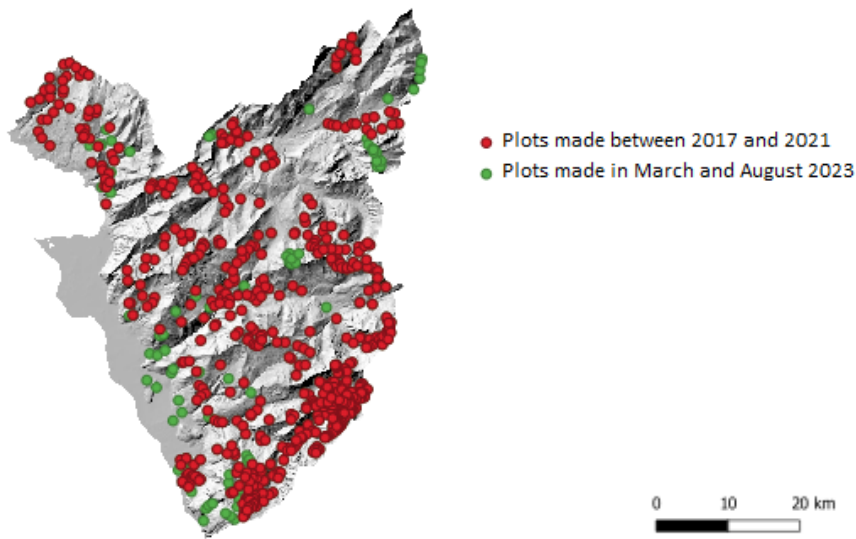


Fig 1: Map of the study area with the plots made before this study (red) and the ones added during this study (green)

3) Evaluation of the models

Here, to evaluate the models, 4 metrics were used, which are the ones that have already been used for the evaluation of the NSDM, which are the AUC, the boyce, the max TSS and the maxSEDI, to see if there was a significant drop in the performance of the models, as it can be expected when using external evaluation (A. Lee-Yaw *et al.*, 2022).

To evaluate the NSDM, the first step was to extract the presence probability of the species on the plots collected for the external dataset using the terra package (Hijmans *et al.*, 2023a) from R. Then, the AUC, the TSS, the boyce and the SEDI were measured with the ecospat (Broennimann *et al.*, 2023) and dismo (Hijmans *et al.*, 2023b) packages by using an external method of validation, using the independent dataset made in the Vaud pre-Alps.

During the evaluation, all the species with less than 5 occurrences in the independent dataset were removed, considering that species with less than 5 occurrences were not pertinent for the evaluation (Collart & Guisan, 2023).

This occurrence filter reduced the total of modeled species from 210 to 128 species. Once the external evaluation was made, the new performances were compared with the ones made from the internal evaluation using a Wilcoxon test (Wilcoxon *et al.*, 1970).

4) Linking species traits to model performance

To assess if biological traits impact the performance of the models on bryophyte species, we selected 7 predictors from the bet database (van Zuijlen *et al.*, 2023), known to potentially impact model performances. Size, to see if the various sizes of the individuals play a role in their predictability in bryophytes as it has been shown to have impact on other group of plants by affecting the final performance of the model, with a better performance for larger species (Syphard & Franklin, 2010), mean of the diameter of the spore, frequency of the spores and the position of the capsules to reflect the dispersal capacities. Lowest altitude at which it can be found, range of altitude where it can be found and sum of the habitats where it can be found to show the niche size.

To reflect species preferences to certain habitat values we used the Landolt values from swissbryophytes. Landolt values (Ivanova & Zolotova, 2023) range from 1 to 5 and they show the preference of the plant in the given variable. Here, the 4 variables using Landolt values are the temperature, the reactivity of the soil, the humidity and the light. For the temperature, a value of 1 represents an alpine or a nival preference for the plant while a temperature of 5 represents a hilly environment. For the reactivity, a value of 1 indicates a really acid soil while in the opposite, a value of 5 indicates a basic one. For the humidity, a value of 1 indicates a dry soil while a value of 5 indicates a drenched one and finally, for the light, a value of 1 indicates an area with a lot of shade while a value of 5 indicates an open one. In this study, Landolt values were used instead of Ellenberg values (as present in the bet data base where the other traits were taken from) to reflect more the local ecological preference of each species (Collart *et al.*, 2023a).

To complete this species trait database, we computed two metrics. The niche breadth, corresponding to the niche size, and the niche centroid distance, translating how far off its optimum niche condition the species actually is. To compute these metrics, all the available occurrences for each species with a resolution <25m were downloaded from swissbryophytes. At these localities, the 19 bioclim variables coupled with the aridity index, the annual evapotranspiration, the growing degree day (0°C) and the Relative Sunshine duration (Srel) (the definitions of the used environmental variables are available in figure S1) available at 25m for the whole Switzerland (CHCLIM25; Broennimann & Guisan, in prep) were also extracted.

After collecting this data, an OMI (Outlying Mean Index) was performed to compute the niche breadth (Dolédec *et al.*, 2000). The OMI is a method used to measure niche optimum and niche breadth along environmental gradients for a given species (Treier *et al.*, 2009). At first, a Principal Component Analysis (PCA) was made and 3 axes were kept, to reduce the total number of dimensions in the variable and to avoid multicollinearity while extracting as much information as possible from these variables, explaining 85% of the total variance. Then from these 3 PCA axes, an OMI was calculated and 2 OMI axes were kept, cumulating 89% of the projected inertia.

From these 2 OMI axes, the total area of the ellipse is calculated, serving as the niche breadth per species.

For the niche centroid distance, the climatic values were extracted for the 650 plots in the Vaud Alps, and then combined with the matrix at the Swiss scale. A PCA was then performed and 3 axes were kept, explaining 85% of the total variance. Then, for each species, the centroid of its climatic niche is calculated both at the Swiss and the Vaud-Alps scale, by calculating the median of each Principal component (PC). Finally, the niche distance was computed using the square root of each of these medians, following the formula:

$$\sqrt{(medPC1S - medPC1V)^2 + (medPC2S - medPC2V)^2 + (medPC3S - medPC3V)^2}$$

where med stands for median, S for Swiss and V for Vaud.

Finally, we also employed the number of occurrences used in the model as this may be an important factor in model fitting. The list of the used variables, their definitions and their source are available in table1.

Table 1: table of the used traits in the correlations, their definition and their source.

Variable	Definition	Source
size	the size of the species	BETdatabase
smeand	mean diameter of the spore	BETdatabase
sfreq	frequency of the spores per year	BETdatabase
capspos	position of the capsules on the plant: 1-erect, 2-erect to inclined, 3-immersed, 4-inclined, 5-inclined to pendulous, 6-pendulous	BETdatabase
limlow	lowest altitude where it can be found	BETdatabase
limrange	range of altitude where it can be found (limmax-limlow)	BETdatabase
habsum	sum of the habitats where it can be found	BETdatabase
npts	number of points used to generate the NSDMs	Valpar
humidity	Landolt value of the wetness, from 1 (very dry) to 5 (very wet)	Swissbryophytes
light	Landolt value of the light, from 1 (very shaded) to 5 (very lighted)	Swissbryophytes
reactivity	Landolt value of the pH, from 1 (very acid) to 5 (basic)	Swissbryophytes
temperature	Landolt value for the temperature, from 1 (nival) to 5 (collineous/very hot)	Swissbryophytes
EllipseArea	evaluation of the niche size	Newly generated
Distance	distance between the niche centroide of the global niche and the study area	Newly generated

To ensure that the predictors were not cross-related, Pearson correlation was calculated between the variables, to avoid multicollinearity in the linear regression (Dormann *et al.*, 2013) (see figure S1).

