Varyingly Hungry Caterpillars: Predictive Models and Foliar Chemistry Suggest How to Eat a Rainforest

Simon T Segar[1,2,*], Martin Volf[1,2], Brus Isua[3], Mentap Sisol[3], Conor M Redmond[1,2], Margaret E Rosati[4], Bradley Gewa[3], Kenneth Molem[3], Chris Dahl[1,2], Jeremy D Holloway[5], Yves Basset[1,2,6], Scott E Miller[4], George Weiblen[7], Juha-Pekka Salminen[8] and Vojtech Novotny[1,2]

[1] University of South Bohemia in Ceske Budejovice, Faculty of Science, Branisovska 1760, 370 05 Ceske Budejovice, Czech Republic. [2]Biology Centre, The Czech Academy of Sciences, Branisovska 31, 370 05 Ceske Budejovice, Czech Republic. [3]New Guinea Binatang Research Center, PO Box 604 Madang, Papua New Guinea. [4]National Museum of Natural History, Smithsonian Institution, Box 37012, Washington, DC 20013-7012, USA. [5]Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, U.K. [6]Smithsonian Tropical Research Institute, Apartado 0843-03092, Panama City, Republic of Panama.[7]Bell Museum of Natural History and Department of Plant Biology, University of Minnesota, 1445 Gortner Avenue, Saint Paul, Minnesota 55108-1095, USA. [8]University of Turku, Department of Chemistry, Vatselankatu 2, FI-20500 Turku, Finland.

* Corresponding Author: Email: simon.t.segar@gmail.com

Abstract

A long-term goal in evolutionary ecology is to explain the incredible diversity of insect herbivores and patterns of plant host use in speciose groups like tropical Lepidoptera. Here we used standardised food-web data, multigene phylogenies of both trophic levels and plant chemistry data to model interactions between Lepidoptera larvae (caterpillars) from two lineages (Geometridae and Pyraloidea) and plants in species-rich lowland rainforest in New Guinea. Model parameters were used to make and test blind predictions for two hectares of exhaustively sampled forest. For pyraloids we relied on phylogeny alone and predicted 54% of species level interactions, translating to 79% of all trophic links for individual insects, by sampling insects from only 15% of local woody plant diversity. The phylogenetic distribution of host plant associations in polyphagous geometrids was less conserved, reducing accuracy. In a truly quantitative food-web only 40% of pair-wise interactions were described correctly in geometrids. Polyphenol oxidative activity (but not protein precipitation capacity), was important for understanding the occurrence of geometrids (but not pyraloids) across their hosts. When both foliar chemistry and plant phylogeny were included, we predicted geometrid-plant occurrence with 89% concordance. Such models help to test macroevolutionary hypotheses at the community level.

Keywords: biodiversity, food-webs, Geometridae, oxidative activity, Papua New Guinea, Pyraloidea.

Introduction

Plants, insect herbivores and their insect natural enemies constitute over 75% of all species on earth [1]; all three are crucial for ecosystem functioning. For decades researchers have sought to explain the incredible diversity of insects on plants, from both evolutionary and ecological perspectives [2,3]. Specifically, the relative contribution to diversification made by i) reciprocal adaptation between plants and insects versus ii) sequential invasions of plant lineages by insects has been a source of much debate [4]. Is global insect diversity an emergent property of inter-species interactions between trophic levels, and selection imposed by insects, or simply a function of plant diversity and multiple changes in resource use? Novel plant traits open up new herbivore-free ‘adaptive zones’. Escalation in toxicity is suggestive of an evolutionary arms race [2,5], indeed some specialist herbivores eventually sequester or overcome these defences and diversify [6,7]. In contrast, multiple sequential invasions by insect lineages across different plant clades may have led to high levels of insect diversification independently of plant diversification [8,9]. Understanding how phylogenetic insect herbivore-plant interaction networks are structured is key to distinguishing between these two processes [8]. Two of the main questions in the field of insect-plant interactions are therefore how and why herbivores vary in their host-specificity and phylogenetic host range [10]. Answering these questions will further our understanding of species richness and pest occurrence; and may allow us to hypothesise how novel networks are structured [11].

The evolutionary history of many herbivorous insect groups consists of frequent minor hosts shifts; 90% of herbivores use no more than three plant families [12]. While major host shifts occur less frequently, they can open up new adaptive zones [8,13]. Central to testing hypotheses of diversification is the assessment of factors determining the variation and taxonomic scale of insect host specificity, and the suitability of novel plant lineages as hosts [14,15]. Studies of host-plant relationships have shown that plant phylogeny, or taxonomy, can predict insect assemblage structure and diversity to a limited extent [16,17]. However, shared evolutionary history of host plants is only a partial explanation for dietary range

[18]. A predictive phylogenetic framework considering two trophic levels along with plant traits is necessary to generate baseline expectations for host use. Such a framework has, so far, been lacking in most studies of host specificity [16,19]. Here we focus on communities, with a rich macroevolutionary literature allowing us to formulate some expectations relevant to this level.

