Ecology and Evolution

Open Access

Combining food web and species distribution models for improved community projections

Loïc Pellissier[1,*], Rudolf P. Rohr[2,*], Charlotte Ndiribe[3], Jean-Nicolas Pradervand[3], Nicolas Salamin[3], Antoine Guisan[3,†] & Mary Wisz[1,†]

[1]Department of Bioscience, The Arctic Research Centre, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus, Denmark
[2]Integrative Ecology Group, Estación Biológica de Doñana (EBD – CSIC), Calle Américo Vespucio s/n, Sevilla 41092, Spain
[3]Department of Ecology and Evolution, University of Lausanne, Bâtiment Biophore, Lausanne CH-1015, Switzerland

Keywords
Biotic interactions, ecological niche modeling, phylogeny, plant–herbivore interactions, trophic network.

Correspondence
Loïc Pellissier, Department of Bioscience, The Arctic Research Centre, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus, Denmark. Tel: +45 8715 8707; E-mail: loic.pellissier@unil.ch

Funding Information
This study was supported by the National Science Foundation (NSF) Grant Nr. 31003A-125145 (BIOASSEMBLE project), the Danish Council for Independent Research Grant no. 12-126430 and FP7-REGPOT 2010-1 EcoGenes Project (Grant No. 264125).

Received: 5 June 2013; Revised: 3 September 2013; Accepted: 9 September 2013

doi: 10.1002/ece3.843

*L. P and R. P. R — Co-first authors.
†A. G and M. W — Co-last authors.

Abstract

The ability to model biodiversity patterns is of prime importance in this era of severe environmental crisis. Species assemblage along environmental gradients is subject to the interplay of biotic interactions in complement to abiotic filtering and stochastic forces. Accounting for complex biotic interactions for a wide array of species remains so far challenging. Here, we propose using food web models that can infer the potential interaction links between species as a constraint in species distribution models. Using a plant–herbivore (butterfly) interaction dataset, we demonstrate that this combined approach is able to improve species distribution and community forecasts. The trophic interaction network between butterfly larvae and host plant was phylogenetically structured and driven by host plant nitrogen content allowing forecasting the food web model to unknown interactions links. This combined approach is very useful in rendering models of more generalist species that have multiple potential interaction links, where gap in the literature may occur. Our combined approach points toward a promising direction for modeling the spatial variation in entire species interaction networks.

Introduction

Sound predictions of the composition and function of future ecosystems are needed to inform decision-making in the face of global change but remain one of the greatest challenges facing ecological scientists (Mokany and Ferrier 2011; Nogués-Bravo and Rahbek 2011). Predicting which species will occur when and where, and their potential interactions requires an understanding of the complex network of trophic linkages that vary in space and time, as well as the associated competitive and facilitative effects (Kissling et al. 2012; Wisz et al. 2013). Species distribution models are spatially explicit models that are used to fill the gaps in our knowledge of spatial distributions of biodiversity, and recent advances in these are used to generate community-level forecasts. These models have only recently begun to incorporate the effect of biotic interactions, but so far, only account for these effects indirectly (e.g., based on correlations in occurrence patterns, Kissling et al. 2012) and cannot divulge information about the way species may or may not interact (Guisan and Thuiller 2005). In contrast to these modeling tools, food web models are used to

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

understand and predict interaction links between species within an interaction network in a given location, as well as measure patterns in the structure of food webs (Naisbit et al. 2012), but these models are generally not spatially explicit (but see Massol et al. 2011). Combining food web models with species distribution models to predict spatial variation in community composition remains an unexplored area in biodiversity modeling but would be a major advancement for predictive ecology.

A common approach to improve species distribution and community models is to consider the distribution of interacting species as a predictor in statistical models (Wisz et al. 2013). This approach is particularly straightforward when examining pairs of species that are known to interact, for example, specialist species when there are at most only a few obligate interactions (e.g., Araújo and Luoto 2007). In such cases, considering positive and negative correlations between species occurrence patterns in models of species distribution and communities generally improves model accuracy (Pellissier et al. 2010; Boulangeat et al. 2012; Giannini et al. 2013; Le Roux et al. 2013). Accounting for correlations in the distributions between species within distribution models can generate useful information on the potential distribution of species given the presence/absence of the known interacting species and has enabled improved forecast of species' fate under future climate change (Schweiger et al. 2012; Giannini et al. 2013). However, this approach is more difficult to apply to more generalist species because it may be unknown how these species interact with others, especially in cases where they do not yet co-occur but may do with global change (Wisz et al. 2013). The difficulty in accounting for species interactions is further heightened when the goal is to model all species in a region and predict community composition or richness. In this case, species distribution modeling lacks the capacity to predict occurrences for all the species in a region, while accounting for multiple and often partially known interactions (Wisz et al. 2013).

Recently, a new approach for modeling food webs was proposed by Rohr et al. (2010). This approach is based on a statistical model that relates the presence and absence of known trophic interactions (links) between species to measureable traits that relate to trophic relationships (such as energy content, size, digestibility, morphological characteristics linked with feeding, etc.) and latent traits (Naisbit et al. 2012). These latent traits are surrogates for properties of interacting species derived from the combined information in the measured traits and are estimated *a posteriori* from the matrix of trophic interactions. The food web model can be interpolated by fitting and predicting linkages between species to the original dataset (here referred as "interpolation") or extrapolated by projecting the model on to an independent dataset to infer potential links between poorly known species, or species in a different or hypothetical community (Clauset et al. 2008). The capacity to extrapolate a food web model has so far never been exploited, nor have food web models been coupled with species distribution models to predict assemblage structure and function in space and time (Thuiller et al. 2013).

Here, we propose combining the strength of food web models and species distribution models to account for multiple interactions between species. This approach aims to improve community projections by filling in gaps in our knowledge about food webs through accounting for multiple and potentially unknown interaction links across species. Biotic interactions in species distribution models among species of higher or lower trophic positions are quite under-represented (Van der Putten et al. 2010) compared to within trophic level (e.g., Ovaskainen et al. 2010; Laughlin et al. 2012). Extrapolation of food web models within a species distribution modeling framework is therefore a promising approach to fill gaps in our knowledge about the interactions between species and to produce more ecologically realistic forecasts. We test our approach on a comprehensive spatial dataset of plant and herbivore communities distributed along broad environmental gradients in the Swiss Alps (Pellissier et al. 2013).

Because phylogenies provide relevant information to understand and predict species trophic associations and assemblages (Mouquet et al. 2012; Whitfeld et al. 2012; Best et al. 2013; Pellissier et al. 2013), we reconstructed two species-level phylogenies for plant species and butterfly species. We also measured several commonly used plant palatability related traits (specific leaf area, leaf dry matter content, leaf carbon, and nitrogen content Wright et al. 2004) that are used to fit the food web models. Leaf palatability traits represent important components mediating plant–herbivore interactions (Ibanez et al. 2013). We then compared the accuracy of species distribution and community models using only abiotic predictors and considering as a predictor the links inferred from the food web models. We use this approach to address the questions (1) Can statistical food web models be used for extrapolation to an independent community? (2) Does accounting for inferred interaction links in species distribution models improve the accuracy of species distribution and community composition projections?

Methods

General framework

The core of our methodology resides in combining statistical models for inferring a plant–butterfly food web and

 © 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

spatial distribution models for projecting the spatial distributions of butterflies. Food webs are only partially known, as observations of herbivores feeding on plants are likely to be incomplete. Furthermore, in the illustration dataset (see description below), 38% of the plant–butterfly pairs of species are never detected in the same place, and for these pairs, the assessment of the presence or absence of a trophic link from field observation is impossible. The inference of the linking probabilities was achieved in two steps. First, we inferred the food web consisting of a species subset with at least one trophic link recorded. We used a statistical model based on latent traits (Rohr et al. 2010) to infer the probabilities of trophic links between plant and butterfly pairs. The model used measurable plant traits as predictors and latent traits to increase the prediction ability. Latent traits are surrogates for properties of interacting species (relating to e.g., foraging sucsceptibility, palatability, or nutrition content) derived from the combined information in the measured traits and are estimated *a posteriori* from the matrix of trophic interactions. The probability of occurrence of species in the landscape was calculated so that they were constrained by their relationship to the abiotic environment, and also the "best-case scenario" trophic link probabilities of the butterfly and plant species present at a given place. This methodology is presented in Fig. 1.

Food web model

The aim of our statistical food web model is to infer the probability of a trophic link between pairs of butterfly–plant species. This inference is based on the know matrix of trophic interaction a_{ij}, where $a_{ij} = 1$ if butterfly species j (as a caterpillar) feeds plant species i. With such binary