To link these predictors to model performance, a GLM with elastic-net regularization was made for each of the evaluation metric, accepting both the variable and its quadratic form as explanatory variables.

The GLMs with a regularization in elastic net were generated using the glmnet (Friedman *et al.*, 2023) and the biomod2 (Thuiller *et al.*, 2023) packages. To do so, the lambda parameter was determined by realizing a 10 fold cross-validation. The

lambda was chosen such that the error is within 1 standard error of the cross-validated errors for the minimum lambda.

The importance of each variable was then computed as the mean value of $1 -$ the correlation between the values predicted, using the usual dataset, and, using the dataset where the variable of interest was shuffled. The shuffling procedure has been applied 100 times following the same procedure as the function `bm_VariablesImportance` from the `biomod2` package (Thuiller *et al.*, 2023). Then, to compare the two models, these variable importance were rescaled from 0 to 1, by dividing each importance value from a model, by their maximal importance value.

All the statistical computations were made using the 4.2 version of R (R core team, 2022).

Results:

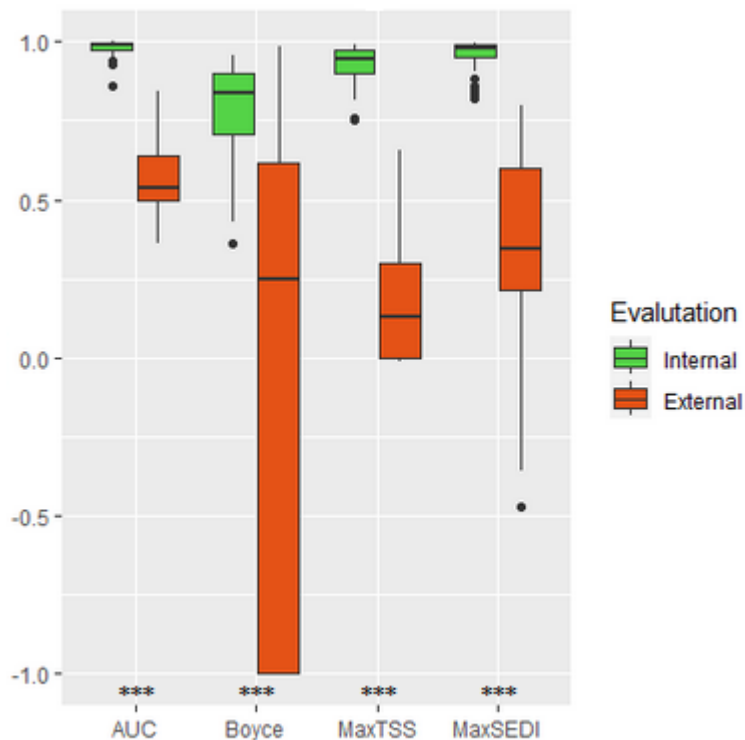


Fig 2: Performances of the NSDM using internal and external evaluation, 3 * means that after a Wilcoxon test, the drop was shown to be significant, with a p-value under 10^{-3} .

Model performances were significantly dropping when evaluating them with an external dataset (p-values < 0.001; Fig2, and with V values of 8256 for the AUC, the MaxTSS and the MaxSEDI, and of 8046 for the Boyce). The medians of these performances decrease from 0.990 to 0.538 for the AUC, 0.942 to 0.089 for the maxTSS, 0.980 to 0.241 for the maxSEDI and 0.840 to 0.29 for the Boyce Index.

Table 2: 5 best performing plant models on the used evaluation metrics, where AUC stands for Area Under the Curve, maxTSS stands for the max True Skill Statistic, maxSEDI stands for the max Symmetric Extremal Dependence Index and Boyce stands for the Boyce Index.

Species	AUC	maxTSS	maxSEDI	Boyce
<i>Orthothecium rufescens</i>	0.841	0.645	0.798	0.787
<i>Campylophyllum halleri</i>	0.824	0.523	0.695	0.82
<i>Scapania cuspiduligera</i>	0.813	0.528	0.726	0.794
<i>Hylocomiastrum pyrenaicum</i>	0.804	0.503	0.666	0.755
<i>Timmia norvegica</i>	0.801	0.654	0.707	0.72

According to our external evaluation, some species were better modeled than others (Table S2). The top 5 of the species having the best models were: *Orthothecium rufescens*, *Campylophyllum halleri*, *Scapania cuspiduligera*, *Hylocomiastrum pyrenaicum* and *Timmia norvegica* (Table2). To explain these differences in model performances between species, we realized linear models with elastic net

regularization to find the best combination of traits explaining those differences. It is shown that the size is positively correlated with the Boyce with an importance of 0.057 for this metric. The lower limitation impacted positively the maxSEDI with an importance of 0.3 and the limitation range impacted both the AUC and the maxTSS positively, with an respective importance of 0.542 and 0.324. The sum of habitats impacted negatively the AUC, the maxTSS and the maxSEDI with a respective importance of 0.252, 0.236 and 0.387. The reactivity impacted all four metrics positively, with an importance of 0.757 for the AUC, 0.062 for the Boyce, 0.92 for the maxTSS and 0.56 for the maxSEDI. The temperature impacted negatively on all the metrics except for the Boyce, with an importance of 0 for the AUC, 0.414 for the maxTSS and 0.43 for the maxSEDI. The number of points impacted positively the Boyce with an importance of 1, but negatively the maxSEDI with an importance of 0.656. Finally, the distance to the niche optimum impacted negatively on all 4 metrics, with an importance of 1 for the AUC, 0.499 for the Boyce, 1 for the maxTSS and 1 for the maxSEDI (Table3).

Table 3: table of the importance of the used variables on the different evaluation metrics used, with their correlation, where a blue case means a negative correlation while a green case means a positive correlation. Importance was scaled between 0 and 1, where the variable with a 1 is the most important one for the evaluation metric, while the others are less important the more they get close to 0. Where size is the size of the individual, smeand is the mean diameter of the spore, sfreq is the frequency of spores, capspos is the position of the capsules on the individuals, limlow is the lowest altitude where it can be found, limrange is the altitude range where it can be found, habsum is the sum of the types of habitats where it can be found, humidity is the preference of the plant toward wetness, reactivity is the preference of the plant for the pH of the soil, light is the preference of the plant regarding the light, temperature is the preference of the plant for the temperature, npts is the number of points used to make the NSDM, ellipseArea is the niche breadth and finally, distance is the distance of centroids.

