**LETTER**

## Predicting novel trophic interactions in a non-native world

Ian S. Pearse[1]*[†] and Florian Altermatt[2][†]

### Abstract

Humans are altering the global distributional ranges of plants, while their co-evolved herbivores are frequently left behind. Native herbivores often colonise non-native plants, potentially reducing invasion success or causing economic loss to introduced agricultural crops. We developed a predictive model to forecast novel interactions and verified it with a data set containing hundreds of observed novel plant–insect interactions. Using a food network of 900 native European butterfly and moth species and 1944 native plants, we built an herbivore host-use model. By extrapolating host use from the native herbivore–plant food network, we accurately forecasted the observed novel use of 459 non-native plant species by native herbivores. Patterns that governed herbivore host breadth on co-evolved native plants were equally important in determining non-native hosts. Our results make the forecasting of novel herbivore communities feasible in order to better understand the fate and impact of introduced plants.

### Keywords

Herbivory, host breadth, invasive species, novel interaction, phylogenetic constraint.

*Ecology Letters* (2013) **16**: 1088–1094

### INTRODUCTION

Non-native plants are becoming a ubiquitous component of most of the world's ecosystems. Despite the growing prevalence of non-native plants, there are few effective tools for predicting the fate of non-native plants or their impacts on native communities (NAS 2002). One major determinant of a non-native plant's success and its effect on native communities is whether and to what degree that plant is colonised by native herbivores (Maron & Vilá 2001; McEvoy 2002). Generally, many non-native plants lose co-evolved herbivores when actively or passively introduced into a new area, and the loss of the co-evolved herbivores ('enemy release') is seen as one important reason for non-native plants' success and invasiveness (Keane & Crawley 2002). However, the colonisation of non-native plants by herbivores that are native in the novel range may negate the effects of enemy release and reduce plant invasiveness (Maron & Vilá 2001; Cappuccino & Carpenter 2005). The inclusion of a novel host by an herbivore may happen immediately due to plasticity in host use, or may develop gradually, based on the evolution of herbivore host use (Singer *et al.* 1993). In either case, the colonisation of non-native plants by herbivores can have a profound effect on biotic communities in invaded areas, plant invasion success and the yield of introduced crops (Maron & Vilá 2001; McEvoy 2002; Agrawal *et al.* 2005; Tylianakis *et al.* 2008).

While the occurrence of novel host–plant interactions is well documented for many case examples (Graves & Shapiro 2003; Pearse & Hipp 2009), there exist no general predictive tools to explain the occurrence of all novel interactions in a community. The lack of such tools is startling, as an accurate *a priori* prediction of novel plant–insect interactions would be of great ecological and economical relevance (McEvoy 2002). The opportunity for local herbivores to interact with novel plants is now the norm, and it is a major goal of ecologists to forecast these interactions even prior to the plant's

introduction. Forecasting of novel herbivore–plant interactions would allow for better screening of potential invasive plants that lack potential herbivores in the novel range, better economic assessment of pest control strategies for non-native crops, and better understanding of how plant introductions will affect native herbivore communities. Specifically, one could predict which plants are likely to be released from top down control when introduced.

One way to predict novel herbivore–plant interactions is to extrapolate from the existing native plant–herbivore interactions in a region. Several patterns common to native herbivore-plant food webs may extend to novel interactions with non-native plants. First, most insect herbivores have a defined host breadth, and consume a small to very small proportion of the total plant hosts that are available to them (Lewinsohn *et al.* 2006; Dyer *et al.* 2007; Melian *et al.* 2009; Novotny *et al.* 2010; Pearse & Altermatt 2013). Next, herbivores often consume a taxonomically defined subgroup of a plant community (Futuyma & Mitter 1996; Odegaard *et al.* 2005; Weiblen *et al.* 2006). Consequently, non-native plants that are highly evolutionary distinct from a native plant community experience less damage from local herbivores (Connor *et al.* 1980; Hill & Kotanen 2009; Pearse & Hipp 2009). Finally, the role of phylogenetic similarity to a native may be weaker when native herbivores have a greater host breadth. Thereby, generalists may include more taxonomically distinct non-native plants in their diet (Bertheau *et al.* 2010; Pearse 2011). While the role of host breadth and phylogenetic similarity have long been thought to influence herbivore associations with non-native plants (Connor *et al.* 1980), there have been few attempts to use these factors to broadly and quantitatively predict the extension of herbivore–plant food networks onto non-native plants.

We extend niche modelling (Peterson *et al.* 2011) to trophic interactions (Fig. 1), by adding host breadth of herbivores and phylogenetic constraint in herbivore host use in an interaction network of 900 European butterflies and moth (Lepidoptera) and
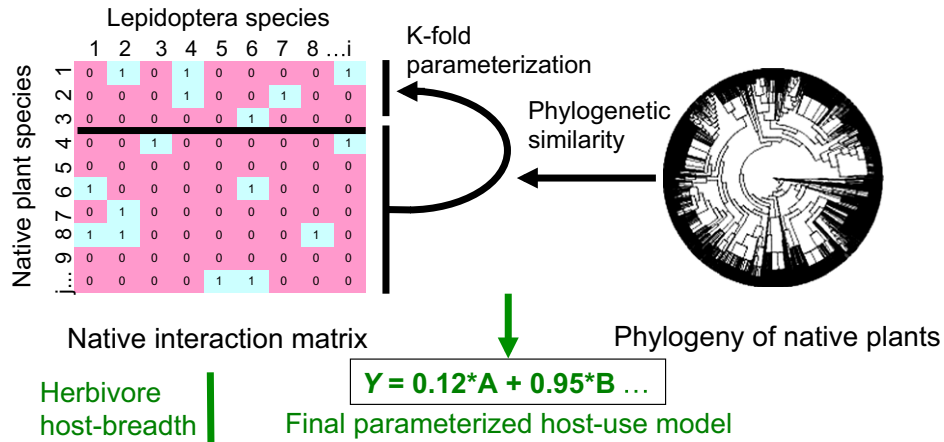
[1]*Lab of Ornithology, Cornell University, Ithaca, NY, USA*
[2]*Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, Dübendorf, CH-8600, Switzerland*
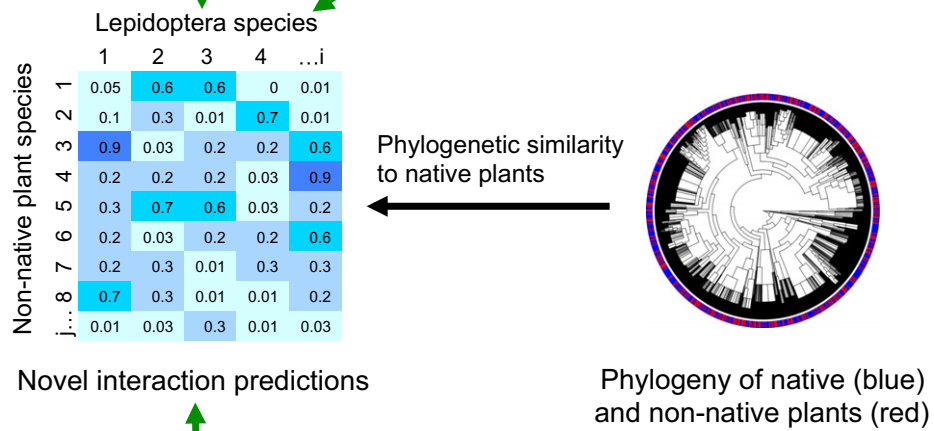
*Correspondence: E-mail: ianspearse@gmail.com*
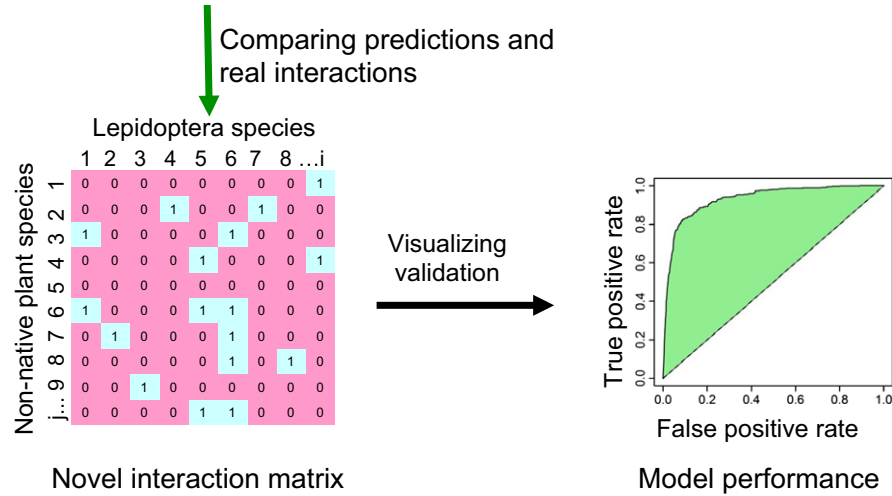[†]*Both authors contributed equally to all parts of this study.*