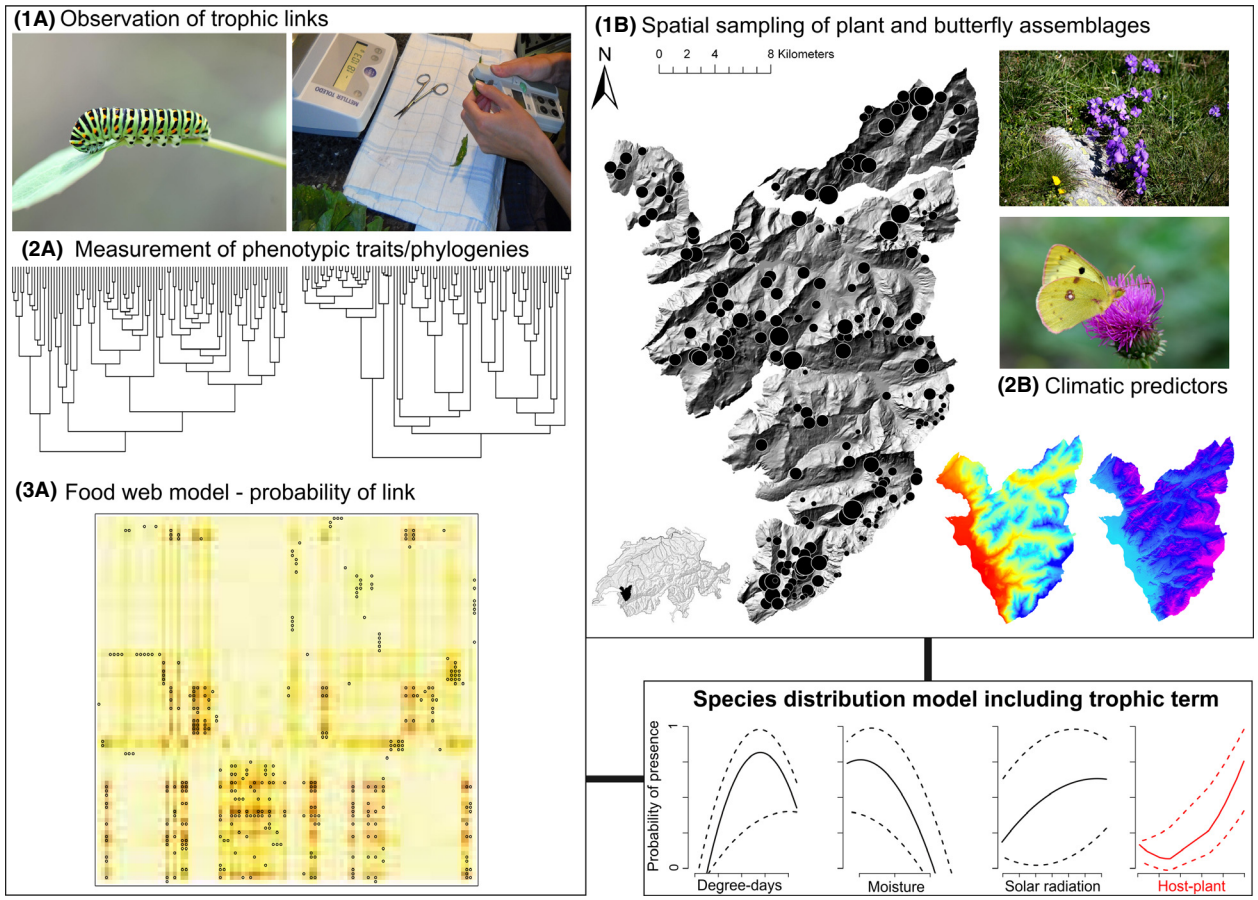


Figure 1. Schema of the methodology and data needed to combine food web and species distribution models. We propose fitting a food web model on observations of trophic interactions between plant and herbivores (1A) in conjunction with traits/phylogenies measured for plant and insects (2A). The food web model provides a probability of link between each pair of plant and herbivore (3A). In parallel, spatial sampling of plant and herbivores (1B, size of the dots proportional to butterfly species richness) allows collecting presence and absence of individual species. The species distribution model finally consists of relating presence and absence of a given herbivore species to climatic predictors (2B), but including a trophic term consisting in the maximal link probability with the plant species at a given site.

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

data, the standard statistical approach is a logistic regression (also called generalized linear model with binomial distribution) and we model the probability P_{ij}^{FW} of a trophic links between species i and j. As a probability takes value between zero and one, the standard approach is to write the model for the logit of the probability. We model this probability using a similar function as in Rohr et al. (2010):

$$\text{logit}\left(P_{ij}^{FW}\right) = \alpha + \beta_1 N_j + \beta_2 \text{Avg}H_j + f_i^1 \delta^1 v_j^1 + f_i^2 \delta^2 v_j^2. \quad (1)$$

This model has two parts: (1) a standard linear part, where N_j and AvH_j denote the nitrogen level and average height of plant species i (β_1 and β_2 are the corresponding slopes, and α is the intercept) and (2) a latent trait part ($f_i^1 \delta_1 v_j^1$ and $f_i^2 \delta_2 v_j^2$ are the corresponding parameter estimates) that quantifies, at the species level, what cannot be captured by the linear part. Each species is characterized by two latent traits, f_i^1 and f_i^2 for the butterfly species j, and v_j^1 and v_j^2 for plants species i, and δ_1 and δ_2 are scaling factors. Leaves with higher nitrogen content are more palatable and nutritious to herbivores, and larger conspicuous plants may potentially be more susceptible to attract herbivores (Pellissier et al. 2012b). Consequently, the two latent traits of each butterfly species can be viewed as species-specific foraging traits, while the plant latent traits can be seen as species-specific susceptibility traits. These latent traits aimed to quantified nonmeasured species' characteristics that are important in explaining the presence or absence of trophic link, while taking into account the measured species' traits (here, the nitrogen level and the average height of plant species). For the latent traits, as the product of the plant vulnerability latent traits and butterfly foraging trait increases, so does the probability that the butterfly feeds on the plant during its caterpillar stage. The value of the latent traits is not know a priori, and they are parameters to be inferred in the food web model from data on observed trophic interactions.

The likelihood function of our latent trait model is given by:

$$L = \prod_{z_{ij}=1} \left(P_{ij}^{FW}\right)^{a_{ij}} \cdot \left(1 - P_{ij}^{FW}\right)^{1-a_{ij}}, \quad (2)$$

where, as the trophic interactions are only partially observed, the product only takes into account pairs of species co-occurring at least once, that is, $z_{ij} = 1$. We calculated the number of co-occurrence using the formula presented in Pellissier et al. (2010). We fit the parameters of the model using a Monte Carlo Markov Chain procedure (specifically the Metropolis–Hasting algorithm), adapted from Rohr et al. (2010), equation (2) as the likelihood function. The advantage of using the likelihood function (2) is all the parameters, and the latent traits for all species can be inferred, even if the trophic network is incompletely known. Based on the inferred values, and using the equation (1), we can predict the probability of interaction between all pairs of butterfly and host plant species in the food web.

Phylogeny as predictor for latent traits

Our observations used to compute values of latent traits come from phylogenetically related species and cannot be considered as independent observations. To account for phylogenetic pseudoreplications, following Felsenstein (1985), we used the phylogenetic correlation to predict the latent traits and also to estimate the trophic link probabilities for all pairs of species.

Here, we first examined whether there was such correlation in the latent traits, using a phylogenetic regression (Freckleton et al. 2002). Such correlation is expected, as plant–herbivore interactions display generally a strong phylogenetic component (Pellissier et al. 2013). A phylogenetic regression is similar to a standard linear regression, only that it does not assume independence of data points, but rather that they are correlated (Freckleton et al. 2002), that is,

$$v_i \sim N(\tilde{v}, C(\lambda)),$$

where the variance–covariance matrix $C(\lambda)$ is given by:

$$C(\lambda)_{ij} = \begin{cases} \sigma^2 & i = j \\ \sigma^2 \lambda t_{ij} & i \neq j \end{cases}.$$

Here, σ denotes the variance. This variance–covariance matrix derives the correlation between two latent traits given by:

$$\text{cor}(v_i, v_j) = \lambda t_{ij},$$

where λ, called Pagel's λ,is the parameter that controls for the strength in the correlation from the phylogeny and t_{ij} is the fraction of time common between species i and j on the phylogenetic tree. When the parameters of the phylogenetic regression are fitted, they can be used to predict the latent traits of a new species using conditional expectation of a normal distribution:

$$v_n = \tilde{v} + \sum_{ij} C(\lambda)_{ni} \left(C(\lambda)^{-1}\right)_{ij} \left(v_j - \tilde{v}\right). \quad (3)$$

where the sum is made over all species for which at least one interaction was recorded. Using equation (3), the latent traits values are predicted for all the species with unknown information. Based on these predicted latent traits, the linking probabilities between all pairs of plant–butterfly species are estimated using equation (1). The phylogenetic regression is fitted in R (R Development Core Team 2013), using the library ape. We tested

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

the predictive power of the food web model both internally (fit) and externally using cross-validation. We measured it with the area under the curve (AUC) of a receiver operating characteristic plot (ROC). We tested whether the ability of the food web model to predict butterfly larvae association along with host plants differed among the main butterfly families using a Kruskal–Wallis test.

Species distribution modeling

We modeled the species distributions using the presence and absence for each species with four different techniques: generalized linear model (GLM), generalized additive model (GAM), gradient boosted model (GBM), and a random forest model (RF). These four statistical techniques are among the most commonly used in species distribution modeling. To produce predictions independent of the calibration data, we ran a tenfold cross-validation. We randomly split the dataset into ten partitions, successively calibrating the models with 90% of the data and sequentially predicting the species distributions and communities based on the remaining 10% data. We used three steps to model the distribution of each species. First, we modeled the distribution of the species using only abiotic predictors.

For the GLM model, the probability of presence of a given butterfly species at a site k is then given by

$$\text{logit}\left(P_k^{\text{PR}}\right) = \alpha + \beta_1 \text{ddeg}_k + \beta_2 \text{mind}_k + \beta_3 \text{srad}_k.$$

where ddeg_k denotes the degree-days, mind_k the moisture, and srad_k the solar radiations. Second, in concert with the abiotic predictors, we added a biotic predictor comprising the presence or absence at a given site of the obligatory trophic partner from the biotic interactions as provided by the literature. Third, alongside the abiotic predictors, we added a biotic predictor comprising the highest probability of an interaction link with any of the plants occurring at the site given by the food web model. For the GLM model, the probability of presence of a given butterfly species at site k is then given by

$$\text{logit}\left(P_k^{\text{PR}}\right) = \alpha + \beta_1 \text{ddeg}_k + \beta_2 \text{mind}_k + \beta_3 \text{srad}_k + \beta_4 \max_{\substack{\text{plant } j \\ \text{present in site k}}} \left(P_{ij}^{\text{FW}}\right).$$

where P_{ij}^{FW} is the maximum or "best-case scenario" trophic link probability predicted by our food web model.

We calculated the predictive power of the SDMs for each species, technique and set of predictors using the AUC. In addition, to obtain the community at each site, we stacked all the species predicted using the threshold independent resampling approach developed in Dubuis et al. (2011) and Pellissier et al. (2013). We evaluated the projected community composition to observed values using the Sørenson index and compared the accuracy between abiotic and biotic SDMs. The Sørenson index is computed as follow:

$$\text{Sørenson index} = \frac{2a}{2a + b + c}$$

where a is the number of species that are both observed and predicted as being present, b is the number of species observed as present but predicted as absent, and c is the number of species observed as absent but predicted as present. Finally, we also compared species richness between observed and predicted values following Pellissier et al. (2012a,b) to assess whether richness predictions are more accurate, once trophic interactions are included in the models. Species distribution models were run in R using the libraries referred in Thuiller et al. (2009). We compared the species distribution and communities composition predictive accuracies between abiotic and biotic models using paired Wilcoxon signed-rank tests.