	AUC	Boyce	maxTSS	maxSEDI
size		0.057		
smeand				
sfreq				
capspos				
limlow				0.3
limrange	0.542		0.324	
habsum	0.252		0.236	0.387
humidity				
reactivity	0.757	0.062	0.92	0.56
light				
temperature	0		0.414	0.43
npts		1		0.656
ellipseArea				
distance	1	0.499	1	1

Discussion:

In this project, we tested the drop of performances of the models between an internal and an external evaluation method. The model showed a significant decrease in performances after using an external evaluation, seen in fig x, where the median passes from 0.990 to 0.538 for the AUC, 0.942 to 0.089 for the maxTSS, 0.980 to 0.241 for the maxSEDI and 0.840 to 0.29 for the boyce. This significant decrease was expected as it is a constant result when comparing internal and external evaluations on SDMs (Consonni *et al.*, 2010; A. Lee-Yaw *et al.*, 2022). This drop has already also been shown in the European bryophytes (Collart *et al.*, 2023), which observed a drop of 0.98 to 0.62 in the AUC, of 0.87 to 0.27 in the maxTSS, and of 0.98 to 0.38 for the Boyce Index, which is a slightly lower drop in performance but similar than what was observed in this study. This slight difference in performance drop may come from the original model realization method (simple SDM in Collart *et al.*, while in this study, N-SDM were used). This observed drop is still higher than the average, with an average drop of 22% for internally evaluated models and 13% for externally evaluated models (Collart *et al.*, 2023). This comes also to question the reliability of the actually made SDMs to actually predict local populations (A. Lee-Yaw *et al.*, 2022) as such a drop in performances makes the median of the performances close to what a random prediction would produce. This drop in performance may be due to the source of data points, coming from GBIF, where the dataset could be biased depending on the investment that the different countries have in funding and sharing their datasets (Zizka *et al.*, 2021). Or the realization of the sampling, which has been made on 2 by 2m plots, to evaluate a model that was projected on a 25m by 25m resolution grid on the studied area. Regarding the results obtained on other species, 64% of the studies report an AUC > 0.70 for at least 75% of species considered after using an external evaluation method (A. Lee-Yaw *et al.*, 2022).

Various biological and ecological traits of the bryophytes showed either a positive or a negative correlation to the performances (see table3).

At first, it is noticeable that the sum of habitats where the species can live have an impact on the models as it affects each evaluation metric but has a rather low importance on the models. This still means here that species with a small number of possible habitats will be better modeled. Specialist species are thus more easily modeled than the more generalist one. This finding is in accordance with (Syphard & Franklin, 2010) that showed that species with a smaller range were better modeled, but is in opposition with the results from (Collart *et al.*, 2023) who found no correlation between model performance and number of occupied habitats for bryophytes in southern Belgium.

It is also remarkable that the niche centroid distance was impacting the models with a large importance on the models this time. Indeed, species closer to their niche

optimum were better modeled than the other ones. This means that for the realization of the models, when projecting a model at local scale, the niche of the species must be known to apply a type of corrector during the computation of the models to take into account a possible bias due to a niche optimum not present in the studied area. A way to apply this corrector would be to use the same method than the used NSDMs (Adde *et al.*, 2023) and compute a third nesting, taking into account the global, regional and finally the local scales, to further improve these models and thus, the conservation plans that depend on them.

On the other hand, the reactivity of the soil was also a variable that impacted the models on each evaluation metric but, with a positive correlation, meaning that the more basic the soil is, the more performing will be the model on the given species. It is also visible that the importance of this variable varies depending on the metric, with globally a high importance except for the boyce index. As the area is a calcareous one, the ground is basic on almost all the studied area (Swiss Geoportal, n.d.). But it is still notable that the pH tends to be even higher on the higher altitude areas (swisstopo, n.d.). This means that plants present in higher areas of the Vaud Alps will tend to be better modeled at higher altitude. These results are in opposition to what was found at the European scale for bryophytes, where the reactivity of the soil, which was a variable of high importance for the model performances, correlated negatively with the performances, meaning that the lower the pH, the more performing were the models (Collart *et al.*, 2023), which may be correlated with the exploitation of the areas where Southern Belgium is vastly exploited for agriculture, while the Vaud Alps are not, making a bias in the nature of the environment during the studies as it was shown that the level of human disturbance was shown to impact negatively on the models (Collart *et al.*, 2023). The last highly correlated variable also tends to prefer species that grow on the top of the mountains, as the temperature is negatively correlated with the performances. This leads to a preference toward cold environment adapted species, naturally present towards the top of the mountains. Finally, the elevation of the species is the last variable that tends to indicate that species in higher altitude are better modeled, even if it has a rather small impact on the global model performances.

Lastly, the other studied variables, the size, the dispersibility and the niche breadth, do not seem to have a significant impact on model performances, as the size and the spore frequency only have a slight impact on the Boyce Index metric. For the size, it may be that bryophytes do not vary enough in size compared to the other studied plant group in (Syphard & Franklin, 2010) which showed that models performed better on shrub size plants than on sub shrub plants, but as the bryophytes only vary in size at the centimeter or even the millimeter scale, such a small variation may not be sufficient to impact the models. For the dispersibility, it may be due to the dispersion method of the bryophytes, which uses wind dispersed spores, giving them the ability to fully englobe their ecological niche (Frahm, 2009) unlike the angiosperms studied in (McCune *et al.*, 2020), relying on other dispersal methods,

which showed that the species with the more dispersing seeds were better modeled. As bryophytes all use the same dispersal method which was proven to be sufficient enough to englobe their whole habitat, the difference between the studied species was not large enough to observe variation in the performances of the models. Finally, the niche breadth did not impact the model performances at all which is surprising as it would have been expected to be negatively correlated, following the results of (Syphard & Franklin, 2010) which showed that species with smaller range sizes were better modeled.

In the context of climate change, there is a need to conserve the endangered species efficiently. In this study, a few species were greatly modeled, with performances that were still considered as good even after the external evaluation.

The species that performed the best (see Table2) has in common their preference towards cold environment (ranging from 1-nival to 3 mountainous temperatures in Landolt values) as well as basic soils (ranging from 4- neutral to basic to 5-basic for the reaction in Landolt values), with the values extracted per species on swiss bryophytes. Also, these species are all present in the same type of environment which is the mountainous south of Switzerland, toward the top of the mountains. These models are well fitted to the actual distribution of the species according to the external evaluation, as well as the threshold (Guisan *et al.*, 2017). As these performances are above 0.8 for the AUC, the models were considered very good on this metric while only good for the other evaluation metrics which is still encouraging compared to the drastic drop that was observed in the total dataset. These species, or species containing the traits that have been proven to impact the models, could be used to evaluate the impact of climate change on the studied area as key species by seeing their repartition evolve by using environmental maps predicting the evolution of the climate in the next decades.

Conclusion:

Biological traits impact the models, by having a direct impact on the model performances. Traits such as the preference towards the cold, basic soils, a small number of suitable habitats and a short distance to the niche optimum all improve the model performances, making these plants more reliable on their repartition and more globally, on the evolution of the climate conditions on the future using future predictors.

On a conservation aspect, these results are encouraging as bryophytes living in the summits of the Vaud's Pre Alps tend to be in a higher risk of extinction due to the melt of the glaciers and the permafrost, in a context of climate change (Cannone *et al.*, 2008; Chersich *et al.*, 2015; Dagnino *et al.*, 2020), giving more value to these areas to be protected.

Bryophytes are plants that are on the centimeter or even the millimeter scale (Vanderpoorten & Goffinet, 2010). However, to generate SDMs, the same climatic and other biotic variables as for the other reign of plants of a much greater size. As a group sensitive to climate, they may be more dependent on microclimate from their niche that is not taken into account while using a 25m buffer as it was used for the realization of these models (Stewart & Mallik, 2006). To improve these models, reducing the scale of the models to ones that fit more the studied species, such as the plot size (here, 2 by 2m) may improve the models for future projects (such as suggested in Collart et al, in prep).

Acknowledgment and Contributions:

During this master's work, I would like to first thank my supervisor, Flavien Collart, which helped me during the whole course of this project, by explaining me the used concepts, advising me on the methods, providing the datasets, helping me during the statistical analysis and also during the redaction of this paper, by giving me regular feedbacks, permitting to increase the quality of this report.