## STEP 1: Model building: native interactions

Lepidoptera species

K-fold parameterization

Phylogenetic similarity

Native interaction matrix

Phylogeny of native plants

Herbivore host-breadth

$$Y = 0.12*A + 0.95*B \ldots$$

Final parameterized host-use model

## STEP 2: Extrapolation to novel interactions

Lepidoptera species

Phylogenetic similarity to native plants

Novel interaction predictions

Phylogeny of native (blue) and non-native plants (red)

## STEP 3: Validation of predictions with real novel host-use

Comparing predictions and real interactions

Lepidoptera species

Visualizing validation

Novel interaction matrix

Model performance

True positive rate

False positive rate

**Fig. 1** A diagram of predictive modelling techniques. Step 1: Looking only at native interactions, we divided the native plant–Lepidoptera interaction matrix into predictor and evaluation data set using K-fold data-splitting. With this, we parameterised a logistic model where herbivore host breadth crossed with phylogenetic similarity of the evaluation plant to a host of the herbivore within the predictor plant data set determined host use. Step 2: Next, we extrapolated this model to non-native plants using native herbivore host breadth and the phylogenetic distance of a non-native plant to a native host. Step 3: We assessed the performance of our novel host predictions by comparing the predicted model scores for each novel herbivore–plant interaction with the novel herbivore–plant interactions that have been observed in Central Europe. We visualised the model performance as an ROC curve, plotting the rate of true positive predictions vs. the rate of false-positive predictions.

1944 native European plants (Ebert 1991–2005; Altermatt 2010; Altermatt & Pearse 2011). We first constructed a phylogenetic supertree of all 2403 native and non-native plants within our 35 752 km$^2$ study region in Central Europe (Fig. S1). We parameterised a logistic model where host phylogenetic similarity, herbivore host breadth, and the interaction of those two terms explained the native host use of herbivores using data partitioning (k-fold data-splitting, Fig. S2). As a retrospective analysis, we then used this model and the phylogenetic similarity of introduced plants to native host plants to predict the likelihood of interaction between the 900 native herbivores and the 459 non-native plant species introduced since 1492.

## METHODS

### Study area

Our study area is situated in Germany, Central Europe. We focused on Baden-Württemberg (centre coordinates: 48°32′16″ N 9°2′28″ E), which is a southwestern German state to the east of the Upper Rhine. The study area covers 35 752 km$^2$. Extensive faunistic records on plants and Lepidoptera are available from Baden-Württemberg (Ebert 1991–2005; Bundesamt für Naturschutz 2010).

### Data on Lepidoptera and plants

Our study considers all Lepidoptera species (butterflies and moths) belonging to the clades of Bombycoidea, Cossoidea, Drepanoidea, Geometroidea, Hepialoidea, Lasiocampoidea, Noctuoidea, Papilionoidea (including Hesperiidae), Psychoidea, Sesioidea, Thyridoidea and Zygaenoidea that have been recorded in Baden-Württemberg and were all considered by Ebert (1991–2005). These families do not form a monophyletic subgroup within the Lepidoptera (e.g. Regier et al. 2013), but have been traditionally treated together in Central Europe (e.g. Schweizerischer Bund für Naturschutz 1987; Ebert 1991–2005; Koch & Heinicke 1991; Pro Natura - Schweizerischer Bund für Naturschutz 2000). For a phylogenetic classification of all these butterfly and moth species, see Altermatt & Pearse (2011).

We used published data on insect–plant interactions of larval Lepidoptera (Ebert 1991–2005; Koch & Heinicke 1991). Interactions with larval Lepidoptera describe observations of food plants used by the larvae. All data are based on observations made under natural, un-manipulated field situations. In total, the data set contains 5520 species-specific insect–plant interactions, based on > 2.35 million larval individuals recorded (Altermatt & Pearse 2011). Each specific insect–plant interaction may be based on one to many (> 100) observations in the field. We only considered records resolved to the plant species level, and did not consider the (relatively few) records that were only giving genus or higher taxonomic resolution. This approach was not only most conservative but also the only one that allowed us to classify each host plant into native or non-native.

We considered in total 2403 different plant species found in Baden-Württemberg as potential host plants for the Lepidoptera. These 2403 plant species consist of all vascular plants known from Baden-Württemberg (Bundesamt für Naturschutz 2010), including native, non-native and widely grown ornamental plants. Three plant species are known to include a large number of closely related (sub) species or apomictic species (Rubus fruticosus agg., Hieracium agg., Taraxacum officinalis agg.). These taxa are often not differentiated in field-observations of plant–insect interactions or the taxonomic status is

disputed. We thus used the aggregated species-complex for each of these plants. We did not pool further groups of plants into species complexes (e.g. Alchemilla sp., Ranunculus sp. or Carex sp.), as for these species specific records were generally given by Ebert (1991–2005).

Of the 2403 plants, 1944 are native species of Baden-Württemberg (including archaeophytes), and 459 are non-native species (i.e. species that arrived/were introduced in Baden-Württemberg after the year 1492). For the classification of non-native vs. native plants, we followed Bundesamt für Naturschutz (Bundesamt für Naturschutz 2010), but used additional references to confirm a few disputed cases (Landolt et al. 2010). Non-native plants include neophytes and introduced ornamental plants. Populations of the non-native plants can range from sporadic occurrence to established or invasive populations. We estimated a species-level supertree phylogeny for the 2403 native and non-native plant species of Baden-Wuerttemberg using Phylomatic (Webb & Donoghue 2005) with the most resolved current seed plant (Spermatophyte) phylogeny as the backbone. We used the Phylocom BLADJ procedure to scale branch lengths according to inferred node ages for seed plant lineages (Wikstrom et al. 2001). To assess how susceptible our analyses were to inaccuracies in branch lengths in the plant phylogeny, we created a permuted phylogeny, which retained phylogenetic topology, but set all branch lengths to one.

We compiled an extensive database on larval food plants for all Lepidoptera species, based on a series of publications that are focusing on Lepidoptera in Baden-Württemberg or immediate surroundings (Schweizerischer Bund für Naturschutz 1987; Ebert 1991–2005; Koch & Heinicke 1991; Pro Natura - Schweizerischer Bund für Naturschutz 2000; Altermatt et al. 2006). Most (> 95%) of the plant–insect interactions in our data set are based on an extensive monograph on the Lepidoptera of Baden-Württemberg (Ebert 1991–2005), but were complemented by the other references and personal observations.

All plant–insect interactions are based on naturally observed records. Based on > 2.35 million larval individuals recorded (Altermatt & Pearse 2011), the 900 Lepidoptera species develop on 759 vascular plant species as larval food plants. The 759 plant species include plants from 386 genera and 88 families. Of these plants, 586 are native to Baden-Württemberg, 173 are non-native (23%). One lepidopteran species can use one to many plant species as a food plant, and one plant species can be used by one to many lepidopteran species. We described the host breadth of the larva as the number of plant species used by a lepidopteran species.