Species data

To test our framework, we used a dataset collected in the Western Swiss Alps spanning an elevation gradient from 1000 m to 3210 m a.s.l. (Fig. 1, Pellissier et al. 2013). Site selection, which only included open vegetation grasslands, was conducted following a balanced stratified random sampling design based on elevation, slope, and aspect in open grasslands (Hirzel and Guisan 2002). Sampling in forests was avoided because they constitute suboptimal habitats for the majority of Papilionoidea species in the study area. Between May and September 2009 and 2010, 192 plots of 50 × 50 m were sampled across the whole study area (700 km^2, Fig. S1). All butterfly species belonging to the Papilionoidea super-family (sensu Heikkilä et al. 2012) were monitored. We collected adult specimens instead of caterpillars because they are more conspicuous, more reliable to survey and easier to identify. For modeling purposes, we only considered species with at least 10 occurrences and excluded very mobile species, as these are not necessarily related to environments or resources where they occur and may constitute noise in our analyses (77 species in total). Because occurrences of the butterfly species from our dataset had a low degree of spatial autocorrelation (mean Moran's I = 0.14, range −0.04 to 0.55), we can be confident about the statistical independence of the inventoried sites (see also Randin et al. 2009). Plant species as potential hosts for butterfly larvae were exhaustively inventoried in 4 m^2 subplots at the center of each butterfly plot. We ensured that the vegetation was representative of the entire 2500 m^2 area from which the butterflies were monitored. For each site, we extracted climatic predictors known to influence plant and insect establishment (Hodkinson 2005). All predictors

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

were interpolated from meteorological stations using a digital elevation model (DEM) at 100 m resolution. Values for degree-days (above 0°C) and moisture index (computed as the difference between precipitation and evapotranspiration) were interpolated following Zimmermann and Kienast (1999). Solar radiation values were calculated using the tool implemented in ArcGIS 10. For each plot location, we extracted the values of environmental variables from the associated environmental layers.

To model the trophic interactions between each butterfly and host plant using the food web model, we collected feeding preferences for the butterfly species recorded in the plots from the literature (Pellissier et al. 2013). In addition, leaf traits were gathered for all the 215 most frequent and abundant vascular plant species occurring in the study area (from a total of ~700 recorded species) and span major portions of the angiosperm phylogeny (Pellissier et al. 2013). Leaf nitrogen (N) was measured from grounded dried leaves and was analyzed by combustion with an elemental analyser. Canopy height was measured from 10 individuals spanning the species ecological conditions. In addition, the phylogenetic relationships of all butterfly species found in the study area, and the 215 plant species were inferred using DNA sequences following methods described in Pellissier et al. (2013) (see also Supplementary materials, Figures S1 and S2).

Results

Plant traits and phylogeny predict butterfly and host plant interactions

We fitted the food web model on 89 plant species that are known to be eaten by at least one butterfly species (Fig. 2). Based on this calibration dataset, the food web model was able to predict the trophic associations between butterfly species and their respective host plants both internally (AUC fit = 0.96) and externally (AUC cross-validation = 0.84). We found a significant effect of plant nitrogen content in the susceptibility of being a

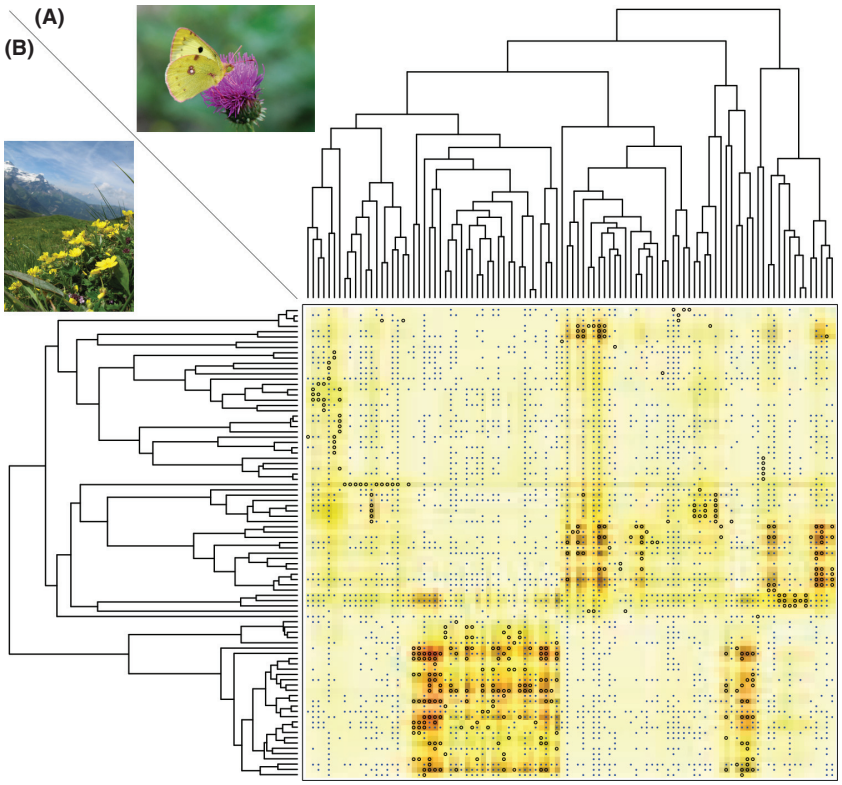


Figure 2. Trophic interactions matrix between butterfly species (A) and plant species (B). Each column and row represents a butterfly and a plant species, respectively. A black dot at an intersection represents a trophic interaction between the two corresponding species, while a blue dot indicates the absence of co-occurrence. Plants and butterfly phylogenies are presented on the left and bottom of the trophic interaction matrix. Phylogenetically closely related species tend to interact with the same subset of species. Our statistical food web model uses the trophic information on co-occurring species to infer the probability of a trophic link between all pair of species. The magnitudes of these probabilities are represented by the background color, increasing from pale yellow to red.

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

trophic resource for butterfly larvae (slope = 0.145, $P = 0.002$). In addition, we found that the two latent traits were significantly correlated with the plant and butterfly phylogenies. Based on the phylogenetic regression using Pagel's-λ correlation structure, we found for plants latent traits f_j^1: Pagel's-$\lambda = 0.8217$ P-value < 0.001, plants latent traits f_j^2: Pagel's$-\lambda = 0.6497$ P-value < 0.001, and for butterfly latent traits v_j^1: Pagel's$-\lambda = 0.7257$ P-value < 0.001, butterfly latent traits v_j^2: Pagel's$-\lambda = 0.9230$ P-value < 0.001. The ability of the food web model to predict butterfly larvae association along with host plants differed among the main butterfly families (Kruskal–Wallis test, KW = 17.06, $P = 0.004$). The highest predictive performance was achieved for the families Nymphalidae and Pieridae, followed by Hesperiidae, while the weakest performance was achieved for Lycaenidae, Papapilionidae, and the Riodinidae species (*Haemearis lucina*). (Fig. S3).

Species distribution

Accounting for modeled trophic links between each butterfly and host plant in models of species distributions significantly improved the predictive power of the models (measured using AUC) compared with a model with only abiotic components (paired Wilcoxon test, GLM: $V = 790$, $P = 0.01$, GAM: $V = 858$, $P = 0.03$, GBM: $V = 843$, $P = 0.03$, RF: $V = 758$, $P = 0.008$, Fig. S4). Model accuracy for some species was largely improved by accounting for the potential trophic links with host plants (e.g., *Parnassius appolo* +21%, *Pontia callidice* +15%, *Papilio machaon* +10%). In contrast, the predictive power of SDMs for some species declined by incorporating the biotic component (e.g., *Pyrgus andromedae* −9%, *Polyommatus eros* −7%, *Plebejus argus* −4%).

We found no general improvement in predictive power when species interactions included in the models were directly based on the literature (paired Wilcoxon test, GLM: $V = 1150$, $P = 0.55$, GAM: $V = 1167$, $P = 0.57$, GBM: $V = 1289$, $P = 0.82$, RF: $V = 1129$, $P = 0.47$) or when the food web was only interpolated to the 89 plant species used for calibration and those used in the SDMs (paired Wilcoxon test, GLM: $V = 1050$, $P = 0.48$, GAM: $V = 1172$, $P = 0.59$, GBM: $V = 1340$, $P = 0.84$, RF: $V = 1127$, $P = 0.47$).

Only models for species from the Nymphalidae (paired Wilcoxon test, GLM: $V = 161$, $P = 0.04$, GAM: $V = 166$, $P = 0.04$, GBM: $V = 149$, $P = 0.03$, RF: $V = 139$, $P = 0.01$) and Pieridae (paired Wilcoxon test, GAM: $V = 2$, $P = 0.04$, GBM: $V = 2$, $P = 0.04$, RF: $V = 2$, $P = 0.01$) families were improved using the food web model compared with the model using the raw literature data. Hesperiidae (paired Wilcoxon test, GLM: $V = 44$, $P = 0.85$, GAM: $V = 46$, $P = 0.87$, GBM: $V = 31$, $P = 0.45$, RF: $V = 25$, $P = 0.25$) and Lycaenidae (paired Wilcoxon test, GLM: $V = 72$, $P = 0.56$, GAM: $V = 69$, $P = 0.53$, GBM: $V = 70$, $P = 0.55$, RF: $V = 69$, $P = 0.53$) showed no significant difference in predictive power.

Species richness and communities composition

Community predictions were also improved with a relative increase of 11% in the median composition similarity (paired Wilcoxon test on Sørenson index: GLM: $V = 6950$, $P = 0.008$, GAM: $V = 6867$, $P = 0.006$, GBM: $V = 7230$, $P = 0.02$, RF: $V = 7157$, $P = 0.01$, Fig. 3). Across the communities, we found a Sørenson index of 0.46 between observed and predicted species when trophic interactions were accounted for in the models, while 0.41 when interactions were not accounted for. In addition, we found a 7% decrease in the median of absolute values of modeled richness residuals (paired Wilcoxon test on richness residuals: GLM: $V = 10350$, $P = 0.007$, GAM: $V = 10420$, $P = 0.004$, GBM: $V = 9967$, $P = 0.04$; RF: $V = 9779$, $P = 0.05$).

Discussion

In nature, species' interactions and interdependencies are dynamic under the processes of global change, and ecological constraints within food web are among those biotic interactions demonstrated to affect species' spatial distributions (Van der Putten et al. 2010; Wisz et al. 2013). Although there have been recent advances in predicting the distributions of species that potentially contribute to community function (Thuiller et al. 2013), it is still a challenge to predict the identity of the species that will interact with each other across trophic levels at a given place or time (Mokany and Ferrier 2011; Mokany et al. 2011). This study presents a novel approach that addresses this challenge by combining the strengths of food web and species distribution modeling (SDMs) to produce more ecologically meaningful predictions of species distribution and assemblage composition. As such, it can be applied to any ecosystem where occurrence data and environmental predictors are available for generating species distribution predictions, and where a general food web model can be constructed based on quantifiable species' traits and phylogenetic information complementing a partial food web. Here, we illustrate the utility of this combined approach, examining 125 butterfly species and 215 potentially interacting plant species in the western Swiss Alps. By combining a food web model based on plant resource characteristics, or phylogenetic information, and SDMs, we show that accounting for trophic interactions between a butterfly and associated food plant host improves the quality of community forecasts.