I would also like to mention Antoine Adde, that provided the NSDMs and their performances, Olivier Broennimann, who helped me by giving me part of the R script, and Alain Venderpoorten for his expertise in the field and identification of the bryophytes species along the years that composed the independent dataset of this study.

Finally, I would like to thank the whole laboratory of Professor Antoine Guisan, as well as Antoine Guisan himself, for their feedback during the presentations of my early work, permitting its evolution on the right path.

References:

- A. Lee-Yaw, J., L. McCune, J., Pironon, S. & N. Sheth, S. 2022. Species distribution models rarely predict the biology of real populations. *Ecography* **2022**: e05877.
- Adde, A., Rey, P., Brun, P., Külling, N., Fopp, F., Altermatt, F., *et al.* 2023a. *N-SDM*: a high-performance computing pipeline for Nested Species Distribution Modelling. *Ecography* **2023**: e06540.
- Adde, A., Rey, P.-L., Fopp, F., Petitpierre, B., Schweiger, A.K., Broennimann, O., *et al.* 2023b. Too many candidates: Embedded covariate selection procedure for species distribution modelling with the covsel R package. *Ecol. Inform.* **75**: 102080.
- Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., *et al.* 2019. Standards for distribution models in biodiversity assessments. *Sci. Adv.* **5**: eaat4858. American Association for the Advancement of Science.
- Breiman, L. 2001. Random Forests. *Mach. Learn.* **45**: 5–32.
- Broennimann, O., Cola, V.D., Petitpierre, B., Breiner, F., Scherrer, D., D`Amen, M., *et al.* 2023. ecospat: Spatial Ecology Miscellaneous Methods.
- Callaghan, T.V., Johansson, M., Brown, R.D., Groisman, P.Ya., Labba, N., Radionov, V., *et al.* 2011. Multiple Effects of Changes in Arctic Snow Cover. *AMBIO* **40**: 32–45.
- Cannone, N., Diolaiuti, G., Guglielmin, M. & Smiraglia, C. 2008. Accelerating Climate Change Impacts on Alpine Glacier Forefield Ecosystems in the European Alps. *Ecol. Appl.* **18**: 637–648.
- Chersich, S., Rejšek, K., Vranová, V., Bordoni, M. & Meisina, C. 2015. Climate change impacts on the Alpine ecosystem: an overview with focus on the soil. *J. For. Sci.* **61** (2015): 496–514.
- Collart, F., Broennimann, O., Guisan, A. & Vanderpoorten, A. 2023a. Ecological and biological indicators of the accuracy of species distribution models: lessons from European bryophytes. *Ecography n/a*: e06721.
- Collart, F. & Guisan, A. 2023. Small to train, small to test: Dealing with low sample size in model evaluation. *Ecol. Inform.* **75**: 102106.
- Collart, F., Kiebacher, T., Quetsch, M., Broennimann, O., Guisan, A. & Vanderpoorten, A. 2023b. *Predicting tomorrow's biodiversity: what do we miss at local scale?*
- Collart, F. & Université de Liège > Master biol. orga. & écol., F. 2018. Modélisation d'une communauté de bryophytes le long d'un gradient altitudinal à fine échelle. Université de Liège, Liège, Belgique.
- Consonni, V., Ballabio, D. & Todeschini, R. 2010. Evaluation of model predictive ability by external validation techniques. *J. Chemom.* **24**: 194–201.
- Cortini-Pedrotti, C. & Lara, F. 2001. Flora dei muschi d'Italia: Orthotrichales. pp. 719–761.
- Dagnino, D., Guerrina, M., Minuto, L., Mariotti, M.G., Médail, F. & Casazza, G. 2020. Climate change and the future of endemic flora in the South Western Alps: relationships between niche properties and extinction risk. *Reg. Environ. Change* **20**: 121.
- Damsholt, K. 2002. *Illustrated flora of Nordic liverworts and hornworts*. Knud Graphic Consult, Odense.
- Däniken, I. von, Guisan, A. & Lane, S. 2014. RechAlp.vd: une nouvelle plateforme UNIL de support pour la recherche transdisciplinaire dans les Alpes vaudoises. *Bull. Société Vaudoise Sci. Nat.* **94**: 175. Société Vaudoise des Sciences Naturelles.
- Danovaro, R., Gambi, C., Dell'Anno, A., Corinaldesi, C., Fraschetti, S., Vanreusel, A., *et al.* 2008. Exponential Decline of Deep-Sea Ecosystem Functioning Linked to Benthic Biodiversity Loss. *Curr. Biol.* **18**: 1–8.
- Deng, H. & Runger, G. 2013. Gene selection with guided regularized random forest. *Pattern Recognit.* **46**: 3483–3489.
- Dolédec, S., Chessel, D. & Gimaret-Carpentier, C. 2000. Niche Separation in Community Analysis: A New Method. *Ecology* **81**: 2914–2927.
- Dormann, C., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., *et al.* 2013.

- Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**: 27–46.
- Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Divers. Distrib.* **17**: 1122–1131.
- Ernakovich, J.G., Hopping, K.A., Berdanier, A.B., Simpson, R.T., Kachergis, E.J., Steltzer, H., *et al.* 2014. Predicted responses of arctic and alpine ecosystems to altered seasonality under climate change. *Glob. Change Biol.* **20**: 3256–3269.
- Evidence, C. 2018. Conservation Evidence | IPBES secretariat.
- Frahm, J.-P. 2009. Diversity, dispersal and biogeography of bryophytes (mosses). In: *Protist Diversity and Geographical Distribution* (W. Foissner & D. L. Hawksworth, eds), pp. 43–50. Springer Netherlands, Dordrecht.
- Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., *et al.* 2023. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models.
- Géosciences et environnement: CHCLIM25: generating high spatial and temporal resolution climate maps for Switzerland on Apple Podcasts. n.d.
- Gignac, L.D. 2001. Bryophytes as Indicators of Climate Change. *The Bryologist* **104**: 410–420. American Bryological and Lichenological Society.
- Gignac, L.D. & Vitt, D.H. 1994. Responses of Northern Peatlands to Climate Change: Effects on Bryophytes. *J. Hattori Bot. Lab.* **75**: 119–132.
- Grimm, N.B., Chapin III, F.S., Bierwagen, B., Gonzalez, P., Groffman, P.M., Luo, Y., *et al.* 2013. The impacts of climate change on ecosystem structure and function. *Front. Ecol. Environ.* **11**: 474–482.
- Guisan, A., Thuiller, W. & Zimmermann, N.E. 2017. *Habitat Suitability and Distribution Models: with Applications in R*. Cambridge University Press.
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., *et al.* 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* **16**: 1424–1435.
- He, X., He, K.S. & Hyvönen, J. 2016. Will bryophytes survive in a warming world? *Perspect. Plant Ecol. Evol. Syst.* **19**: 49–60.
- Hijmans, R.J., Bivand, R., Pebesma, E. & Sumner, M.D. 2023a. terra: Spatial Data Analysis.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**: 1965–1978.
- Hijmans, R.J., Phillips, S. & Elith, J.L. and J. 2023b. dismo: Species Distribution Modeling.
- Hirzel, A. & Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecol. Model.* **157**: 331–341.
- Ivanova, N. & Zolotova, E. 2023. Landolt Indicator Values in Modern Research: A Review. *Sustainability* **15**: 1–22. MDPI.
- Karger, D.N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., *et al.* 2017. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**: 170122. Nature Publishing Group.
- Kasprzyk, T. & Université de Liège > Master biol. orga. & écol., F. 2020. Mécanismes écologiques et évolutifs de la variation des communautés de bryophytes le long d'un gradient altitudinal. Université de Liège, Liège, Belgique.
- Kass, J.M., Pinilla-Buitrago, G.E., Paz, A., Johnson, B.A., Grisales-Betancur, V., Attali, D., *et al.* 2023. wallace: A Modular Platform for Reproducible Modeling of Species Niches and Distributions.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., *et al.* 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Kirschbaum, M.U.F. & Fischlin, A. 1996. Climate change impacts on forests.
- Külling, N., Adde, A., Fopp, F., Schweiger, A.K., Broennimann, O., Rey, P.-L., *et al.* 2024. SWECO25: a cross-thematic raster database for ecological research in Switzerland. *Sci. Data* **11**: 21. Nature Publishing Group.