### Model-building and predictions

We built and assessed a predictive model of novel herbivore–plant interactions in a three-step process (Fig. 1). In step 1, we built a model of host use based on herbivore interactions with native plants (Fig. 1). Specifically, we used a K-fold procedure to split the native plant–herbivore matrix into five (k = 5) random partitions, where each partition contained 1/5 (1/k) of the native plants and their interactions with all Lepidoptera. For each of the five partitions, we treated four partitions as the 'calibration' data and the fifth as the 'evaluation' data, noting that meaning of these terms differ slightly from previous uses (Peterson et al. 2011). We estimated the parameters of a generalised linear model using two predictors and their interaction: (1) the number of hosts of a given herbivore, (2) the 'phylogenetic similarity', i.e. the minimum phylogenetic distance of a given 'evaluation' plant to any 'prediction' host plant of a given her-

bivore, to predict the binomial response variable of host use of herbivores with 'evaluation' plants. We then averaged the parameter estimates from the five data partitions, and used deviation between partitions to estimate model uncertainty (Table S1). Both parameters and their interaction significantly predicted in-sample host use (Table S1) so we retained all parameters in our final model. In a separate analysis, we tested whether different numbers of data partitions (varying k from 2 to 10) had a significant effect on parameter estimates, and we found no evidence for this (Fig. S2), and thus used k = 5 for all subsequent analyses. In step 2, we projected the averaged host-use model onto all plants introduced into Baden-Württemberg after the year 1492 (Fig. 1). We obtained values for the two predictor variables, herbivore host breadth (on native plants) and 'phylogenetic similarity', and we estimated prediction scores for each Lepidoptera-non-native plant interaction using the linear formula:

$$P_{hp} = m_h * H + m_s * S + m_{sh} * S * H,$$

where $P_{hp}$ is the model score for a given novel herbivore–plant interaction, H is the host breadth of a given herbivore and S is minimum phylogenetic distance (similarity) of a non-native plant to any native host of a given herbivore. In step 3, we validated our predictions using a data set of actual interactions of Lepidoptera with non-native plants. Model scores partially separated realised herbivore-non-native plant interactions from non-realised interactions (Fig. S3). We used model scores to create a receiver–operating characteristic (ROC) curve that plots the rate of true positive predictions against the rate of false-positive predictions, and to calculate the area under the ROC curve (AUC), a statistic of model predictiveness. In a separate, *post hoc* analysis, we constructed models using either host breadth or phylogenetic similarity as single predictors of host use, and we found that the model with both factors more accurately predicted novel interactions than models based on either factor alone (Kolmogorov–Smirnov Tests; P < 0.001; Fig. S4). In a final analysis, we combined the last k-fold portion of within-sample native predictions with the out-of-sample non-native predictions. From this, we constructed a logistic model to ask whether 'native-

ness' interacted with either herbivore host breadth or plant relatedness. All statistics and predictive modelling was conducted using R (R Core Development Team 2008) with packages ape (Paradis *et al.* 2004) for dealing with phylogenetic information and ROCR (Sing *et al.* 2009) for visualising ROC curves.

For the evaluation of the use of non-native plants by insects at genus or family level, we calculated the number and percentage of all non-native taxa (at genus or family level) which are used by the native Lepidopteran herbivores. We did these calculations separately for all plants with and without a native confamilial and with and without a native conger plant species. To avoid pseudo-replication, we used a conservative approach and counted each plant genus or family only once, irrespective of the number of species it has. By having the genus/family level as the unit of replication, we avoided that species-rich genera/families that have or have not many insect interactions are biasing the comparison. We used a Fisher's exact test to analyse the contingency tables of non-native taxa used as host plants.

## RESULTS

In accord with previous studies (Odegaard *et al.* 2005; Weiblen *et al.* 2006), the final native host-use model showed that Lepidoptera species used a phylogenetically constrained set of native hosts (in all data partitions P < 0.001) and that phylogenetic host constraint was greater for more specialised Lepidoptera than for generalists (in all data partitions P < 0.001, Table S1).

The interactions between native hosts and insects accurately predicted novel herbivore–plant interactions that have developed with non-native plants in Central Europe since 1492 (Fig. 2; solid line). Model scores accurately separated realised herbivore–plant interactions from non-interactions (Fig. S3). A non-informative null model (Fig. 2; dashed line) predicted novel host use in an equal proportion of cases where host use was observed (True Positives) and cases where host use was not observed (False Positives). Increased predictive ability beyond the null model (green shaded area) indicates that a greater portion of true interactions were predicted with
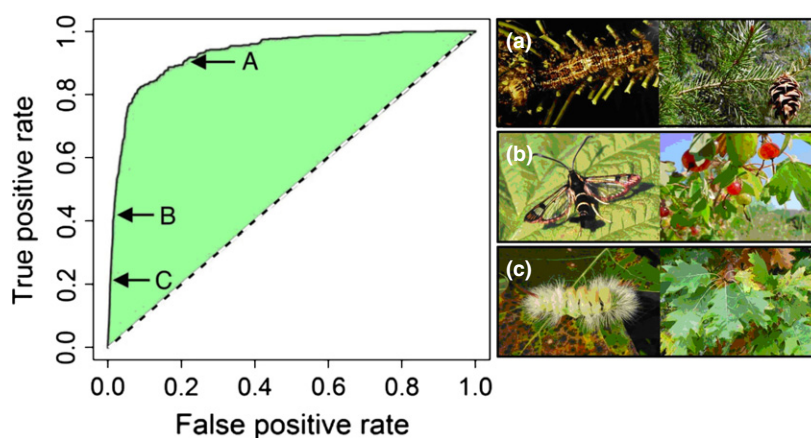


**Fig. 2** Predicted incorporation of 459 non-native plants into the larval diet of 900 European Lepidoptera species. The ROC curve of the final model predicting host use of larval Lepidoptera on non-native plants. The solid line and green shading indicate the predictive value of the model, the dashed line represents a null model. When the true positive rate quickly approaches 1, the model discriminates well between hosts and non-hosts. (a-c) Three examples of European moth species that extended their larval host range to non-native (North-American) plants. Arrows on the ROC curve indicate the model scores of each example. Interactions between native hosts and insects in the model accurately predict (c) the larval diet extension of the Tussock moth to the non-native red oak and (b) the larval diet extension of a sesiid moth to introduced gooseberry. The diet extension (a) of the Pine-tree Lappet moth onto the non-native Douglas fir is poorly predicted, i.e. at a higher rate of false-positives. Photos courtesy of F. Altermatt, Ruedi Bryner and the Dale A. Zimmerman Herbarium at Western New Mexico University.

**Table 1** Predictions for Lepidoptera subgroups. As the relationship between host breadth and phylogenetic host constraint likely differs between Lepidopteran subgroups, we made a native host-use model for the eight major Lepidoptera subgroups within our data set and projected that model onto non-native host use. We present the model performance (AUC) in predicting non-native interactions, where AUC = 1 represents a perfectly predictive model and AUC = 0.5 represents an uninformative model

| Lepidoptera taxon | # species | Number of host plants (mean ± SD) | Model performance (AUC) |
|---|---|---|---|
| All Lepidoptera | 900 | 5.75 ± 7.91 | 0.930 |
| Hesperioidea + Papilionoidea (i.e. butterflies) | 126 | 4.79 ± 5.68 | 0.986 |
| Hepialoidea | 5 | 4.20 ± 2.17 | 0.533 |
| Psychidae | 16 | 5.63 ± 4.13 | 0.884 |
| Zygaenoidea + Cossoidea + Sesioidea | 55 | 4.20 ± 4.30 | 0.982 |
| Bombycoidea + Lasiocampidae | 23 | 8.57 ± 8.22 | 0.927 |
| Drepanidae | 14 | 2.64 ± 1.82 | 0.997 |
| Geometroidea | 283 | 5.00 ± 7.39 | 0.929 |
| Noctuoidea | 360 | 6.56 ± 9.04 | 0.909 |

a lower false-positive rate (Fig. 2). The interactions between native hosts and insects was significantly more predictive than the null model (AUC = 0.930, $P_{randomised} < 0.001$) and predicted 82.8% of all actual novel interactions with a low (10%) false-positive prediction rate (Fig. 2). The full model that included both herbivore host breadth and phylogenetic similarity was more predictive than a model including either host breadth or phylogenetic similarity alone (Fig. S4). To assess whether our predictions were susceptible to inaccuracies in the time-calibration of the plant phylogeny, we ran our analysis using a tree with all equal branch lengths, and we found that this did not reduce our predictive ability, but rather increased it marginally (AUC = 0.942). To assess whether alternate definitions of herbivore host breadth affected host-use model predictions, we ran another analysis defining herbivore host breadth as the number of plant genera consumed, but we found that this has little bearing on the predictions of the model (AUC = 0.932), likely because these two estimates of host breadth were highly correlated (Fig. S5). As host records are likely incomplete for some Lepidoptera taxa, especially for polyphagous noctuid and geometrid moths, we conducted an identical analysis constrained to plant interactions with butterflies (i.e. superfamily Papilionoidea including Hesperiidae), which have nearly complete host sampling. In the butterfly-only data set, the interactions between native hosts and insects predicted 100% of all actual novel interactions with a false-positive prediction rate of only 7% (Fig. S6). We repeated this analysis for seven other subgroups of Lepidoptera with differing average host breadths, and presumably sampling intensity (Table 1). We found that our model accurately predicted novel host use of most Lepidopteran groups with the exception of the species-poor Hepialoidea (Table 1). We conducted an additional *post hoc* analysis to assess whether the factors that constrained host associations differed between native and novel hosts, and we found no evidence for this (phylogenetic similarity $P_{simulated} = 0.35$; host breadth: $P = 0.13$).