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

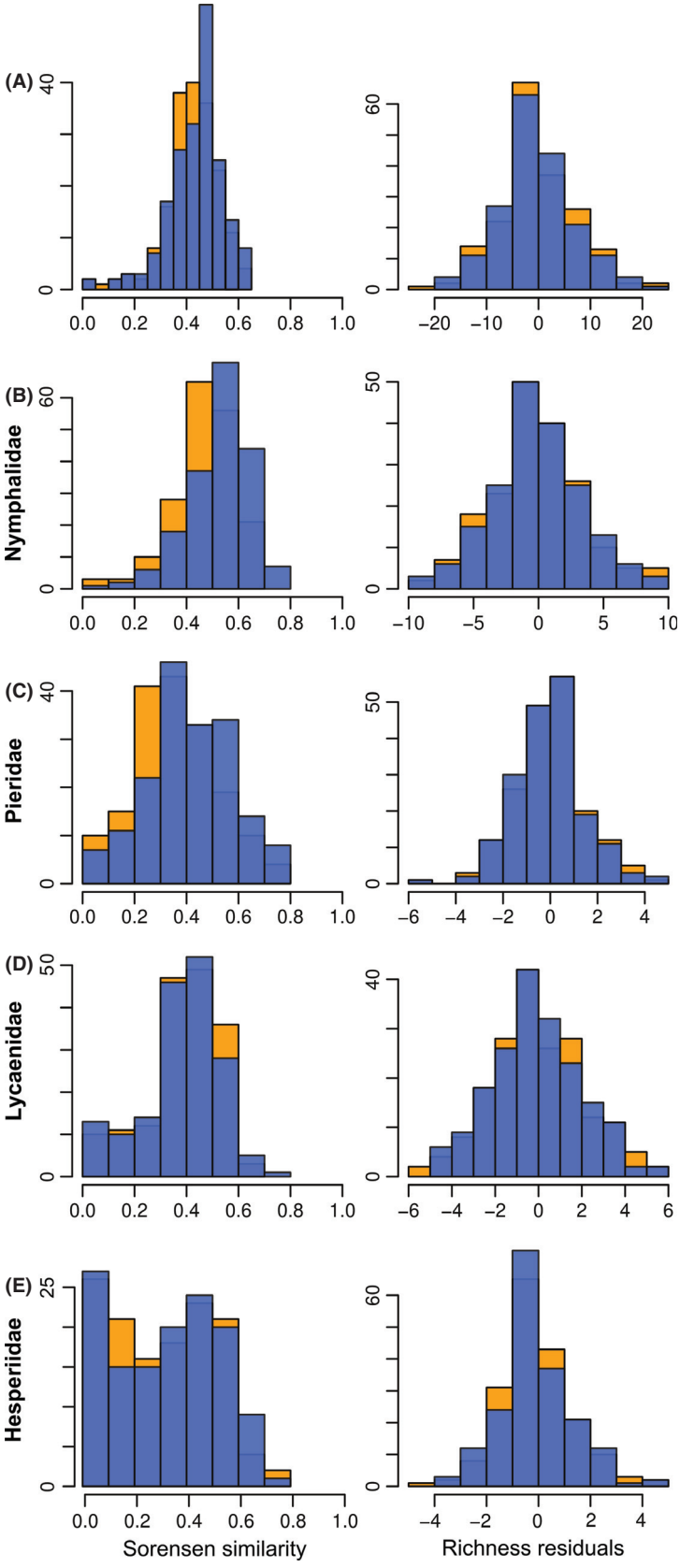


Figure 3. Histograms of Sørenson similarity with observations and richness residuals of the stacked species distribution models. Shown are the results for the random forest modeling technique with abiotic predictors (in orange below) and considering in addition the trophic links with host plant (in blue), for all species together (A) and the four main families, Nymphalidae (B), Pieridae (C), Lycaenidae (D), and Hesperiidae (E).

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Extrapolation of food webs

Given the prevalence of functionally constrained food web networks in several ecosystems across the globe (e.g., Trussell et al. 2002; Petchey et al. 2005), our approach shows great promise for providing more ecologically informed and realistic spatial distributions for species including those that have incomplete distributional information or when it is unknown how they interact with other species. A limited number of studies have attempted to integrate trophic interactions in biogeographical modeling, including Gravel et al. (2011), who extended the island biogeography theory of extinction-colonization processes to account for trophic interactions. Similarly, Jabot and Bascompte (2012) integrated a spatial meta-community structure to food web network approaches by studying a bitrophic plant–pollinator interaction. This is the first attempt to combine a food web model with a niche-based species distribution model. The ability to extrapolate food web models is useful in the context of climate change, to predict potential interactions of species that do not currently co-occur or interact, but may do so in the future. For example, a recent study documented a temperature-dependent shift of a butterfly species to a novel host plant impacting species range (Pateman et al. 2012; but see Pelini et al. 2010). Our food web model SDM approach could be used to predict shifts such as this one.

A unique feature of our approach is the extrapolation of the food web model to predict missing interactions in networks and inform species distribution models. Our food web modeling approach yields a continuous index of interaction probability, based on the quantitative support for the presence or absence of interactions between species derived from information on species' traits and phylogeny. Information directly derived from the literature is less reliable, as there is typically underrepresentation of information on the interactions among species. Here, based on known interactions involving 125 butterfly and 89 host plants, we extrapolate a food web model to 215 host plants (~35% of the regional species pool) for which relevant functional traits and phylogenetic positions were available. The extrapolation was based on the strong associations between nitrogen content and the probability of a trophic interaction between species and the correlation between the latent traits and the species phylogeny. This supports the view that herbivore diet shows strong phylogenetic constraints (Pellissier et al. 2013) and is driven by plant resource content (Pellissier et al. 2012a,b). Previous studies also showed that food webs may be phylogenetically structured (Ives and Godfray 2006). We found no general improvement of species distribution models when we considered only interpolated values on the 89 host plants (~15% of the plant species pool) or using the raw links in the literature. This may have arisen because too many potential host plants were missing resulting in biased inferences of butterfly presence and absences at sites. We speculate that more accurate assemblage predictions may have been realized if we had information for the full regional species pool (~700 plant species), but for which trait information is currently unavailable.

Our approach is particularly suited for group of species that are neither specialist nor extreme generalists, but still interact with a large number of species, where current knowledge on species interactions is probably incomplete. From our dataset, the models built for the butterfly families Nymphalidae and Pieridae showed the most consistent improvement in model predictive performance (Fig. 3). This may perhaps be due to the Nymphalidae family being composed of more polyphagous species (larger number of potential host plants) and the display of lower phylogenetic variability in host plant association (e.g., Satyrinae subfamily feeding on the grass family, Poaceae and Cyperaceae) (Pellissier et al. 2013). Consequently, the food web model performed relatively well across this family and could explain several missing links in the interaction matrix required to predict communities using SDMs. Models for species of the Lycaenidae and Hesperiidae were not improved when information from the food web model was included in SDMs. Those families display higher degree of shifts in host-plant associations with closely related lineages specialized on distinct plant clades (Megens et al. 2005; Vila et al. 2011). As a consequence, trophic interactions are more difficult to model and including uncertain interaction probabilities results in poorest predictions. The overall trend of improvement at the species and community levels is likely to arise because the more generalist Nymphalidae family is dominant in our butterfly dataset, while specialist species are less frequent (Fig. 3).

One caveat of the approach proposed in this study is that it is bound to a statistical framework and cannot demonstrate causality. Even if species ecological knowledge supports the relevance of including trophic interactions in model of species distribution, adding further abiotic predictors may also have led to model improvement. Hence, we are unable to demonstrate that the inclusion of trophic interactions actually improves the models in a causal and mechanistic way. Nevertheless, despite the fact that our results are correlative, they suggest new research areas related to the inclusion of trophic interactions in studies of species distributions.

Future perspectives

In addition to improving model predictions, the next-generation models combining SDMs and food web

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

models should allow novel and complex ecological questions to be addressed (Thuiller et al. 2013). For instance, how might recurrent perturbations in the food web (e.g., the extinction of a predator) affect the distribution of species in a landscape? Nevertheless, the approach presented here is a timely contribution in using food web models to infer the geographical distribution of species in the landscape, as recently proposed by Massol et al. (2011) and Jabot and Bascompte (2012). In addition, we considered a food web that only accounts for bottom-up interactions that potentially influence species' distributions (i.e., the host plant resources that simultaneously influence butterfly distribution). Our approach can be extended to both bottom-up and top-down interactions within food webs. However, this would require a statistical model that simultaneously fits both food web structure and the presences/absences of interactions along environmental gradients, while considering the turnover in these interactions in time and space.

A key improvement of our approach is that with widespread species, trophic interactions usually shift across a species range, and as yet cannot be accounted for in SDMs that do not incorporate food web models. With the plant–butterfly system, we have illustrated, as plant leaf palatability (SLA, LDMC) shift along environmental gradients, this may in turn influence the interactions with other species along those gradients (Linder et al. 2012). In such a landscape, food web models can potentially account for shifts in prevailing environmental conditions, an area to be considered in future modeling efforts. As with many interdisciplinary modeling efforts, the refinement of our models is constrained by the availability of data to fit and evaluate them. For instance, the dataset used to test these methods are hindered by a spatial scale mismatch between plant and butterfly taxa. While the butterfly communities were sampled in plots of 50 $\times$ 50 m, only 2 $\times$ 2 m grasslands were sampled for plant species presences. It is probable that if we had collected host plant data at a scale more representative of the butterfly plots, better results may have been achieved. In addition, because of the difficult detection of larvae in the field, we used presence and absence of adult butterfly specimens, while we modeled the food web of larvae. It is likely that the difference between abiotic models and those considering trophic interactions would have been more pronounced if records of larvae were used.

Conclusion

Here, we facilitate the integration of ecological rules based on trophic interactions to constrain prediction of species assemblages in space and time. The models built using our combined approach are ecologically more sounds as they consider biotic interaction that are known a priori to influence the life history of butterflies. This framework extends the boundaries of SDMs by accounting for known and unknown species ecological interactions and produces more realistic models of species' assemblages. It also offers an innovative way to improve our understanding of community assembly by accounting for the multiple interactions among different species based on extensive datasets. Moreover, it informs where species can be found as a function of their relation to the abiotic environment, as well as how these species will interact with novel components of the biotic environment. Our approach draws upon the strengths of spatial ecology and trophic ecology, and should encourage the further integration of these disciplines to advance efforts to predict species assemblages.

Acknowledgments

We would like to thank those who helped with the field work, and Yannick Chittaro for providing pictures. This study was supported by the NSF Grant Nr. 31003A-125145 (BIOASSEMBLE project) to AG. LP was supported by The Danish Council for Independent Research Grant n° 12-126430. RPR was supported by the FP7-REGPOT 2010-1 EcoGenes Project (Grant No. 264125). This project also received support from the Vital-IT facilities of the Swiss Institute of Bioinformatics at the University of Lausanne.