- Luan, J., Zhang, C., Xu, B., Xue, Y. & Ren, Y. 2020. The predictive performances of random forest models with limited sample size and different species traits. *Fish. Res.* **227**: 105534.
- Marra, G. & Wood, S.N. 2011. Practical variable selection for generalized additive models. *Comput. Stat. Data Anal.* **55**: 2372–2387.
- McCune, J.L., Rosner-Katz, H., Bennett, J.R., Schuster, R. & Kharouba, H.M. 2020. Do traits of plant species predict the efficacy of species distribution models for finding new occurrences? *Ecol. Evol.* **10**: 5001–5014.
- Nyholm, E. 1986. *Illustrated Flora of Nordic Mosses*. Nordic Bryological Society.
- Phillips, S.J. & Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161–175.
- Phillips, S.J., Dudík, M. & Schapire, R.E. 2004. A maximum entropy approach to species distribution modeling. In: *Proceedings of the twenty-first international conference on Machine learning*, p. 83. Association for Computing Machinery, New York, NY, USA.
- Pörtner, H.-O., Roberts, D.C., Tignor, M.M.B., Poloczanska, E.S., Mintenbeck, K., Alegría, A., *et al.* (eds). 2022. *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.
- R core team (2022). R: A language and environment for statistical computing. R foundation for Statistical computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rastorfer, J.R. & Rastorfer, J.B. 2010. Vanderpoorten, A. and B. Goffinet. Introduction to Bryophytes. *Trans. Ill. State Acad. Sci.* **103**: 159–164. Illinois State Academy of Science.
- Reynard, E., Grêt-Regamey, A. & Keller, R. 2021. The ValPar.CH project – Assessing the added value of ecological infrastructure in Swiss Parks. *Ecomont J. Prot. Mt. Areas Res.* **13**: 64–68.
- Shaw, A.J., Szövényi, P. & Shaw, B. 2011. Bryophyte diversity and evolution: windows into the early evolution of land plants. *Am. J. Bot.* **98**: 352–369.
- Shivanna, K.R. 2020. The Sixth Mass Extinction Crisis and its Impact on Biodiversity and Human Welfare. *Resonance* **25**: 93–109.
- Smith, A.J.E. 1990. *The Liverworts of Britain and Ireland*. Cambridge University Press.
- Smith, A.J.E. 2004. *The Moss Flora of Britain and Ireland*. Cambridge University Press.
- Stewart, K.J. & Mallik, A.U. 2006. Bryophyte Responses to Microclimatic Edge Effects Across Riparian Buffers. *Ecol. Appl.* **16**: 1474–1486.
- Swiss Geoportal - <https://map.geo.admin.ch>
- swisstopo - <https://swisstopo.admin.ch>
- Syphard, A.D. & Franklin, J. 2010. Species traits affect the performance of species distribution models for plants in southern California. *J. Veg. Sci.* **21**: 177–189.
- Thuiller, W., Georges, D., Gueguen, M., Engler, R., Breiner, F., Lafourcade, B., *et al.* 2023. biomod2: Ensemble Platform for Species Distribution Modeling.
- Treier, U.A., Broennimann, O., Normand, S., Guisan, A., Schaffner, U., Steinger, T., *et al.* 2009. Shift in cytotype frequency and niche space in the invasive plant *Centaurea maculosa*. *Ecology* **90**: 1366–1377.
- Trenberth, K.E. 2011. Changes in precipitation with climate change. *Clim. Res.* **47**: 123–138.
- Tuba, Z., Slack, N.G. & Stark, L.R. 2011. *Bryophyte Ecology and Climate Change*. Cambridge University Press.
- van Zuijlen, K., Nobis, M.P., Hedenäs, L., Hodgetts, N., Calleja Alarcón, J.A., Albertos, B., *et al.* 2023. Bryophytes of Europe Traits (BET) data set: A fundamental tool for ecological studies. *J. Veg. Sci.* **34**: e13179.
- Vanderpoorten, A., Patiño, J., Désamoré, A., Laenen, B., Górski, P., Papp, B., *et al.* 2019. To what extent are bryophytes efficient dispersers? *J. Ecol.* **107**: 2149–2154.
- Wilcoxon, F., Katti, S.K. & Wilcox, R.A. 1970. Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. *Sel. Tables Math. Stat.* **1**: 171–259.

Zanatta, F., Engler, R., Collart, F., Broennimann, O., Mateo, R.G., Papp, B., *et al.* 2020. Bryophytes are predicted to lag behind future climate change despite their high dispersal capacities. *Nat. Commun.* **11**: 5601.

Zou, H. & Hastie, T. 2005. Regularization and Variable Selection Via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**: 301–320.

Supplementary materials:

Table S1: Used environmental variables and their definition for the computation of the niche breadth and niche centroids distance.

Variable	Definition
Bio1	Annual mean temperature
Bio2	Mean diurnal range (mean of monthly(max temp – min temp))
Bio3	Isothermality (Bio2/Bio7)*100
Bio4	Temperature seasonality (standard deviation*100)
Bio5	Max temperature of the warmest month
Bio6	Min temperature of the coldest month
Bio7	Temperature annual range (Bio5-Bio6)
Bio8	Mean temperature of the wettest quarter
Bio9	mean temperature of the driest quarter
Bio10	mean temperature of the warmest quarter
Bio11	mean temperature of the coldest quarter
Bio12	Annual precipitation
Bio13	Precipitation of the wettest month
Bio14	Precipitation of the driest month
Bio15	Precipitation seasonality (coefficient of variation)
Bio16	Precipitation of the wettest quarter
Bio17	Precipitation of the driest quarter
Bio18	Precipitation of the warmest quarter
Bio19	Precipitation of the coldest quarter
Slope reliability (Srel)	Duration of daily sunshine
Aridity Index (AI)	Index of aridity (evapotranspiration/precipitation values)
Growing Degree Day (0°C)	sum of growing degree-days above 0°C
Annual Evapotranspiration	Annual evapotranspiration