In our native host-use model, most of the variation in insect–plant interactions was explained by evolutionary relationships between plants. Thus, we further explored how taxonomic similarity of non-native plants to local natives affects host affiliations of
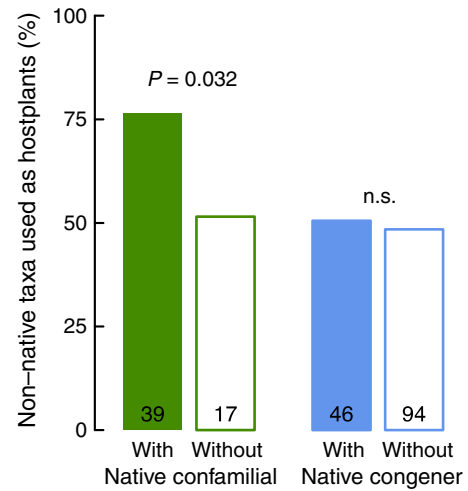


**Fig. 3** Observed percentage of non-native plant taxa used as larval host plants by Central European butterflies and moths. The percentage of used plants is given separately for non-native plants with (filled bars) and without (open bars) native plants belonging to the same family (green bars) or the same genus (blue bars). Non-native plants that have native confamilial host plants are significantly more often used as larval host plants compared to non-native plants that do not have native confamilials. No such difference was found at the genus level. Numbers at the bottom of the bars give the absolute values of non-native plant families and plant genera used as larval host plants.

native insects. We divided novel plants into those that had a native congener, a native con-familial, or no native relatives within Central Europe. We found that non-native plants with a native con-familial were more likely to be colonised by native Lepidoptera than those without native con-familials (Fisher's Exact Test, $P = 0.032$, Fig. 3). However, at finer level of taxonomic specificity, non-native plants with native congeners were just as likely to be colonised by native Lepidoptera as those without native congeners (Fisher's Exact Test, $P = 0.8$), suggesting that the plant family level is most important in determining the host breadth of herbivores (Fiedler 1998).

## DISCUSSION

We demonstrated that novel interactions between herbivores and non-native plants can be predicted based on plant evolutionary relationships and properties of the native herbivore–plant food web. We were able to predict the majority of novel interactions between herbivores and non-native plants with a small rate of false predictions (Fig. 2). The high predictability of novel herbivore–plant interactions demonstrates the primacy of phylogenetic host constraint for most herbivores even in non-co-evolved systems, and suggests that the properties that govern native herbivore–plant food webs are the same for novel food webs. From a practical standpoint, we predicted numerous novel herbivore–plant interactions of economic or conservation importance, and similar techniques can be used to predict the consequences of plant introductions.

### Non-native plants as food web components

Numerous recent studies have explored the mechanisms by which herbivores affiliate with particular host plants (Stamp 2003; Futuyma & Agrawal 2009), and others have noted that explanations of herbivore–plant interactions must be transferable to novel (non-

co-evolved) interactions (McEvoy 2002). Previous studies showed that plant evolutionary history (presumably a proxy for conserved plant traits) constrains herbivore host affiliations with native, co-evolved plants (Odegaard *et al.* 2005; Weiblen *et al.* 2006) (Table S1) and suggest that phylogenetically defined host use is a central tenet of co-evolution (Janz 2011). Our study shows that plant evolutionary history also constrains herbivore interactions with non-co-evolved, non-native plants (Fig. 2). Moreover, deep evolutionary history (corresponding roughly to the 'family' level of organisation) is the greatest constraint of novel herbivore–plant interactions (Fig. 3).

Introduced plants represent a non-random sample of plants, biased towards those species or individuals that are able to disperse to and colonise a new region (Colautti *et al.* 2006). As such, introduced plants may interact with herbivores in a fundamentally different way from native plants that have not experienced a colonisation bias (e.g. Parker *et al.* 2006). On the contrary, we found that phylogenetic host constraint and inherent herbivore host breadth equally described native (co-evolved) and novel (non-co-evolved) herbivore plant interactions.

## Practical uses of predicting novel food webs

On the basis of interactions between native hosts and insects, we also successfully predicted specific diet extensions of potential European pest insects to plants of forestry or agricultural interest introduced from North America, as well as the diet extension of European insects onto non-native plants that are of invasive concern. The goal of this approach is to correctly identify specific important interactions between a novel plant and native herbivore with the lowest possible false-positive rate, where a null model would result in a 50% false-positive rate. For example, we predicted that the tussock moth (*Calliteara pudibunda*) colonises red oak (*Quercus rubra*; a common introduced tree throughout Europe) with a false-positive rate of only 0.7% (Fig. 2c). The tussock moth is an herbivorous insect of forestry concerns, having mass-outbreaks (Schwenke 1978; Mazzoglio *et al.* 2005), and it is thus critical to understand its diet extension to novel host plants. Similarly, we predicted that the specialist Sessiid moth *Synanthedon tipuliformis* colonises *Ribes aureum*, a cultivated gooseberry introduced from North America, with a false-positive rate of only 2.0% (Fig. 2b). *S. tipuliformis* is known to cause damage in agricultural gooseberry plantations, and an accurate prediction of host switch to introduced agricultural gooseberries is thus economically important. Only when forecasting interactions of herbivores with introduced conifers, false-positive rates became consistently higher. For example, the model reached a false-positive rate of 22% before correctly predicting the interaction of the Pine-tree Lappet moth (*Dendrolimus pini*), a potential forest pest (Schwenke 1978), with the introduced Douglas fir (*Pseudotsuga menziesii*, Fig. 2a), i.e. our model (incorrectly) predicted that 22% of non-native plants were better hosts for *D. pini* than *P. menziesii*. The more limited prediction of herbivore interactions with conifers (gymnosperms) may reflect differences in co-evolutionary histories of insects with gymnosperms than with angiosperms (Farrell 1998). As a rough simplification, angiosperms diversified together with phytophagous insects, while the main gymnosperm lineages were already established before this major radiation of insects. Thus, there is no (or less) potential for a co-evolutionary signal to predict host switches in conifers. This is consistent with findings showing that introduced conifers without native congeners are generally not

very extensively used as host plants by native arthropods (Roques *et al.* 2006), suggesting that such switches are not only harder to predict, but possibly also harder to make.

We also accurately predicted which native insects feed on non-native invasive plants such as *Solidago* sp. (Fig. S7), and may have a substantial negative fitness effect on those invasives. In Central Europe, the geometrid moths *Eupithecia virgaureata* and *E. absinthiata* feed (among other plants) on various native mugworts and ragworts (*Artemisia* and *Senecio* sp.). Over the last decades, they extended their diet onto invasive goldenrods (*Solidago canadensis* and *S. gigantea*). We predicted this host shift with a false-positive rate of only 0.08%. As the larvae of *Eupithecia* moths selectively feed on flowers and developing seeds, this shift opens up the possibility for ecological and evolutionary dynamics leading to a more pronounced use of these plants and the moths may eventually even have a negative fitness effect on invasive *Solidago* sp. In each of these cases, our model only predicts the incidence of the novel interaction, but neither the prevalence nor the importance of the interaction to plant fitness or invasiveness. A more robust understanding of how common the novel interaction will be and how well a given non-native plant tolerates herbivory will be needed before these models can predict the importance of novel herbivores as fitness limiting enemies of non-native plants (Burghardt *et al.* 2009).

## CONCLUSIONS

In a retrospective analysis, we demonstrated that it is possible to accurately predict the majority of novel host use of non-native plants with only information about a native food web and the evolutionary relationships of plants. Our model can be used to predict herbivore-use of novel crops, screen potential biocontrol agents, or anticipate the impact of plant invasions or climate-induced range expansions on insect populations. Similar to innovations in geographical niche modelling (Peterson *et al.* 2011), future trophic modelling techniques for novel host interactions will likely include more information about plant defensive traits, more sophisticated computational techniques to account for greater nonlinearities and interactions among predictors, and techniques that incorporate herbivore relationships and traits into predictions. It has been very difficult to understand consequences of plant introductions for native systems (NAS 2002), but our study shows that it is actually possible to predict most of the novel trophic interactions between native herbivores and introduced plants.

## REFERENCES

Agrawal, A.A., Kotanen, P.M., Mitchell, C.E., Power, A.G., Godsoe, W. & Klironomos, J. (2005). Enemy release? An experiment with congeneric plant pairs and diverse above- and belowground enemies. *Ecology*, 86, 2979–2989.