Conflict of Interest

None declared.

References

Araújo, M. B. and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. Glob. Ecol. Biogeogr. 16:743–753.

Best, R. J., N. C. Caulk, and J. J. Stachowicz. 2013. Trait vs. phylogenetic diversity as predictors of competition and community composition in herbivorous marine amphipods. Ecol. Lett. 16:72–80.

Boulangeat, I., D. Gravel, and W. Thuiller. 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecol. Lett. 15:584–593.

Clauset, A., C. Moore, and M. Newman. 2008. Hierarchical structure and the prediction of missing links in networks. Nature 101:98–101.

Dubuis, A., J. Pottier, V. Rion, L. Pellissier, J.-P. Theurillat, and A. Guisan. 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. Divers. Distrib. 17:1122–1131.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

 © 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Freckleton, R., P. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160:712–726.

Giannini, T. C., D. S. Chapman, A. M. Saraiva, I. Alves-dos-Santos, and J. C. Biesmeijer. 2013. Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants. Ecography 36:649–656.

Gravel, D., E. Canard, F. Guichard, and N. Mouquet. 2011. Persistence increases with diversity and connectance in trophic metacommunities. PLoS ONE 6:e19374.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8:993–1009.

Heikkilä, M., L. Kaila, M. Mutanen, C. Peña, and N. Wahlberg. 2012. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. Proc. R. Soc. B Biol. Sci. 279:1093–1099.

Hirzel, A., and A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecol. Mod. 157:331–341.

Hodkinson, I. D. 2005. Terrestrial insects along elevation gradients: species and community responses to altitude. Biol. Rev. 80:489–513.

Ibanez, S., S. Lavorel, S. Puijalon, and M. Moretti. 2013. Herbivory mediated by coupling between biomechanical traits of plants and grasshoppers. Funct. Ecol. 27:479–489.

Ives, A. R., and H. C. J. Godfray. 2006. Phylogenetic analysis of trophic associations. Am. Nat. 168:E1–E14.

Jabot, F., and J. Bascompte. 2012. Bitrophic interactions shape biodiversity in space. Proc. Natl Acad. Sci. USA 109:4521–4526.

Kissling, W. D., C. F. Dormann, J. Groeneveld, T. Hickler, I. Kühn, G. J. McInerny, et al. 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. J. Biogeogr. 39:2163–2178.

Laughlin, D. C., C. Joshi, P. M. van Bodegom, Z. A. Bastow, and P. Z. Fulé. 2012. A predictive model of community assembly that incorporates intraspecific trait variation. Ecol. Lett. 15:1291–1299.

Le Roux, P., J. Lenoir, L. Pellissier, M. Wisz, and M. Luoto. 2013. Horizontal, but not vertical, biotic interactions affect fine-scale plant distribution patterns in a low energy system. Ecology 94:671–682.

Linder, H. P., O. Bykova, J. Dyke, R. S. Etienne, T. Hickler, I. Kühn, et al. 2012. Biotic modifiers, environmental modulation and species distribution models. J. Biogeogr. 39:2179–2190.

Massol, F., D. Gravel, N. Mouquet, M. W. Cadotte, T. Fukami, M. Leibold, et al. 2011. Linking community and ecosystem dynamics through spatial ecology. Ecol. Lett. 14:313–323.

Megens, H.-J., R. De Jong, and K. Fiedler. 2005. Phylogenetic patterns in larval host plant and ant association of Indo-Australian Arhopalini butterflies Lycaenidae: Theclinae. Biol. J. Linn. Soc. 84:225–241.

Mokany, K., and S. Ferrier. 2011. Predicting impacts of climate change on biodiversity: a role for semi-mechanistic community-level modelling. Divers. Distrib. 17:374–380.

Mokany, K., T. D. Harwood, J. M. Overton, G. M. Barker, and S. Ferrier. 2011. Combining α- and β-diversity models to fill gaps in our knowledge of biodiversity. Ecol. Lett. 14:1043–1051.

Mouquet, N., V. Devictor, C. N. Meynard, F. Munoz, L.-F. Bersier, J. Chave, et al. 2012. Ecophylogenetics: advances and perspectives. Biol. Rev. 87:769–785.

Naisbit, R. E., R. P. Rohr, A. G. Rossberg, P. Kehrli, and L.-F. Bersier. 2012. Phylogeny versus body size as determinants of food web structure. Proc. R. Soc. B Biol. Sci. 279:3291–3297.

Nogués-Bravo, D. and C. Rahbek. 2011. Communities under climate change. Science 334:1070–1071.

Ovaskainen, O., J. Hottola, and J. Siitonen. 2010. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology 91:2514–2521.

Pateman, R. M., J. K. Hill, D. B. Roy, R. Fox, and C. D. Thomas. 2012. Temperature-dependent alterations in host use drive rapid range expansion in a butterfly. Science 336:1028–1030.

Pelini, S. L., J. A. Keppel, A. E. Kelley, and J. J. Hellmann. 2010. Adaptation to host plants may prevent rapid insect responses to climate change. Glob. Ch. Biol. 16:2923–2929.

Pellissier, L., K. Anne Bråthen, J. Pottier, C. F. Randin, P. Vittoz, A. Dubuis, et al. 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. Ecography 33:1004–1014.

Pellissier, L., J.-N. Pradervand, J. Pottier, A. Dubuis, L. Maiorano, and A. Guisan. 2012a. Climate-based empirical models show biased predictions of butterfly communities along environmental gradients. Ecography 35:684–692.

Pellissier, L., S. Rasmann, G. Litsios, K. Fiedler, A. Dubuis, J. Pottier, et al. 2012b. High host-plant nitrogen content: a prerequisite for the evolution of ant–caterpillar mutualism? J. Evol. Biol. 25:1658–1666.

Pellissier, L., C. Ndiribe, A. Dubuis, J.-N. Pradervand, N. Salamin, A. Guisan, et al. 2013. Turnover of plant lineages shapes herbivore phylogenetic beta diversity along ecological gradients. Ecol. Lett. 16:600–608.

Petchey, O. L., A. P. Beckerman, J. O. Riede, and P. H. Warren. 2005. Size, foraging, and food web structure. Proc. Natl Acad. Sci. USA 105:4191–4196.

Randin, C. F., H. Jaccard, P. Vittoz, N. G. Yoccoz, and A. Guisan. 2009. Land use improves spatial predictions of mountain plant abundance but not presence-absence. J. Veg. Sci. 20:996–1008.

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Rohr, R. P., H. Scherer, P. Kehrli, C. Mazza, and L.-F. Bersier. 2010. Modeling food webs: exploring unexplained structure using latent traits. Am. Nat. 176:170–177.

Schweiger, O., R. K. Heikkinen, A. Harpke, T. Hickler, S. Klotz, O. Kudrna, et al. 2012. Increasing range mismatching of interacting species under global change is related to their ecological characteristics. Glob. Ecol. Biogeogr. 21:88–99.

Thuiller, W., B. Lafourcade, R. Engler, and M. B. Araújo. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32:369–373.

Thuiller, W., T. Münkemüller, S. Lavergne, D. Mouillot, N. Mouquet, K. Schiffers, et al. 2013. A road map for integrating eco-evolutionary processes into biodiversity models. Ecol. Lett. 16:94–105.

Trussell, G. C., P. J. Ewanchuk, and M. D. Bertness. 2002. Field evidence of trait-mediated indirect interactions in a rocky intertidal food web. Ecol. Lett. 5:241–245.

Van der Putten, W. H., M. Macel, and M. E. Visser. 2010. Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. Phil. Trans. R. Soc. B. Biol. Sci. 365:2025–2034.

Vila, R., C. D. Bell, R. Macniven, B. Goldman-Huertas, R. H. Ree, C. R. Marshall, et al. 2011. Phylogeny and palaeoecology of Polyommatus blue butterflies show Beringia was a climate-regulated gateway to the New World. Proc. R. Soc. B: Biol. Sci. 278:2737–2744.

Whitfeld, T., V. Novotny, S. Miller, and J. Hrcek. 2012. Predicting tropical insect herbivore abundance from host plant traits and phylogeny. Ecology 93:211–223.

Wisz, M. S., J. Pottier, W. D. Kissling, L. Pellissier, J. Lenoir, C. F. Damgaard, et al. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol. Rev. 88:15–30.

Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, et al. 2004. The worldwide leaf economics spectrum. Nature 428:821–827.

Zimmermann, N., and F. Kienast. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. J. Veg. Sci. 10:469–482.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Butterfly phylogeny.
Figure S2. Molecular dated phylogenetic tree of 231 angiosperm species from the Western Swiss Alps.
Figure S3. Trophic interactions matrix between butterfly species and plant species extrapolated on the full dataset comprising 125 butterfly and 215 plant species.
Figure S4. Boxplot of the AUC of the modeled links of the food web model for each of the four main butterfly families, Nymphalidae, Pieridae, Lycaenidae, Hesperiidae.
Figure S5. Histograms of the difference between the predictive power of the models (AUC) without and with the trophic interaction as a predictor.
Table S1. Species belonging to Papilionoidea that were found in the study area.

© 2013 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Supplementary materials

Figure S1 – Butterfly phylogeny. Shown is the consensus tree using a Bayesian ultrametric approach of the species of butterfly species found in the study area, belonging to the six families of butterflies, viz. Nymphalidae, Lycaenidae, Pieridae, Papilionidae, Riodinidae and Hesperiidae. Phylogenetic relationships were inferred using DNA sequences obtained from GenBank and including two nuclear markers (EF1-alpha, Wgl) and four mitochondrial markers (16s, COI, NDH1, NDH5).