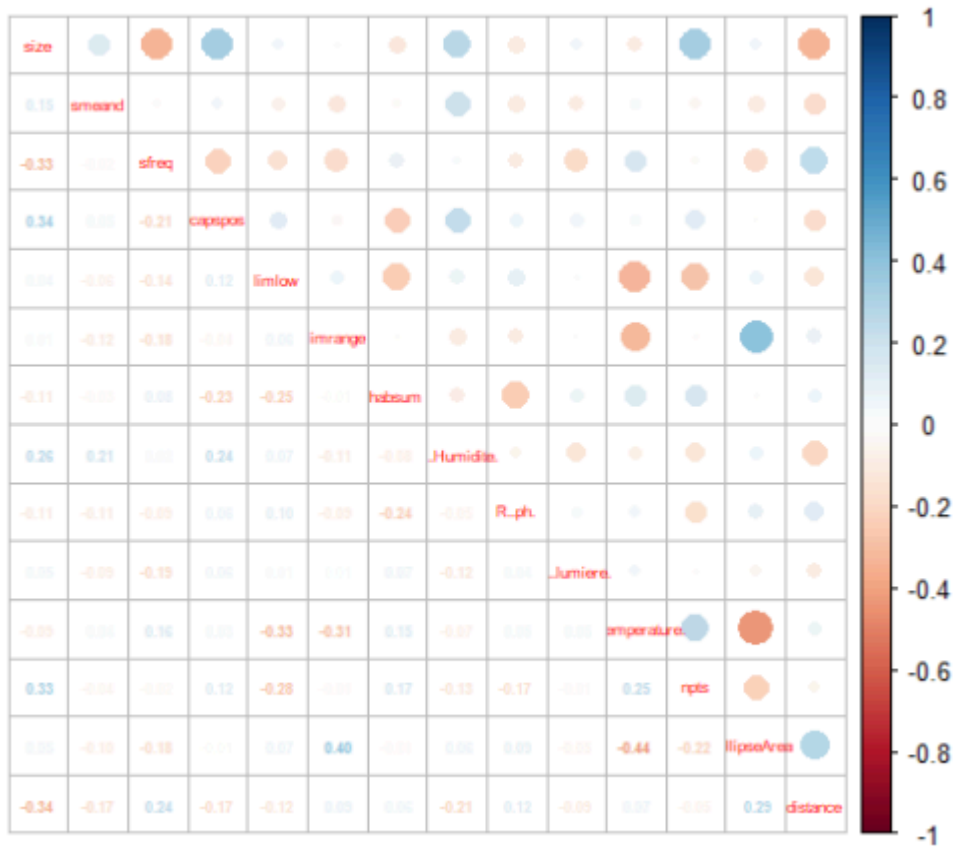


Fig S1: correlation between the selected predictors showed by the colored circles and their numerical equivalent, where size is the size of the individual, smeand is the mean diameter of the spore, sfreq is the frequency of spores, capspos is the position of the capsules on the individuals, limlow is the lowest altitude where it can be found, limrange is the altitude range where it can be found, habsum is the sum of the types of habitats where it can be found, humiditee is the preference of the plant toward wetness, lumiere is the preference of the plant regarding the light, R.ph is the preference of the plant for the pH of the soil, temperature is the preference of the plant for the temperature, npts is the number of points used to make the NSDM, EllipseArea is the niche breadth and finally, distance is the distance of centroids.

Table S2: Performances of all the plants studied after internal and external evaluation. The presence of an (i) means that the method of evaluation was internal, while its absence means that the method of evaluation was external.