Altermatt, F. (2010). Tell me what you eat and I'll tell you when you fly: diet can predict phenological changes in response to climate change. *Ecol. Lett.*, 13, 1475–1484.

Altermatt, F., Fritsch, D., Huber, W. & Whitebread, S. (2006). *Die Gross-Schmetterlingsfauna der Region Basel*. Entomologische Gesellschaft Basel, Basel.

Altermatt, F. & Pearse, I.S. (2011). Similarity and specialization of the larval versus adult diet of European butterflies and moths. *Am. Nat.*, 178, 372–382.

Bertheau, C., Brockerhoff, E.G., Roux-Morabito, G., Lieutier, F. & Jactel, H. (2010). Novel insect-tree associations resulting from accidental and intentional biological 'invasions': a meta-analysis of effects on insect fitness. *Ecol. Lett.*, 13, 506–515.

Bundesamt für Naturschutz (2010). FloraWeb. Available at: http://www.floraweb.de/index.html. Last accessed 15 October 2012.

Burghardt, K.T., Tallamy, D.W. & Shriver, G.W. (2009). Impact of native plants on bird and butterfly biodiversity in suburban landscapes. *Conserv. Biol.*, 23, 219–224.

Cappuccino, N. & Carpenter, D. (2005). Invasive exotic plants suffer less herbivory than non-invasive exotic plants. *Biol Lett.*, 1, 435–438.

Colautti, R.I., Grigorovich, I.A. & MacIsaac, H.J. (2006). Propagule pressure: a null model for biological invasions. *Biol. Invasions*, 8, 1023–1037.

Connor, E.F., Faeth, S.H., Simberloff, D. & Opler, P.A. (1980). Taxonomic isolation and the accumulation of herbivorous insects: a comparison of introduced and native trees. *Ecol. Entom.*, 5, 205–211.

Dyer, L.A., Singer, M.S., Lill, J.T., Stireman, J.O., Gentry, G.L., Marquis, R.J., Ricklefs, R.E., Greeney, H.F., Wagner, D.L., Morais, H.C., Diniz, I.R., Kursar, T.A. & Coley, P.D. (2007). Host specificity of Lepidoptera in tropical and temperate forests. *Nature*, 448, 696–699.

Ebert, G. (ed.) (1991–2005). *Die Schmetterlinge Baden-Wurttembergs*. Ulmer, Stuttgart.

Farrell, B.D. (1998). "Inordinate fondness" explained: why are there so many beetles? *Science*, 281, 555–559.

Fiedler, K. (1998). Diet breadth and host plant diversity of tropical- vs. temperate-zone herbivores: South-east Asian and west Palaearctic butterflies as a case study. *Ecol. Entom.*, 23, 285–297.

Futuyma, D.J. & Agrawal, A.A. (2009). Macroevolution and the biological diversity of plants and herbivores. *PNAS*, 106, 18054–18061.

Futuyma, D.J. & Mitter, C. (1996). Insect-plant interactions: the evolution of component communities. *Phil. Trans. Royal Soc. B*, 351, 1361–1366.

Graves, S.D. & Shapiro, A.M. (2003). Exotics as host plants of the California butterfly fauna. *Biol. Conserv.*, 110, 413–433.

Hill, S.B. & Kotanen, P.M. (2009). Evidence that phylogenetically novel non-indigenous plants experience less herbivory. *Oecologia*, 161, 581–590.

Janz, N. (2011). Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies. *Ann. Rev. Ecol. Evol. System.*, 42, 71–89.

Keane, R.M. & Crawley, M.J. (2002). Exotic plant invasions and the enemy release hypothesis. *Trends. Ecol. Evol.*, 17, 164–170.

Koch, M. & Heinicke, W. (1991). *Wir bestimmen Schmetterlinge*. Ausgabe in einem Band, Neumann/Neudamm, Melsungen.

Landolt, E., Baeumler, B., Erhardt, A., Hegg, O., Kloetzli, F., Laemler, W., Nobis, M., Rudmann-Maurer, K., Schweingruber, F.H., Theurillat, J.P., Urmi, E., Vust, M. & Wohlgemuth, T. (2010). *Floria Indicativa: Ecological Indicator Values and Biological Attributes of the Flora of Switzerland and the Alps*. Haupt Verlag, Bonn.

Lewinsohn, T.M., Prado, P.I., Jordano, P., Bascompte, J. & Olesen, J.M. (2006). Structure in plant-animal interaction assemblages. *Oikos*, 113, 174–184.

Maron, J.L. & Vilá, M. (2001). When do herbivores affect plant invasion? Evidence for the natural enemies and biotic resistance hypotheses. *Oikos*, 95, 361–373.

Mazzoglio, P., Paoletta, M., Patetta, A. & Currado, I. (2005). *Calliteara pudibunda* (Lepidoptera Lymantriidae) in Nowrthwest Italy. *Bull. Insectol.*, 58, 25–34.

McEvoy, P.B. (2002). Insect-plant interactions on a planet of weeds. *Entomol. Exp. Appl.*, 104, 165–179.

Melian, C.J., Bascompte, J., Jordano, P. & Krivan, V. (2009). Diversity in a complex ecological network with two interaction types. *Oikos*, 118, 122–130.

NAS (2002). *Predicting Invasions of Nonindigenous Plants and Plant Pests*. National Academy Press, Washington, D.C.

Novotny, V., Miller, S.E., Baje, L., Balagawi, S., Basset, Y., Cizek, L., Craft, K.J., Dem, F., Drew, R.A.I., Hulcr, J., Leps, J., Lewis, O.T., Pokon, R., Stewart, A.J.A., Samuelson, G.A. & Weiblen, G.D. (2010). Guild-specific patterns of species richness and host specialization in plant-herbivore food webs from a tropical forest. *J. Anim. Ecol.*, 79, 1193–1203.

Odegaard, F., Diserud, O.H. & Ostbye, K. (2005). The importance of plant relatedness for host utilization among phytophagous insects. *Ecol. Lett.*, 8, 612–617.

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

Parker, J.D., Burkepile, D.E. & Hay, M.E. (2006). Opposing effects of native and exotic herbivores on plant invasions. *Science*, 311, 1459–1461.

Pearse, I.S. (2011). Leaf defensive traits in oaks and their role in both preference and performance of a polyphagous herbivore, *Orgyia vetusta*. *Ecol. Entom.*, 36, 635–642.

Pearse, I.S. & Altermatt, F. (2013). Extinction cascades partially estimate herbivore losses in a Lepidoptera-plant food web. *Ecology*, DOI: 10.1890/12-1075.1.

Pearse, I.S. & Hipp, A.L. (2009). Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. *PNAS*, 106, 18097–18102.

Peterson, A.T., Soberon, J., Pearson, R.G., Anderson, R.P., Martinez-Meyer, E., Nakamura, M. & Araujo, M.B. (2011). *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton, NJ.

Pro Natura - Schweizerischer Bund für Naturschutz (2000). *Schmetterlinge und ihre Lebensräume*. Fotorotar AG, Zurich.

R Core Development Team (2008). R v. 2.10.1. In. The R Foundation.

Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.C., Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C. & Mitter, K.T. (2013). A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS One*, 8, e58568.

Roques, A., Auger-Rozenberg, M.A. & Boivin, S. (2006). A lack of native congeners may limit colonization of introduced conifers by indigenous insects in Europe. *Can. J. For. Res.-Rev. Can. Rech. For.*, 36, 299–313.

Schweizerischer Bund für Naturschutz (1987). *Tagfalter und ihre Lebensräume*. Fotorotar AG, Zurich.

Schwenke, W. (1978). *Die Forstschädlinge Europas. Ein Handbuch in fünf Bänden Band 3: Schmetterlinge*. Paul Parey, Hamburg and Berlin.

Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. (2009). ROCR: Visualizing the performance of scoring classifiers. Available at: http://CRAN.R-project.org/package=ROCR. In. R package version 1.0-4. Last accessed 15 October 2012.

Singer, M.C., Thomas, C.D. & Parmesan, C. (1993). Rapid human-induced evolution of insect host associations. *Nature*, 366, 681–683.

Stamp, N. (2003). Out of the quagmire of plant defense hypotheses. *Q. Rev. Biol.*, 78, 23–55.

Tylianakis, J.M., Didham, R.K., Bascompte, J. & Wardle, D.A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecol. Lett.*, 11, 1351–1363.

Webb, C.O. & Donoghue, M.J. (2005). Phylomatic: tree assembly for applied phylogenetics. *Mole. Ecol. Notes*, 5, 181–183.