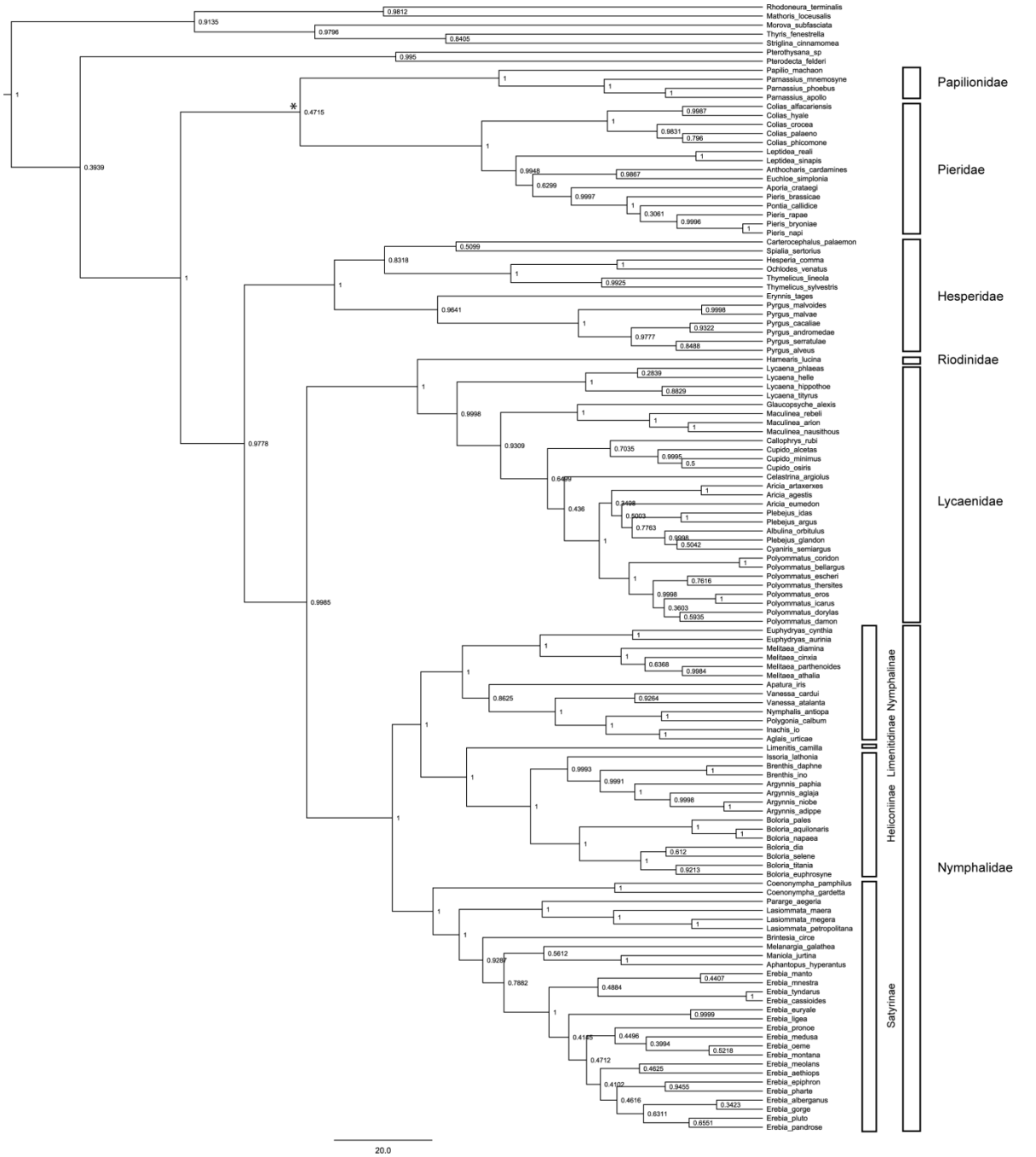
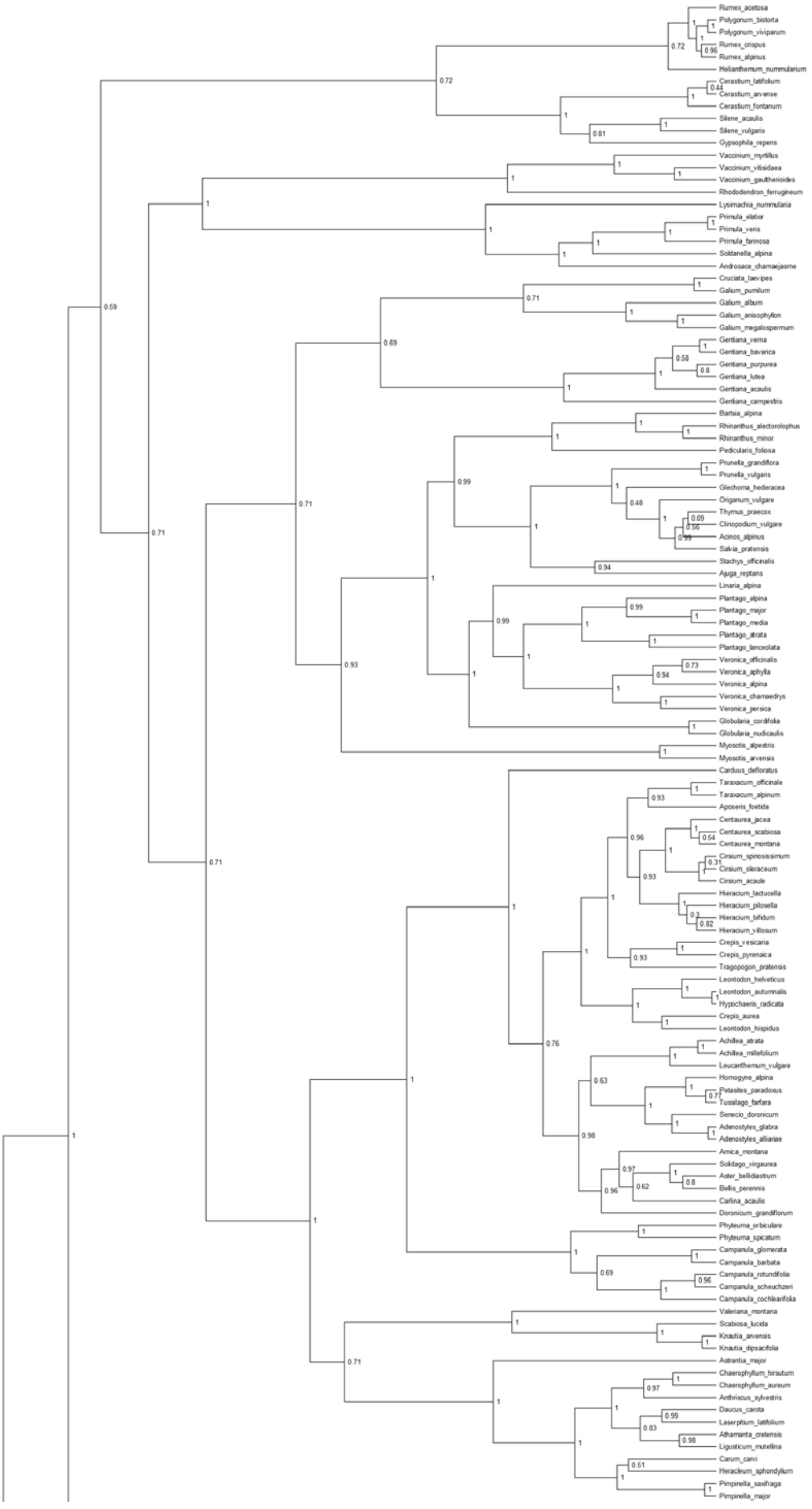


Figure S2 – Molecular dated phylogenetic tree of 231 angiosperm species from the Western Swiss Alps. Posterior means of divergence are in relative time units with all nodes scaled from a common basal divergence of 130 Mya. The nodes were well resolved and supported, with 77% nodes receiving posterior probabilities greater than 90%. Two outgroup species *Abies alba* Mill. and *Picea abies* (L.) H. Karst. were used to root the tree. The resulting phylogenetic trees were checked against the Angiosperm Phylogeny Group tree for accepted relationships among plant orders and families The phylogenetic reconstructions, including divergence time estimations, produced well-supported phylogenetic trees with nodes congruent to taxonomic groups defined by the APG III classification. A total of 56% of the nodes in the dated tree had 100% bootstrap support. A total of 77% nodes had posterior probabilities greater than 90%. In general, only a few nodes showed low support, such as the placement of *Cirsium spinosissimum* (L.) Scop. and *Cirsium oleraceum* (L.) Scop. within the Asteraceae family, and some nodes within the Cyperaceae and Lamiaceae family.


Rumex_acetosa
Polygonum_bistorta
Polygonum_viviparum
Rumex_crispus
Rumex_alpinus
Helianthemum_nummularium
Cerastium_latifolium
Cerastium_arvense
Cerastium_fontanum
Silene_acaulis
Silene_vulgaris
Gypsophila_repens
Vaccinium_myrtillus
Vaccinium_vitisidaea
Vaccinium_gaultherioides
Rhododendron_ferrugineum
Lysimachia_nummularia
Primula_elatior
Primula_veris
Primula_farinosa
Soldanella_alpina
Androsace_chamaejasme
Cruciata_laevipes
Galium_pumilum
Galium_album
Galium_anisophyllon
Galium_megalospermum
Gentiana_verna
Gentiana_bavarica
Gentiana_purpurea
Gentiana_lutea
Gentiana_acaulis
Gentiana_campestris
Bartsia_alpina
Rhinanthus_alectorolophus
Rhinanthus_minor
Pedicularis_foliosa
Prunella_grandiflora
Prunella_vulgaris
Glechoma_hederacea
Origanum_vulgare
Thymus_praecox
Clinopodium_vulgare
Acinos_alpinus
Salvia_pratensis
Stachys_officinalis
Ajuga_reptans
Linaria_alpina
Plantago_alpina
Plantago_major
Plantago_media
Plantago_atrata
Plantago_lanceolata
Veronica_officinalis
Veronica_aphylla
Veronica_alpina
Veronica_chamaedrys
Veronica_persica
Globularia_cordifolia
Globularia_nudicaulis
Myosotis_alpestris
Myosotis_arvensis
Carduus_defloratus
Taraxacum_officinale
Taraxacum_alpinum
Aposeris_foetida
Centaurea_jacea
Centaurea_scabiosa
Centaurea_montana
Cirsium_spinosissimum
Cirsium_oleraceum
Cirsium_acaule
Hieracium_lactucella
Hieracium_pilosella
Hieracium_bifidum
Hieracium_villosum
Crepis_vesicaria
Crepis_pyrenaica
Tragopogon_pratensis
Leontodon_helveticus
Leontodon_autumnalis
Hypochaeris_radicata
Crepis_aurea
Leontodon_hispidus
Achillea_atrata
Achillea_millefolium
Leucanthemum_vulgare
Homogyne_alpina
Petasites_paradoxus
Tussilago_farfara
Senecio_doronicum
Adenostyles_glabra
Adenostyles_alliariae
Arnica_montana
Solidago_virgaurea
Aster_bellidiastrum
Bellis_perennis
Carlina_acaulis
Doronicum_grandiflorum
Phyteuma_orbiculare
Phyteuma_spicatum
Campanula_glomerata
Campanula_barbata
Campanula_rotundifolia
Campanula_scheuchzeri
Campanula_cochlearifolia
Valeriana_montana
Scabiosa_lucida
Knautia_arvensis
Knautia_dipsacifolia
Astrantia_major
Chaerophyllum_hirsutum
Chaerophyllum_aureum
Anthriscus_sylvestris
Daucus_carota
Laserpitium_latifolium
Athamanta_cretensis
Ligusticum_mutellina
Carum_carvi
Heracleum_sphondylium
Pimpinella_saxifraga
Pimpinella_major