species	AUC(i)	Boyce(i)	MaxTSS(i)	MaxSEDI(i)	AUC	Boyce	MaxTSS	MaxSEDI
<i>Abietinella abietina</i>	0.990203863439121	0.901268	0.954707702435813	0.988428594898261	0.59008229323824	0.48	0.182416678096283	0.268193893516267
<i>Amblystegium serpens</i>	0.996452516876406	0.868242	0.975227281799109	0.995057384204564	0.518773605865363	-0.4	0.0464341257498333	0.0981963057257381
<i>Aneura pinguis</i>	0.993462915906847	0.824845717171717	0.975671148844579	0.996575540267065	0.568868407578085	0.613	0.161290322580645	0.45997108615237
<i>Atrichum undulatum</i>	0.99669114078811	0.87053	0.975431889400922	0.994878116981192	0.508364644838075	0.617	0.0249597423510466	0.0962965765276638
<i>Barbilophozia hatcheri</i>	0.993562605733083	0.804682	0.960425657894737	0.992312940935707	0.497695852534562	-1	-0.00153609831029189	-Inf
<i>Barbilophozia lycopodioides</i>	0.977708968819659	0.919354	0.900737096256341	0.966900940808086	0.752216748768473	0.565	0.4395656068062986	0.596311344731631
<i>Barbula unguiculata</i>	0.996641326539954	0.860084	0.978167101552772	0.996119695386115	0.511212325145412	0.214	0.0226409652453972	0.167506979997354
<i>Blepharostoma trichophyllum</i>	0.98654063410818	0.941978	0.92418296939753	0.986057259896557	0.539622481081541	-0.246	0.105603265028484	0.151812504664226
<i>Brachythecium velutinum</i>	0.998761425973995	0.863116897959184	0.985878696044055	0.995895186403699	0.515760233918129	0.495	0.0453508771929825	0.22658293625155
<i>Brachythecium glareosum</i>	0.930495389674468	0.958616	0.812336126274252	0.92008188489044	0.597702205882353	0.481	0.194393382352941	0.310304702205343
<i>Brachythecium mildeanum</i>	0.96939326765192	0.797256	0.908359154351396	0.97284653220492	0.50968792866941	0.402	0.104938271604938	0.405189028870347
<i>Brachythecium rivulare</i>	0.991685656661921	0.929193191919192	0.946701997073695	0.984203620747931	0.610034715706512	0.546	0.229452054794521	0.33846541233885
<i>Brachythecium rutabulum</i>	0.999148392177451	0.875242	0.986856536290852	0.99746139367809	0.650131251726996	-0.039	0.300013815971263	0.471924456141398
<i>Brachythecium salebrosum</i>	0.991195538774165	0.90902	0.952107793732018	0.987450156754631	0.614701291329644	0.421	0.202171136653995	0.2875216980078
<i>Bryoetrichophyllum recurvirostrum</i>	0.99428679100667	0.921856	0.836406652784875	0.934589525943169	0.664914089347079	0.911	0.311408693747904	0.610776688018196
<i>Bryum argenteum</i>	0.9892503428062	0.93558	0.952877271454826	0.988428492809823	0.565335753176044	0.558	0.159709618874773	0.281182221093875
<i>Bryum elegans</i>	0.94028679100667	0.921856	0.836406652784875	0.934589525943169	0.664914089347079	0.911	0.311408693747904	0.610776688018196
<i>Bryum pallens</i>	0.989941712345129	0.548140429610046	0.843034160387896	0.84806905060569	0.5	-1	-Inf	-Inf
<i>Bryum pseudotriquetrum</i>	0.988712793671335	0.930018	0.94876689036925	0.986803616880562	0.601084438408692	0.031	0.300427384566225	0.445159337195757
<i>Bryum rubens</i>	0.99808904901672	0.702911586206897	0.972022212352095	0.984187214610209	0.496632996632997	-1	-0.00168350168350173	-Inf
<i>Bryum weigelii</i>	0.984152759490892	0.457740683544304	0.927230635631775	0.940015004410573	0.469325153374233	-1	-0.00153374233128833	-Inf
<i>Calliergonella cuspidata</i>	0.99847977259136	0.832244363636364	0.984806723609286	0.995774623134535	0.510748170235836	0.651	0.022499323209691	0.075182925044814
<i>Calypogeia azurea</i>	0.958795350875849	0.909756	0.856883124627311	0.945865570303003	0.738780729252599	0.659	0.459660122092064	0.622150291362079
<i>Campylodictyon chrysophyllum</i>	0.995809119786438	0.76173	0.970994407482902	0.994036527107517	0.613406966864911	0.186	0.235140186918588	0.375036700814292
<i>Campylopus stellatum</i>	0.997563533484223	0.867552	0.976638694359384	0.9970662424321	0.515367903696844	0.544	0.0450174513854262	0.205130148311962
<i>Campylopus halleri</i>	0.937833574244414	0.896558	0.821676025697182	0.92818132541739	0.824450549405049	0.82	0.522370486656201	0.695299111543751
<i>Cephalozia bicuspudata</i>	0.994814295100434	0.867644	0.967850110467119	0.992900596538158	0.494565217391304	-1	-0.00155279503105588	-Inf
<i>Cephalozia pleneiceps</i>	0.976705571274746	0.574760956521739	0.927111254037676	0.968578766687571	0.462432312883436	-1	-0.00153374233128833	-Inf
<i>Ceratodon purpureus</i>	0.99705333484223	0.867552	0.976638694359384	0.9970662424321	0.515367903696844	0.544	0.0450174513854262	0.205130148311962
<i>Climacium dendroideum</i>	0.9885256233792	0.950188	0.932921053604787	0.98034566444839	0.611317447111956	0.29	0.233885878063038	0.347326026964122
<i>Critoneuron filicinum</i>	0.986799602721917	0.936606	0.928244148102254	0.978570768879042	0.640337019483399	0.699	0.265007898894155	0.40883862353039
<i>Critoneuron filicinum</i>	0.996182634886353	0.898913690721649	0.967377139685059	0.989720321123249	0.557270233196159	-0.5	0.1296296296296	0.243287966194715
<i>Cratichneumon mollicum</i>	0.99756098062504	0.92045042553192	0.966340181933619	0.98677419922338	0.564921707736871	0.034	0.133369715795986	0.204866738243045
<i>Dichodontium pellucidum</i>	0.996480820749753	0.482928972995259	0.97465077004275	0.979426505793019	0.498341625207297	-1	-0.00165837479270314	-Inf
<i>Dicranella subulata</i>	0.96222853318928	0.786275102040816	0.889808995874149	0.961834290574753	0.518605459920014	0.411	0.074334898786603	0.107179349407518
<i>Dicranella varia</i>	0.996953157273103	0.714889673469388	0.974971222937867	0.9924047210219	0.486760124610592	-1	-0.00155763239875384	-Inf
<i>Dicranum scoparium</i>	0.9981436239404378	0.88699	0.985696428260852	0.99733442272947	0.455953768453768	0.605	0.00320512820512819	0.0522002345060786
<i>Dicranum spadiceum</i>	0.97393058660019	0.783806	0.913396410894107	0.97335615336243	0.74183256491528	0.717	0.487687431425727	0.709201598821249
<i>Didymodon fallax</i>	0.996251364494954	0.638916648648649	0.948092044707429	0.95835631264519	0.506914607948443	0.143	0.0171186898810955	0.191277616915174
<i>Didymodon ferrugineus</i>	0.9855708331375	0.862026	0.9296344624792	0.9935082168852	0.79776793528505	0.818	0.43586285947612	0.63729032050605
<i>Didymodon rigidulus</i>	0.99799214294562	0.652078421052632	0.98568135951748	0.99790019652775	0.387090483619345	-1	-0.00156006240294607	-0.358517210747244
<i>Distichium capillaceum</i>	0.979738770303092	0.925278	0.9104913489026	0.971694299614584	0.80050908786828	0.634	0.519085788561525	0.736868318562587
<i>Ditrichum flexicaule</i>	0.982729594634826	0.748212	0.942758050110469	0.991577863941566	0.519490484960098	0.814	0.0384317146454703	0.162893541303009
<i>Encalypta alpina</i>	0.986916148342823	0.586474	0.952749360270112	0.988992808975058	0.726569096538137	0.294	0.514776254962464	0.674983472878484
<i>Encalypta thaptocarpa</i>	0.979279783637593	0.463233696969697	0.96052985600363	0.991732679906469	0.5	NA	-Inf	-Inf
<i>Encalypta rhaptocharpa</i>	0.99692578181205	0.843062	0.9753366314063	0.994574977056968	0.551896434249375	0.621	0.103849647967295	0.215691650416452
<i>Entodon concinnus</i>	0.988099277077116	0.889836	0.945496251841007	0.984938872510456	0.683798053126547	0.615	0.317934432468273	0.485172547813241
<i>Ephemeron minutissimum</i>	0.9502039386657	0.786142	0.871911312060467	0.94867111254488	0.53061851237639	-1	-0.077114279806966	0.124151779026486
<i>Eurhynchium striatum</i>	0.997290414878399	0.748546536585366	0.967530749769199	0.976231503200665	0.58312883435828	-0.345	0.19846576586812	0.612474946604242
<i>Fissidens dubius</i>	0.98976371783052	0.917648	0.942119808215529	0.983716293030395	0.641960154848456	0.678	0.293503918421301	0.436706947103885
<i>Fissidens osmundioides</i>	0.988026232693709	0.528655621368323	0.987763205646602	0.995315851892155	0.542266803840878	0.04	0.0864197530864197	0.264323387505
<i>Fissidens taxifolius</i>	0.999010050597918	0.665390111111111	0.971405695213392	0.97544325958071	0.53986395471379	0.496	0.0820177801724138	0.145347256500432
<i>Fissidens viridulus</i>	0.982389459691923	0.361313919872601	0.816390147783251	0.81805943947941	0.518614808652246	0.675	0.0392550594247682	0.087487833723406
<i>Gymnostomum aeruginosum</i>	0.976077812584956	0.868942	0.899925204006652	0.967605742529509	0.63852544132918	0.425	0.293769470404984	0.42921400105286
<i>Heterocladium dimorphum</i>	0.982148145882254	0.877688	0.920726178914015	0.975641737808722	0.47585	0.49	0.302124609982857	0.302124609982857
<i>Homalothecium lutescens</i>	0.991469045497597	0.893184	0.947639085385918	0.986264217253687	0.642357253943568	0.766	0.297822706065319	0.427542299407339
<i>Homalothecium incurvatum</i>	0.988274182243001	0.848638	0.954473807181505	0.988236816377958	0.703630734590487	0.763	0.48001688713763	0.64347996452911
<i>Hylocomium pyrenaicum</i>	0.97065345067532	0.927204	0.876798366022888	0.955903598698156	0.804801077932668	0.755	0.502621053857803	0.665737241935623
<i>Hylocomium splendens</i>	0.997581045405476	0.873432	0.983552067353349	0.996873276355311	0.535767218831735	0.685	0.087576285963828	0.169649486031495
<i>Hypnum cupressiforme</i>	0.997601541042231	0.841932	0.98489869376763	0.997703026548961	0.544458323870089	0.704	0.0902793549852372	0.168603262374235
<i>Hypnum vaucherii</i>	0.935535752390638	0.768964	0.850894911603697	0.943731020747716	0.68343582822086	0.685	0.401226993865031	0.551065607618416
<i>Isoterygopsis pulchella</i>	0.988452504443957	0.615804	0.927289502520272	0.952837584681119	0.496056782334385	-1	-0.00157728706624605	-Inf
<i>Jungfermannia atroviridis</i>	0.99273792741148	0.853306	0.950403785871391	0.989885885109317	0.48608964513138	-1	-0.00154559505409579	-Inf
<i>Kiaeria starkei</i>	0.95659593461268	0.86777	0.867297571251076	0.952582728701307	0.736042944785276	0.047	0.599693251533742	0.613885014595752
<i>Lophocolea bidentata</i>	0.996405884821843	0.838774202020202	0.980290185336544	0.99688571442076	0.474299065420651	-1	-0.00155763239875384	-Inf
<i>Lophocolea minor</i>	0.972710654018343	0.90435	0.896645740584202	0.966116652417584	0.718581713054815	-0.248	0.402783288929199	0.678508619935254
<i>Lophozia ventricosa</i>	0.99454121638849	0.6157						