Weiblen, G.D., Webb, C.O., Novotny, V., Basset, Y. & Miller, S.E. (2006). Phylogenetic dispersion of host use in a tropical insect herbivore community. *Ecology*, 87, S62–S75.

Wikstrom, N., Savolainen, V. & Chase, M.W. (2001). Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. B-Biol. Sci.*, 268, 2211–2220.

## SUPPORTING INFORMATION

Supplementary Information for:  : " Predicting novel trophic interactions in a non-native world".  **Figures S1-S7**, **Table S1**
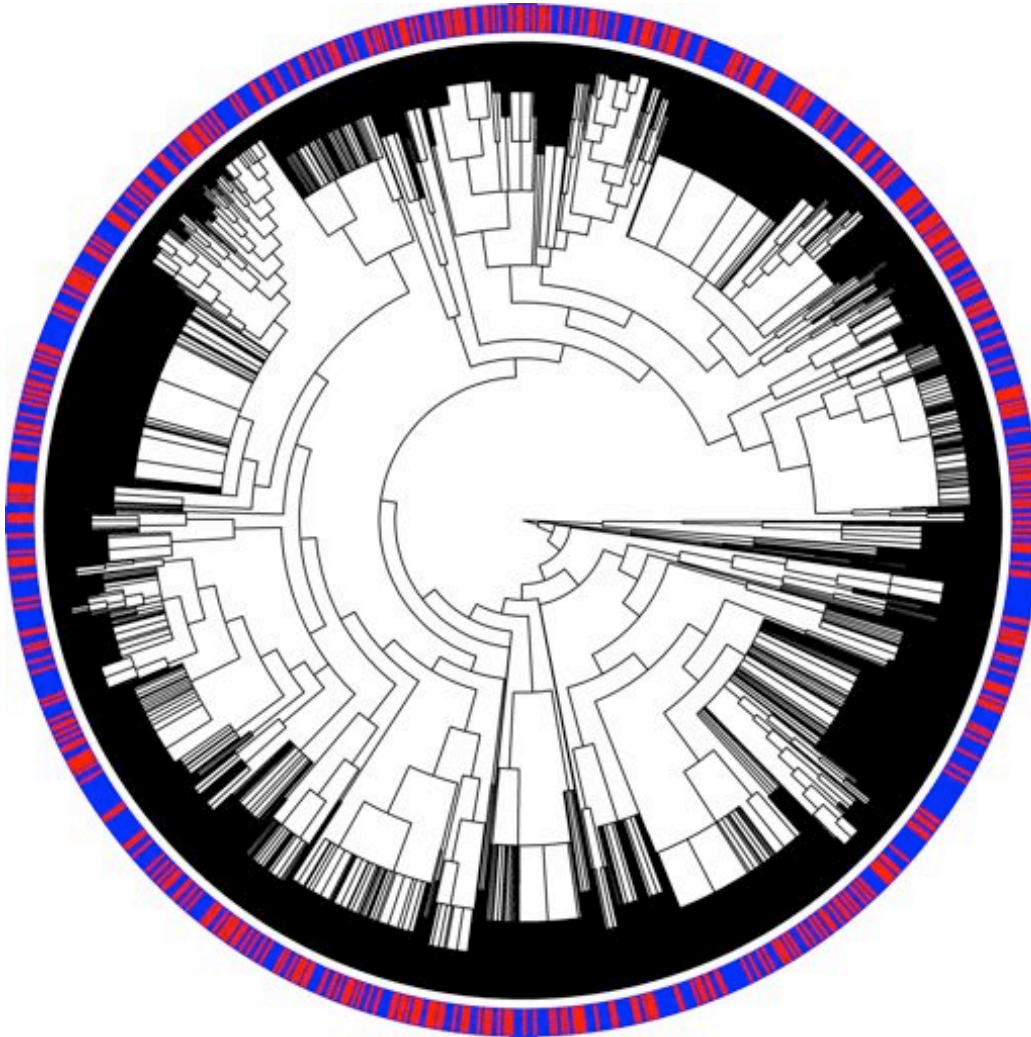