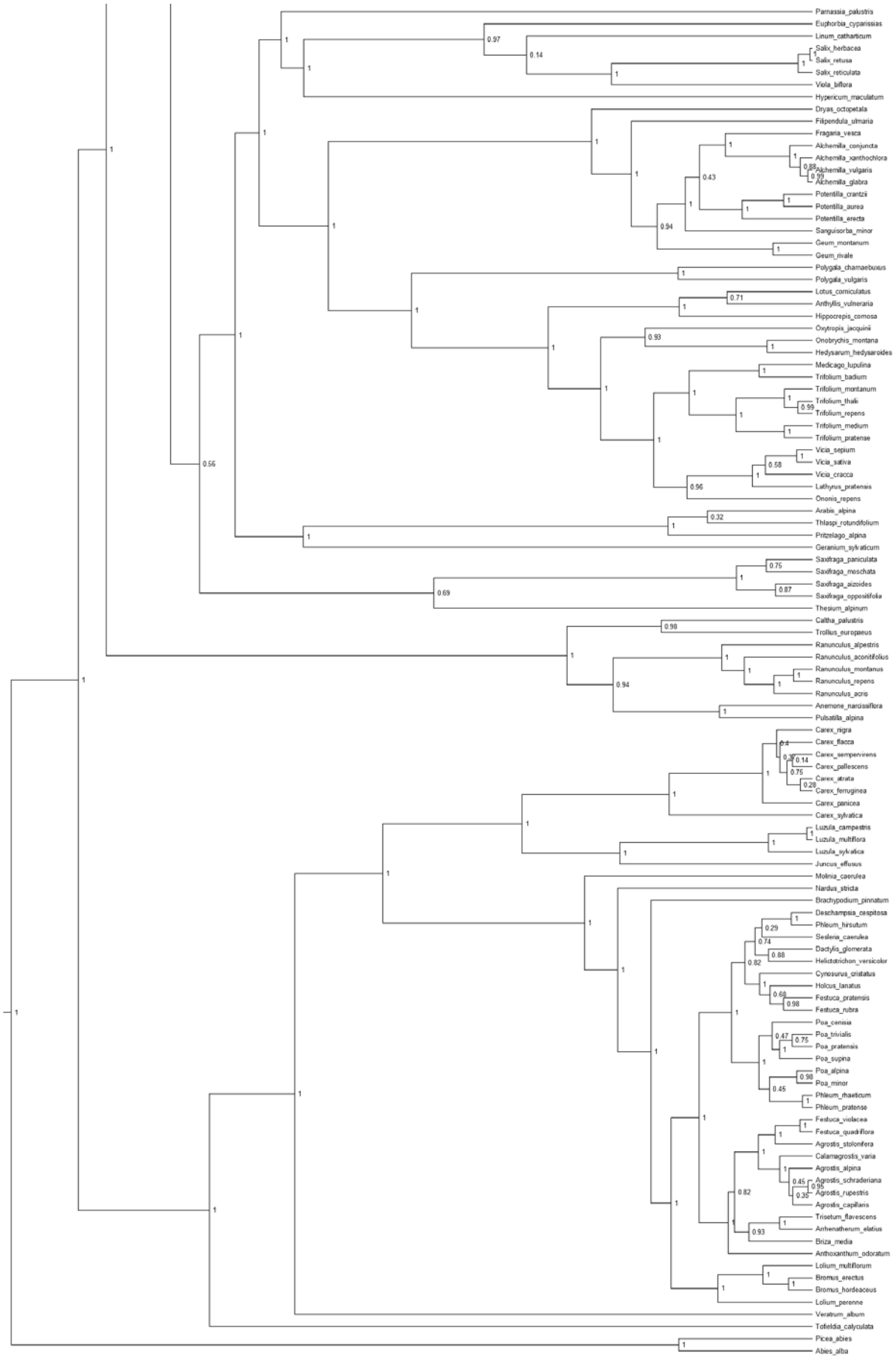

Parnassia_palustris
Euphorbia_cyparissias
Linum_catharticum
Salix_herbacea
Salix_retusa
Salix_reticulata
Viola_biflora
Hypericum_maculatum
Dryas_octopetala
Filipendula_ulmaria
Fragaria_vesca
Alchemilla_conjuncta
Alchemilla_xanthochlora
Alchemilla_vulgaris
Alchemilla_glabra
Potentilla_crantzii
Potentilla_aurea
Potentilla_erecta
Sanguisorba_minor
Geum_montanum
Geum_rivale
Polygala_chamaebuxus
Polygala_vulgaris
Lotus_corniculatus
Anthyllis_vulneraria
Hippocrepis_comosa
Oxytropis_jacquinii
Onobrychis_montana
Hedysarum_hedysaroides
Medicago_lupulina
Trifolium_badium
Trifolium_montanum
Trifolium_thalii
Trifolium_repens
Trifolium_medium
Trifolium_pratense
Vicia_sepium
Vicia_sativa
Vicia_cracca
Lathyrus_pratensis
Ononis_repens
Arabis_alpina
Thlaspi_rotundifolium
Pritzelago_alpina
Geranium_sylvaticum
Saxifraga_paniculata
Saxifraga_moschata
Saxifraga_aizoides
Saxifraga_oppositifolia
Thesium_alpinum
Caltha_palustris
Trollius_europaeus
Ranunculus_alpestris
Ranunculus_aconitifolius
Ranunculus_montanus
Ranunculus_repens
Ranunculus_acris
Anemone_narcissiflora
Pulsatilla_alpina
Carex_nigra
Carex_flacca
Carex_sempervirens
Carex_pallescens
Carex_atrata
Carex_ferruginea
Carex_panicea
Carex_sylvatica
Luzula_campestris
Luzula_multiflora
Luzula_sylvatica
Juncus_effusus
Molinia_caerulea
Nardus_stricta
Brachypodium_pinnatum
Deschampsia_cespitosa
Phleum_hirsutum
Sesleria_caerulea
Dactylis_glomerata
Helictotrichon_versicolor
Cynosurus_cristatus
Holcus_lanatus
Festuca_pratensis
Festuca_rubra
Poa_cenisia
Poa_trivialis
Poa_pratensis
Poa_supina
Poa_alpina
Poa_minor
Phleum_rhaeticum
Phleum_pratense
Festuca_violacea
Festuca_quadriflora
Agrostis_stolonifera
Calamagrostis_varia
Agrostis_alpina
Agrostis_schraderiana
Agrostis_rupestris
Agrostis_capillaris
Trisetum_flavescens
Arrhenatherum_elatius
Briza_media
Anthoxanthum_odoratum
Lolium_multiflorum
Bromus_erectus
Bromus_hordeaceus
Lolium_perenne
Veratrum_album
Tofieldia_calyculata
Picea_abies
Abies_alba

Figure S3 - Trophic interactions matrix between butterfly species and plant species extrapolated on the full dataset comprising 125 butterfly and 215 plant species. Each column and row represents a butterfly and a plant species respectively. A black dot at an intersection represents a trophic interaction between the two corresponding species, while a blue dot indicates the absence of co-occurrence. Plants and butterfly phylogenies are presented on the left and bottom of the trophic interaction matrix.

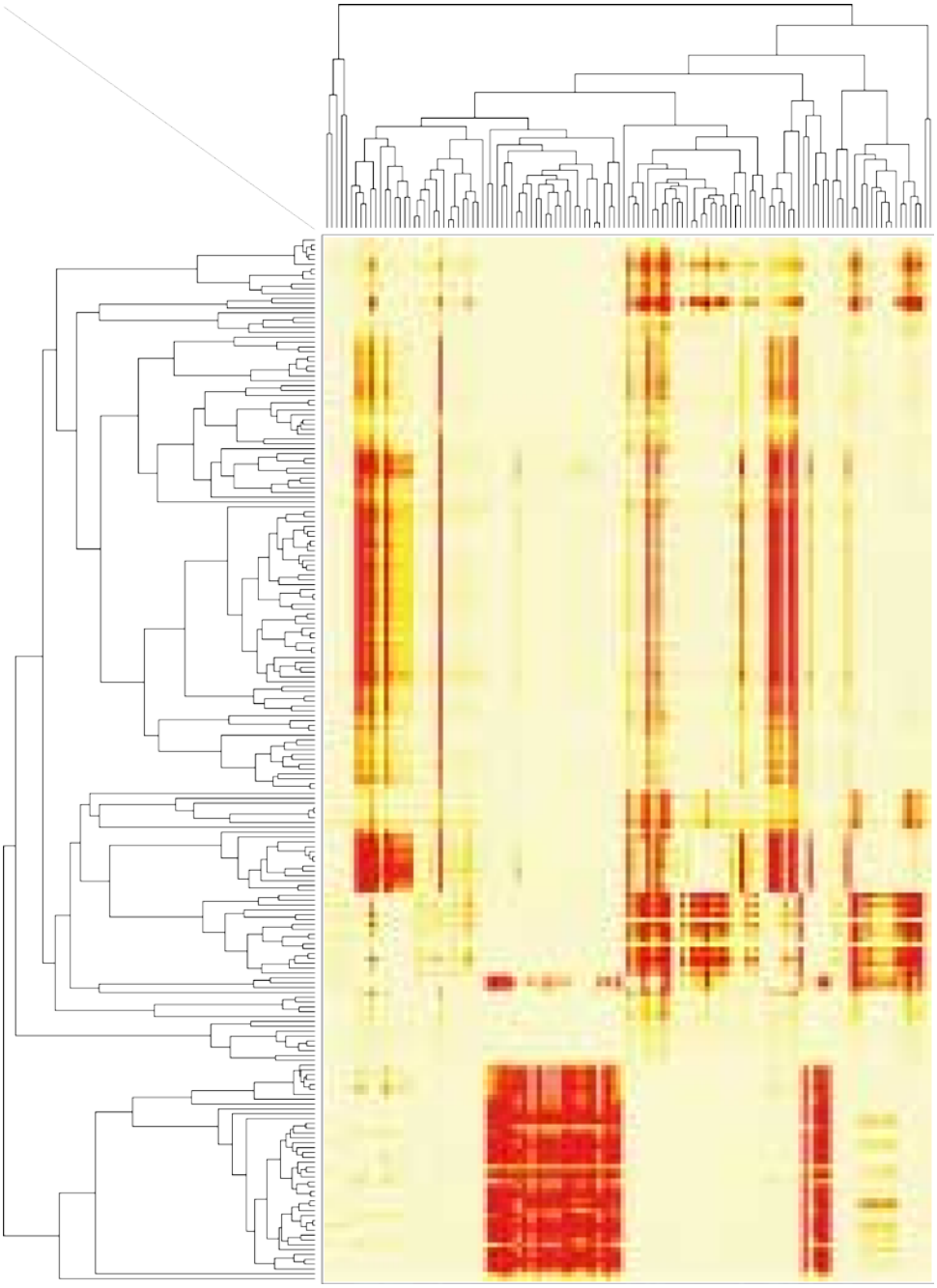


Figure S4 - Boxplot of the AUC of the modelled links of the food-web model for each of the four main butterfly families, Nymphalidae, Pieridae, Lycaenidae, Hesperiidae. The food web model was more limited in predicting the interactions links of Lycaenidae, while it scored the highest for Nymphalidae and Pieridae. Nevertheless, a few more specialized Nymphalidae species had a lower score especially in the *Boloria*, *Euphydryas* and *Melitea* genera.