Platydictya jungermanniioides	0.95851835990252	0.859662	0.880353241834215	0.958239817691644	0.759784106568672	0.836	0.431832797427653	0.609371740418775
Pleurodium subulatum	0.992574134340931	0.439133746278575	0.918993421052632	0.925133610367915	0.4984375	-1	-0.00156250000000002	-Inf
Pleurozium schreberi	0.98448727201034	0.948024	0.933523655455972	0.981122206457796	0.512517146776406	-0.2	0.0817901234567902	0.241776956477772
Pohlia cruda	0.987568554085163	0.898094	0.933069078450368	0.980572903491307	0.626349650349665	0.516	0.234933248569612	0.335639085368209
Pohlia drummondii	0.961504391064867	0.775105	0.886240896572827	0.961688835847552	0.772169059011164	0.112	0.525199362041467	0.736198862837384
Pohlia wahlenbergii	0.994731242129916	0.684616348139256	0.941048216381539	0.955143531861136	0.488372093023256	-1	-0.00166112956810627	-Inf
Polytrichum juniperinum	0.997696076177036	0.799562	0.982770690866252	0.997164709267874	0.4984	-1	-0.00160000000000005	-Inf
Polytrichum umigerum	0.995589365340349	0.762034	0.979550825386386	0.996185325283011	0.497560975609756	-1	-0.00162601626016257	-Inf
Preissia quadrata	0.983436598366669	0.859008	0.921339685178846	0.977223281435271	0.698566985114419	-0.5	0.402910464341258	0.600233293282783
Pseudoleskea incurvata	0.967806272172554	0.928332	0.867671219769304	0.951548614959154	0.769120115080317	0.984	0.499472548549508	0.748531248180444
Pseudoleskeella catenulata	0.933310112812045	0.944006	0.819176693040228	0.925044854646501	0.594017314232712	0.215	0.23482428115016	0.41250379956624
Pseudocleropodium purum	0.996752232142857	0.720872947368421	0.981424500187723	0.992367295371089	0.526307189542484	0.712	0.0617647058823529	0.3301386690495
Ptychodium plicatum	0.946170829288169	0.905714	0.825099454652236	0.928719496569387	0.699060710194731	0.898	0.42213058419244	0.61723769050987
Racomitrium canescens	0.98495473065831	0.899162	0.932348324719819	0.980093179879957	0.606985871271586	-0.5	0.23021978021978	0.337466612370282
Radula complanata	0.99055952384833	0.820484	0.988945972173773	0.998139262871312	0.443697409668372	0.491	-0.0141065830721003	-0.16935155586975
Rhodobryum roseum	0.988343542260214	0.733494	0.958260355029586	0.994135138201874	0.490580847723705	-1	-0.00156985871271587	-Inf
Rhynchostegium murale	0.990296375308898	0.937006	0.94335369061264	0.984089050164648	0.571175321049693	0.246	0.16750418760469	0.24097128316891
Rhytidadelphus squarrosus	0.997147958065173	0.80151	0.982737179186045	0.997548609208577	0.542819639562935	0.578	0.0859514687100893	0.242252140179427
Rhytidadelphus triquetrus	0.998222565541866	0.875534	0.982790823924229	0.996676312420443	0.565007716049383	0.012	0.131365740740741	0.27337207181019
Rhytidium rugosum	0.997826178751161	0.56516962962963	0.955233946737511	0.961083254456469	0.491433021806854	-1	-0.00155763239875384	-Inf
Sanionia uncinata	0.988418636624207	0.948238	0.931042621687184	0.979515808302996	0.577496641289745	0.657	0.197044334975369	0.284641501334536
Scapania aequiloba	0.939276664169322	0.914354	0.816559286271272	0.923086770574474	0.651078482328482	0.644	0.271335758835759	0.380805890788131
Scapania cuspiduligera	0.958139008822972	0.736296	0.888215280645767	0.962739243599526	0.8125935090895478	0.794	0.528301886792453	0.726499567723827
Scapania iriquia	0.995736525521215	0.564516343434343	0.983906486494048	0.99458682466056	0.498405103668262	-1	-0.00159489633173848	-Inf
Schistidium apocarpum	0.998097994125377	0.841316912482066	0.943608371338023	0.9439684185201	0.49607535321821	-1	-0.00156985871271587	-Inf
Schistidium dupretii	0.940088070766311	0.622724	0.757910690436096	0.829672136513435	0.682820395039045	0.576	0.41070280202113	0.620335782589634
Solenostoma gracillimum	0.950673755270489	0.862428	0.853068470565956	0.944027835618544	0.523006134968325	0.616	0.214110429447853	0.427193876229437
Sytrichia norvegica	0.981721914036818	0.675776	0.939755177514793	0.985484541508093	0.361202771819832	-1	-0.00156739811912221	-0.473885830628076
Sytrichia ruralis	0.945141554467576	0.916992	0.835055800252958	0.933930650836042	0.739774436090226	0.898	0.40596992481203	0.648044679330288
Thuidium assimile	0.994040470645387	0.614419662173913	0.903625569732387	0.912727971884726	0.5	-1	-Inf	-Inf
Thuidium austriacum	0.947812763157911	0.9395	0.834551315789474	0.933242756061743	0.578474081284984	0.527	0.139011925042589	0.232167305697975
Thuidium norvegicum	0.966760699183436	0.753064	0.919003062117235	0.977020227226012	0.613988179003659	0.725	0.215029552490853	0.425594759948573
Timmia norvegica	0.862626067323469	0.660362	0.753726190476191	0.861082384908495	0.801203277009729	0.72	0.654377880184332	0.707342152305276
Tortella fragilis	0.97948409227339	0.769188	0.925294400137209	0.980828847950159	0.5	-1	-Inf	-Inf
Tortella inclinata	0.95059556282962	0.898294	0.85909043602333	0.94723913709743	0.376583169152829	-0.286	-0.00154798761609909	-0.136519964574668
Tortella tortuosa	0.995858123092623	0.922694	0.972182458025454	0.99369912982211	0.514449031438552	0.43	0.0624347865535544	0.250346264234517
Tortula hoppeana	0.97132330544952	0.922072	0.888126889299353	0.961277433271366	0.776116303219107	0.361	0.445794392523365	0.608376376380719
Trichodon cylindricus	0.99810208174238	0.505949387755102	0.988646290053208	0.995090296220118	0.5	-1	-Inf	-Inf
Trichostomum crispulum	0.991147749999999	0.569478909090909	0.920368517894737	0.930175886399514	0.480769230769231	-1	-0.00153846153846149	-Inf
Weissia brachycarpa	0.980295174284782	0.681910352941176	0.944157370108517	0.971518955983558	0.490030674846626	-1	-0.00153374233128833	-Inf
Weissia controversa	0.998136840551134	0.706518	0.984366805259874	0.997999456907953	0.507471496997909	-0.101	0.0185185185185186	0.226407679790493
Weissia longifolia	0.96627764504441	0.820794	0.881336061004261	0.950401269466289	0.621668297737609	0.46	0.25801623395314	0.428602616368738