**Fig. S1:  A phylogenetic supertree of the native (blue) and non-native (red) seed plants (Spermatophyta) used in this study**. We constructed a supertree of phylogenetic relationships among 2,403 vascular plants in Central Europe.  Plant taxa were added to the tree based on taxonomic affiliation at the genus or family level.  Within genus, all relationships are analyzed as polytomies.  Plant species from across the evolutionary history of plants have been introduced to Central Europe.  Non-native (red) tips are printed wider than native tips (blue) to allow visualization.
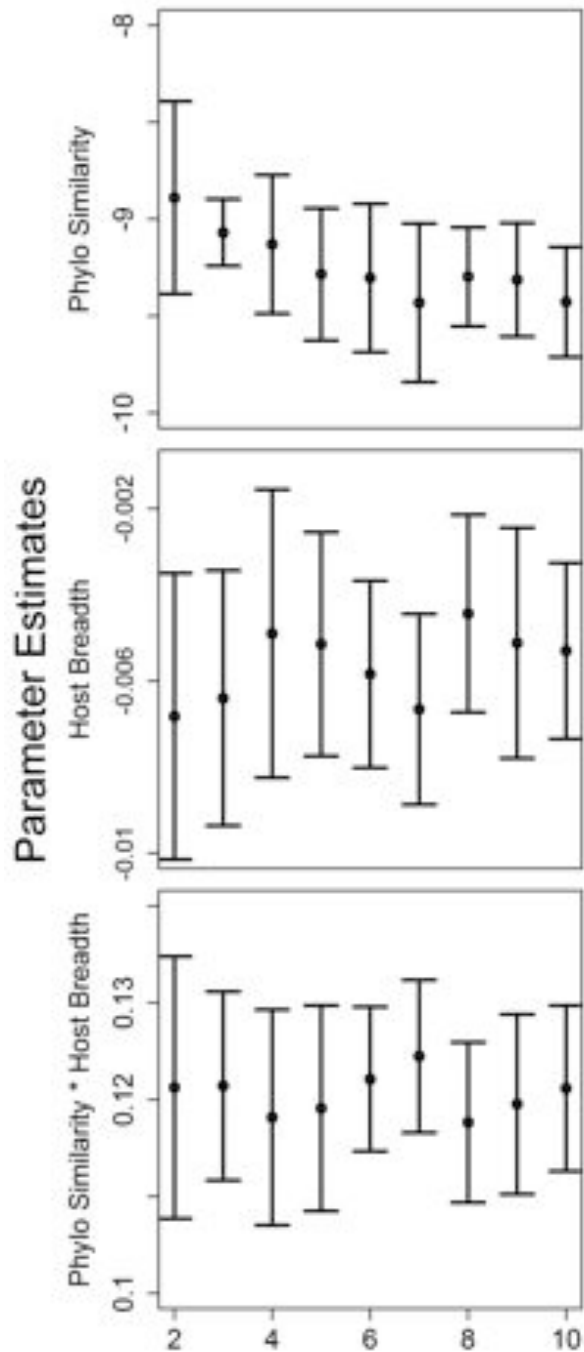
**Fig. S2: Sensitivity of parameter estimates to different numbers of data partitions.**
We partitioned the native plant-herbivore food web into 2-10 partitions (k 2 to 10), and calculated parameter estimates for each of the three factors in our full model. Bars represent the mean parameter estimate ± SE. Parameter estimates were insensitive to the respective partition level, and we subsequently used a partition-value k of five (5-fold partitioning).
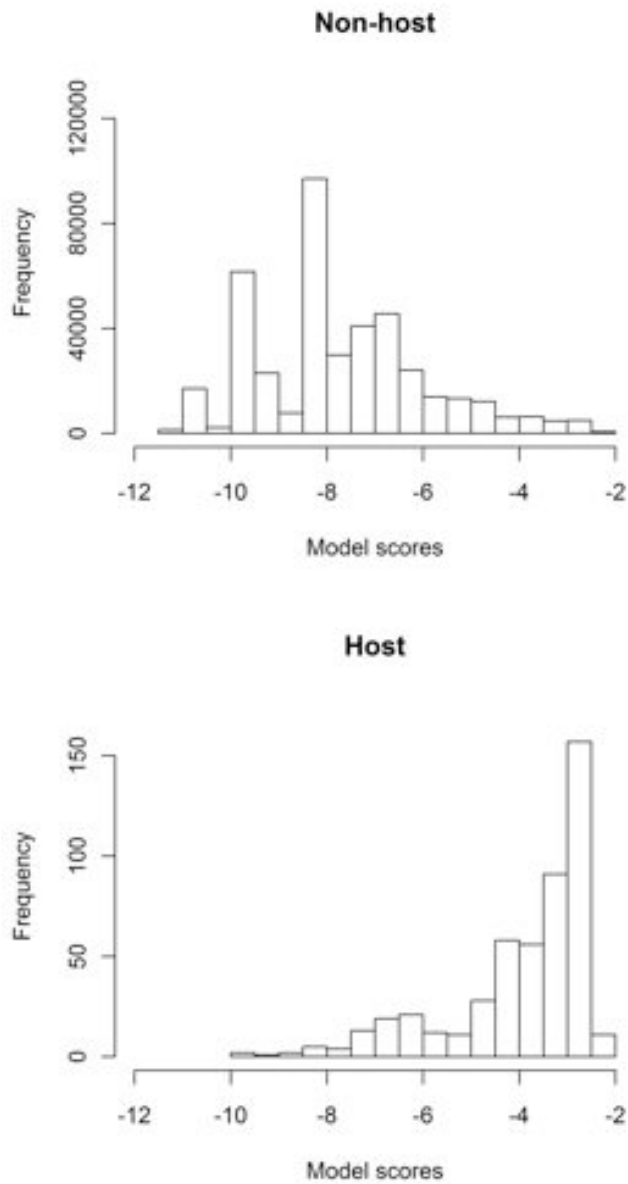
**Fig. S3: Histograms of model scores for herbivore-plant interactions.** A) Model scores for "non-hosts", i.e., non-interacting non-native plants and herbivores. B) Model scores for actual "host" interactions between non-native plants and herbivores. The median model score for actual "host" interactions more than 2-fold higher than the median score for actual non-realized interactions.
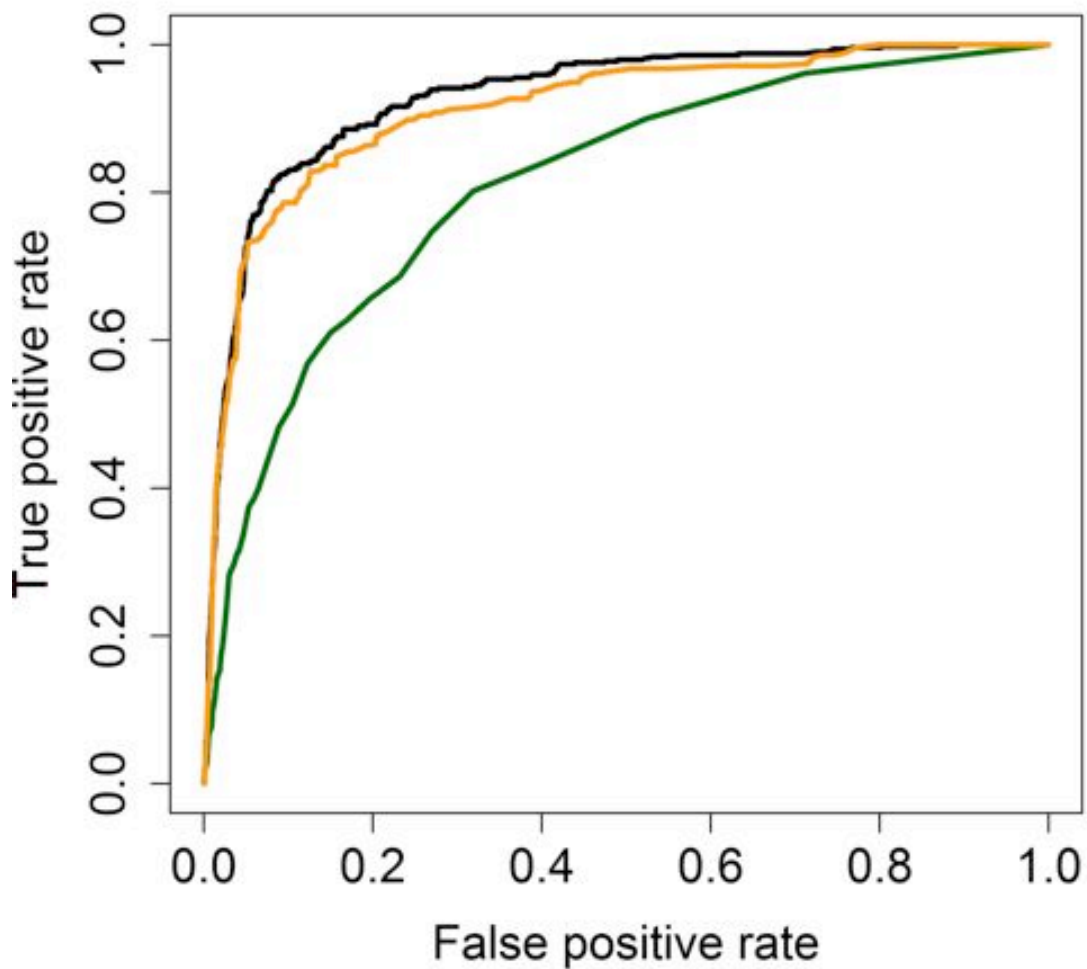
**Fig. S4: ROC curves of simplified models.** Lines show the model accuracy of the full novel interactions model incorporating host breadth and phylogenetic similarity (black; AUC= 0.930), a model incorporating only phylogenetic similarity (orange; AUC=0.914), and a model incorporating only host breadth (green; AUC=0.811). The true positive rate differed among models (Kolmogorov-Smirnov Tests; P<0.001). AUC - Area under the curve, which approaches one for a perfectly predictive model.
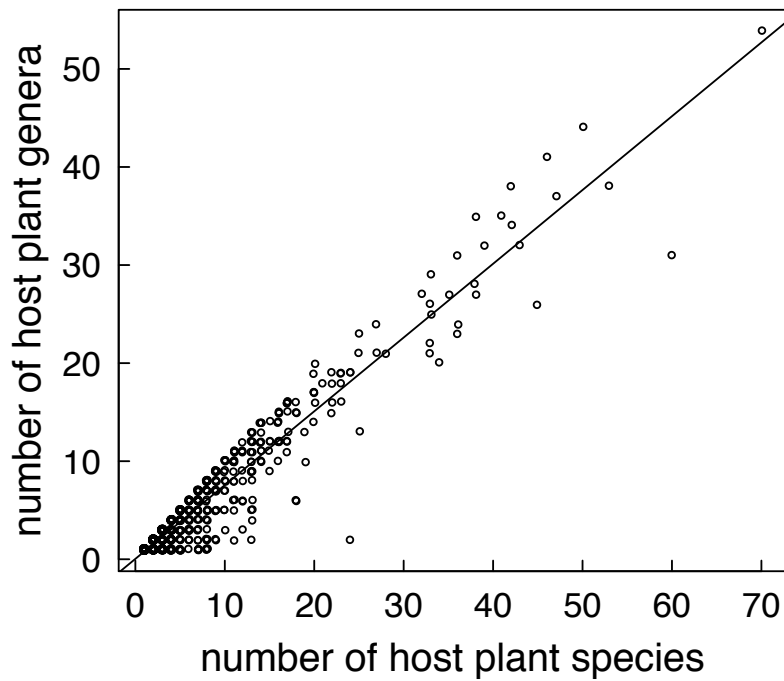
**Fig. S5: Host breadth based on plant species or genera.** A scatterplot of the number of host plant species versus the number of host plant genera for all moth species in our study. Values are slightly jittered on both x- and y- axis (by 0.5). The line is a highly significant linear regression fit (p<0.0001). Because of the strong correlation between host plant species the definition of host breadth as the number of plant species consumed or plant genera did not affect the ability of the host-use model to predict novel hosts (species-based definition: AUC = 0.930, genus-based definition: AUC = 0.932).
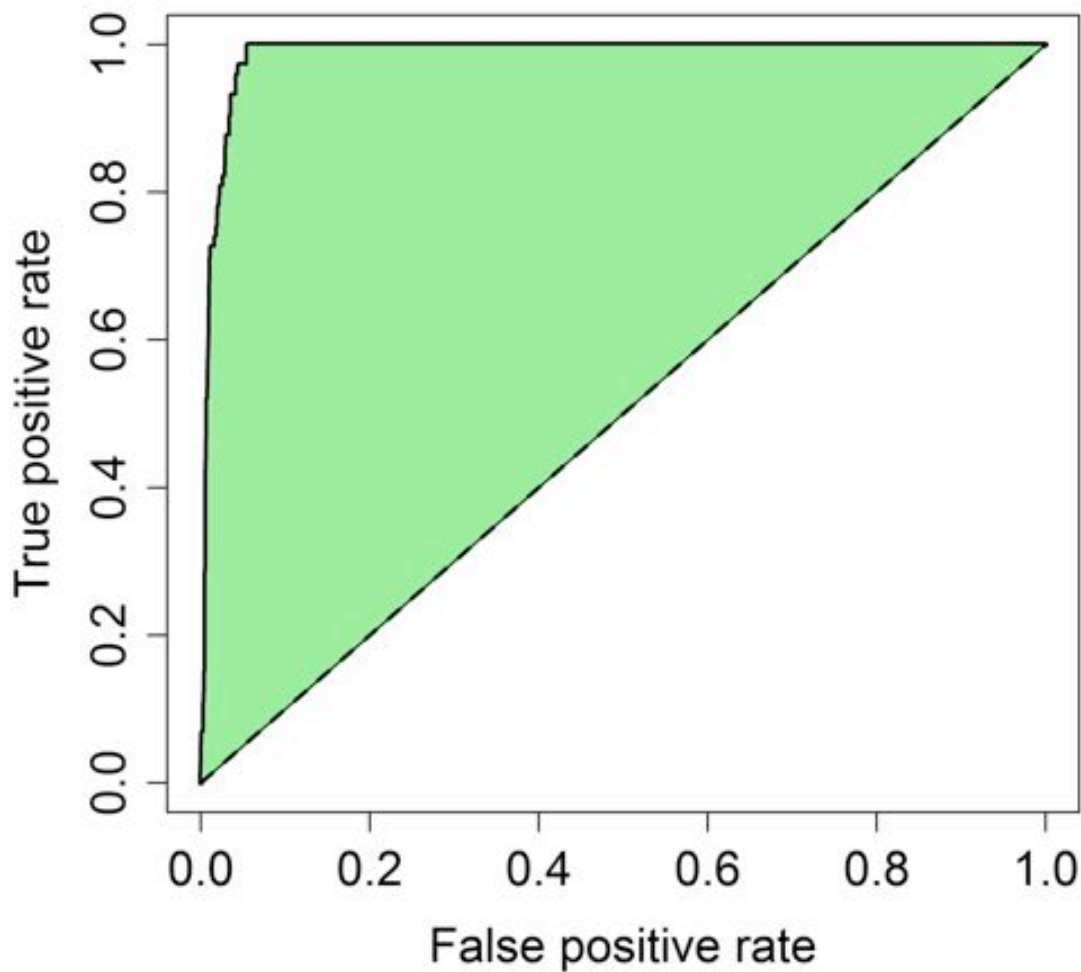