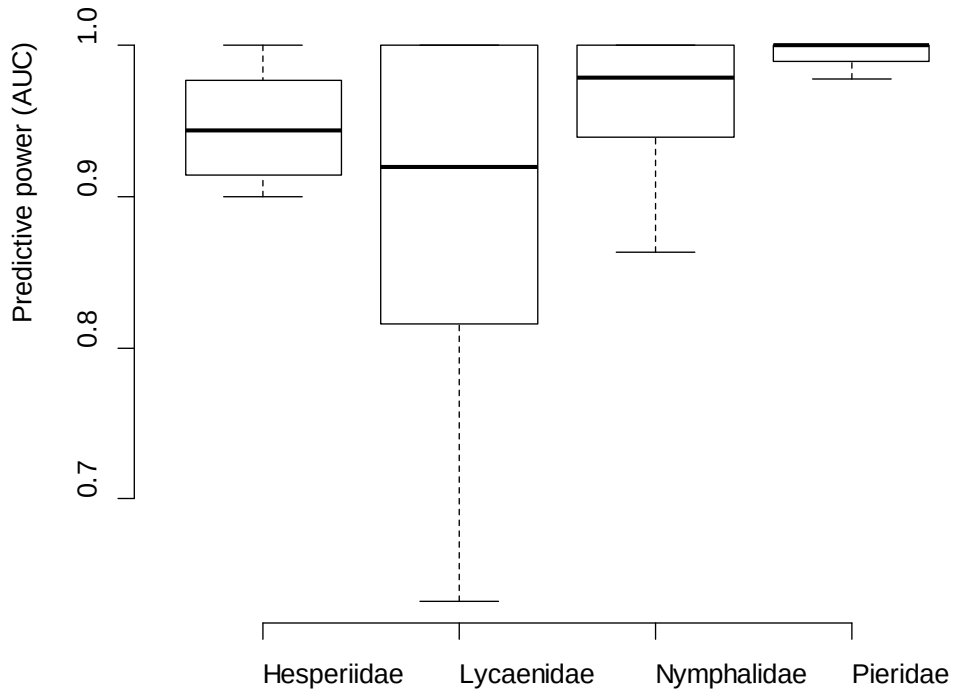
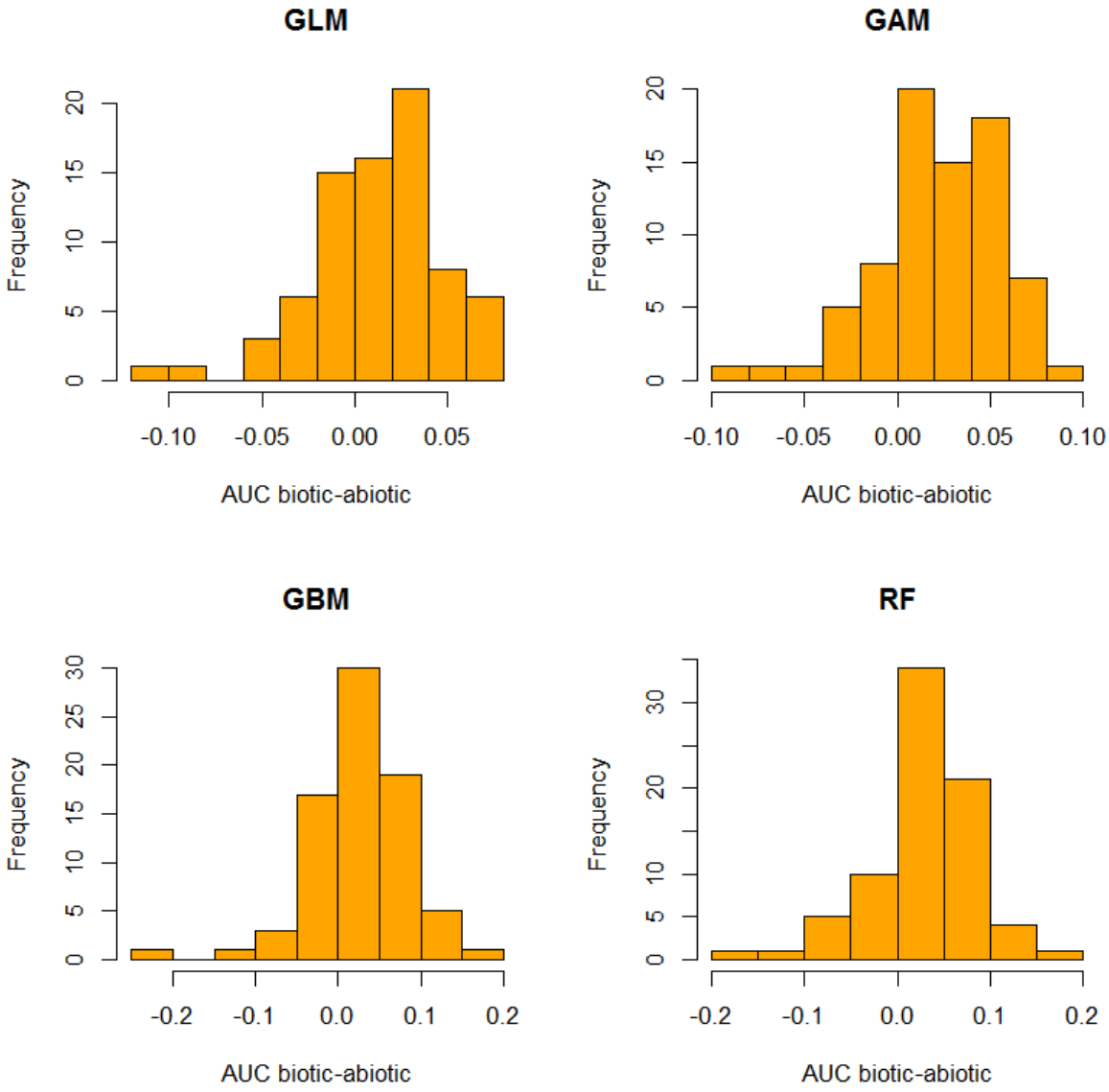


Figure S5 - Histograms of the difference between the predictive power of the models (AUC) without and with the trophic interaction as a predictor. Results for the four modelling techniques (GLM, GAM, GBM and RF) are presented.



44

45

46 **Table S1** - Species belonging to Papilionoidea that were found in the study area.

Species	Observations
Hesperiidae	
Carterocephalus palaemon	18
Erynnis tages	57
Hesperia comma	34
Ochlodes venatus	29
Pyrgus alveus	34
Pyrgus andromedae	10
Pyrgus cacaliae	10
Pyrgus carlinae	1
Pyrgus malvae	23
Pyrgus malvoides	4

Pyrgus serratulae	35
Spialia sertorius	12
Thymelicus lineola	72
Thymelicus sylvestris	31
Lycaenidae	
Aricia agestis	1
Aricia artaxerxes	45
Aricia eumedon	46
Callophrys rubi	28
Celastrina argiolus	2
Cupido alcetas	1
Cupido minimus	131
Cupido osiris	2
Glaucopsyche alexis	1
Lycaena helle	5
Lycaena hippothoe	35
Lycaena phlaeas	2
Lycaena tityrus	29
Maculinea arion	46
Maculinea nausithous	2
Maculinea rebeli	1
Plebeius glandon	18
Plebeius orbitulus	10
Plebeius argus	20
Plebeius idas	7
Polyommatus damon	27
Polyommatus semiargus	101
Polyommatus bellargus	21
Polyommatus coridon	39
Polyommatus dorylas	9
Polyommatus eros	26
Polyommatus escheri	1
Polyommatus icarus	51
Polyommatus thersites	12
Riodinidae	
Hamearis lucina	11
Nymphalidae	
Aglais urticae	166
Apatura iris	1
Argynnis adippe	18
Argynnis aglaja	73
Argynnis niobe	26
Argynnis paphia	9
Boloria dia	6
Boloria euphrosyne	52
Boloria selene	2
Boloria titania	76
Boloria aquilonaris	2
Boloria napaea	40
Boloria pales	70
Brenthis daphne	7

Brenthis ino	17
Euphydryas aurinia	60
Euphydryas cynthias	3
Inachis io	9
Issoria lathonia	8
Limenitis camilla	1
Melitaea athalia	36
Melitaea parthenoides	8
Melitaea cinxia	5
Melitaea diamina	62
Nymphalis antiopa	1
Polygonia c.album	6
Vanessa cardui	56
Vanessa atalanta	28
Aphantopus hyperantus	36
Brintesia circe	3
Coenonympha gardetta	77
Coenonympha pamphilus	47
Erebia aethiops	83
Erebia alberganus	9
Erebia cassioides	12
Erebia epiphron	26
Erebia eriphyle	5
Erebia euryale	40
Erebia gorge	22
Erebia ligea	47
Erebia manto	79
Erebia medusa	6
Erebia melampus	87
Erebia meolans	9
Erebia montana	6
Erebia oeme	77
Erebia pandrose	38
Erebia pharte	57
Erebia pluto	18
Erebia pronoe	22
Erebia tyndarus	28
Erebia mnestra	1
Lasiommata maera	60
Lasiommata megera	5
Lasiommata petropolitana	35
Maniola jurtina	44
Melanargia galathea	31
Oeneis glacialis	4
Pararge aegeria	8
Papilionidae	
Papilio machaon	86
Parnassius apollo	20
Parnassius mnemosyne	7
Parnassius phoebus	6
Pieridae	
Anthocharis cardamines	46

Aporia crataegi	58
Colias alfacariensis	20
Colias crocea	20
Colias hyale	9
Colias palaeno	6
Colias phicomone	72
Euchloe simplonia	5
Leptidea reali	7
Leptidea sinapis	5
Pieris brassicae	46
Pieris bryoniae	72
Pieris napi	66
Pieris rapae	70
Pontia callidice	12

47

48

49