**Fig. S6: Butterfly-only ROC curve.** As butterfly-host associations are the best-documented herbivore-plant interactions, we restricted our herbivore dataset to only the 126 butterfly species and ran an analysis identical to that with the full dataset (Figure 1). The novel host predictions for the butterfly-only dataset were even better than for the full dataset, where we predicted all observed novel butterfly-plant interactions with only a 7% false-positive rate. The dashed line indicates the null expectations from an uninformative model, and the shaded, green area represents the increased predictive ability beyond that null model.
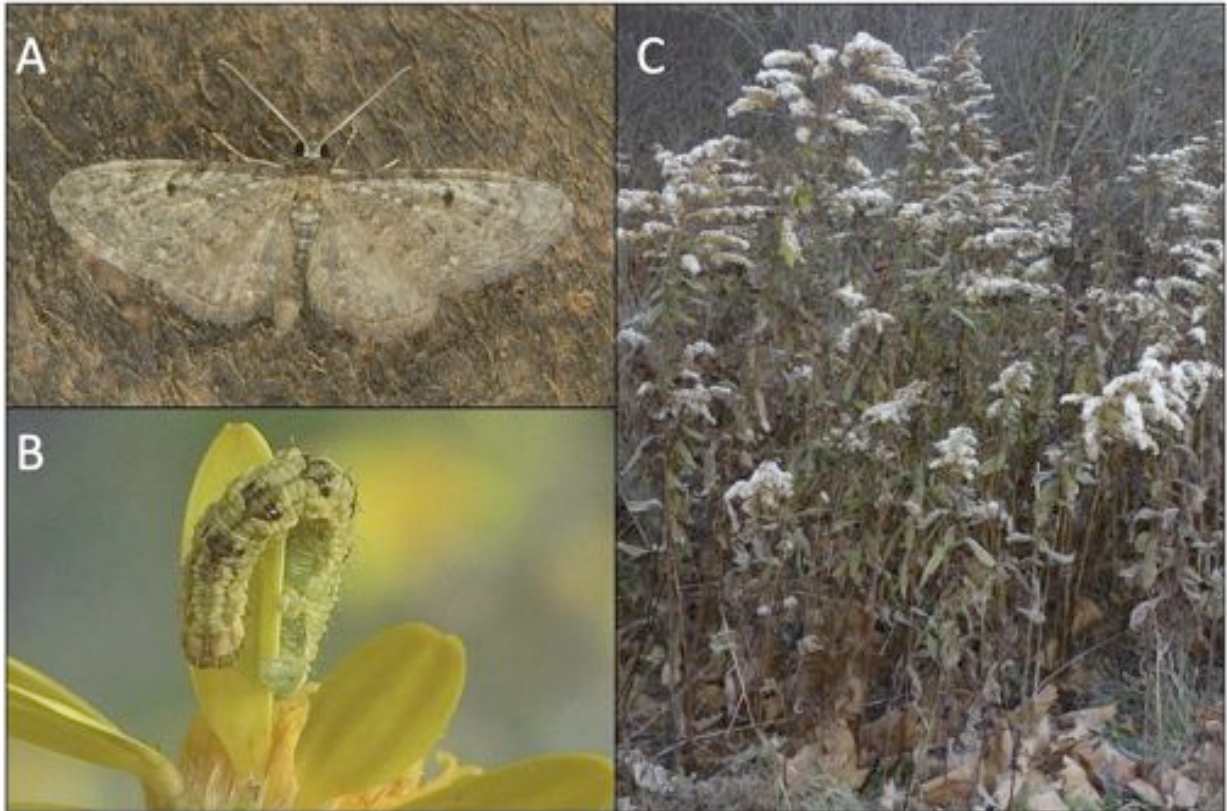
**Fig. S7: Native herbivores (A,B) feed on goldenrod (C), an invasive plant throughout Europe.** A) adult moth of *Eupithecia virgaureata* (Foto H. Melzer, Lepiforum), B) larva of *Eupithecia absinthiata* (Foto F. Nantscheff, Lepiforum), and C) senescent goldenrod plants with seed heads (*Solidago canadensis*; Foto F. Altermatt). In Central Europe the geometrid moths *Eupithecia virgaureata* and *E. absinthiata* feed (among others) on various native mugworts and ragworts (*Artemisia* and *Senecio* sp.). Over the last decades they extended their diet onto invasive goldenrods (*Solidago canadensis* and *S. gigantea*) (*25*). We predicted this host-shift with a false-positive rate of only 0.08%. As the larvae of *Eupithecia* moths selectively feed on flowers and developing seeds, they may have a negative fitness effect on invasive *Solidago* sp. With our data we unfortunately cannot draw conclusions on temporal dynamics, and it is up to now unclear how effective host-switches are and how much they affect the non-native plants. However, the diet switch to a non-native plant is the first and probably most important step to use that plant more extensively. It is basically putting "the foot in the

door", which then may open up the possibility for ecological and evolutionary processes to use that novel plant more extensively.

**Table S1: Model parameters** from data partitions (K=5) of the native plant-herbivore data (see STEP 1 in Fig. 1) as estimated using a generalized linear model where phylogenetic distance to a "native" (i.e., "predictor dataset") plant, herbivore host breadth, and the interaction of these terms predicted the binomial response of host-use. The averaged model parameters were used in generating predictions of non-native host use (STEP 2 in Fig. 1).

| K | Phylo-similarity to "Native" Plant | Herbivore host breadth | Interaction: Phylo Sim. x Host breadth |
|---|---|---|---|
| 1 | -8.59101 | -0.01256 | 0.144882 |
| 2 | -9.1036 | -0.00619 | 0.115433 |
| 3 | -9.11003 | -0.00835 | 0.142009 |
| 4 | -10.5968 | -0.00000 | 0.09207 |
| 5 | -9.02491 | 0.001372 | 0.10116 |
| | | | |
| **Ave** | -9.28528 | -0.00515 | 0.119111 |
| **SE** | 0.341594 | 0.002601 | 0.010619 |