There is, as envisaged by Ehrlich & Raven [2], an arms race between herbivorous insects and plants which drives co-adaptation and patterns of host use through 'escape and radiate' diversification. Recent evidence adds support to this hypothesis; co-adaptation between insects and plants has been demonstrated [6,7], and selection for defensive trait escalation exists [20]. One clear expectation from this 'coevolutionary framework' [3] is a high level of herbivore community similarity and host specificity at the plant genus level [14]. Phylogenetic signal in both plant traits and insect host use is driven by co-adaptation, and should lead to predictable network structure. Convergence in chemistry between plant clades will drive long distance host shifts and lead to adaptive radiations. The support for some of the predictions made by Ehrlich & Raven [2] has been mixed, with clear cases of sequential radiation and phylogenetic tracking [21] being more prevalent than co-diversification [4,9]. Two additional scenarios have been proposed: i) the 'oscillating radiation' and ii) 'musical chairs' hypotheses. Oscillating radiation [8] involves repeated yet phylogenetically scattered shifts by derived generalist insect species to specialise on new host plants, after an initial expansion of host range. The ability to use ancestral hosts is retained. Under this scenario long distance host shifts will structure communities, although insects will use both derived and ancestral host 'nodes'. This should lead to highest phylogenetic signal through the host network. Such dynamics have been reported for the butterfly family Nymphalidae [22], and more recently in polyphagous lymantriine moths [23]. In turn, the generality of the oscillation hypothesis has also been called into question [24], and the 'musical chairs' hypothesis relaxes the assumption that shifting species undergo changes in niche breadth or stem from generalist taxa [but see 25]. Shifts are continuous and within specialist clades, being phylogenetically local in their scale. In this context, we expect to detect many short-range host shifts within communities, allowing for accurate predictions of host use and high phylogenetic signal in the herbivore network.

In this study, we demonstrate how simultaneously considering the evolutionary history of hosts and herbivores, alongside host traits that may vary partially independently from host phylogeny, can provide novel and perhaps unexpected insights into how both groups have diversified in complex natural systems. We focus on insect-herbivore food-webs in natural forest in Papua New Guinea, making and testing predictions of network structure across two hectares of exhaustively sampled forest. Furthermore, we untangle the contribution of shared evolutionary history and plant defensive traits by including data on the major defensive attributes of plant polyphenols.

Polyphenolics are fantastically diverse and phylogenetically widespread compounds. Polyphenols have been implicated in insect herbivore defence, but their mode of operation and exact role is not always clear, especially as measures of total content without detailed compositional or activity data are usually not sufficient [26,27]. Therefore, we have included two types of defensive activities connected with tannins and other polyphenols: oxidative activity shows how easily polyphenols are oxidized in the alkaline gut of insect herbivores thus causing oxidative stress; and protein precipitation capacity shows how well polyphenols, especially tannins, may bind with dietary proteins before entering the alkaline gut regions of insects, thus making their diet less nutritive [27].

We aim to test the predictive power of phylogenetic models and plant traits to detect network structure, with the added expectation that such a predictive approach can distinguish between underlying evolutionary processes, providing support for some of the hypotheses outlined above. We expect that incorporating data on one of the most widespread groups of plant defensive compounds will help us to detect convergence in trait space across hosts, improving our predictions for insect species with broad host use. In comparison, insect phylogenetic relationships may be a better predictor of occurrence for more phylogenetically specialised or more highly coevolved insect lineages. The relative contribution of host and herbivore phylogenies, as well as the predictive power of plant chemistry, is therefore tested for two families of caterpillar with contrasting host use patterns.

Methods

Sampling Insects

We sampled all caterpillars from a locally representative selection of 88 host plants in a 10 x 20 km area matrix of primary and secondary lowland rainforest in Madang province, Papua New Guinea. We refer to this standardised data set as the 'Madang' data set. This selection of plants reflected the local diversity of vegetation, and focused on three families that are locally species-rich (Moraceae, Euphorbiaceae and Rubiaceae) as well as a selection of 28 plant families represented by one or more species. Our sampling was standardised across all host trees and is described elsewhere in detail [14,28] and here in Appendix 1. We focused our analyses on the food-webs of feeding individuals from species in two ecologically dominant lineages: Geometridae and Pyraloidea (Lepidoptera). In combination they comprise 50% of all caterpillar individuals sampled by Novotny *et al.* [14], placing them among the most species-rich caterpillar lineages sampled from the PNG flora. All insects were identified to morpho-species, whilst a subset was barcoded to confirm species boundaries [14]. We also destructively sampled two hectares of lowland rainforest (one ha primary and one ha secondary) around Wanang village (75 km from the Madang sampling area), this forest was contiguous with the Madang sampling area until the onset of commercial logging in 2005, and both sites share many plant and Lepidoptera species [29]. All trees >5cm DBH were felled in order of size, felled trees and the surrounding area were immediately searched for caterpillars by teams of 15 local assistants; for detailed methods see [30,31] and Appendix 1. The forest in Wanang was sampled for caterpillars in accordance to local host abundance so that every tree in the two one hectare plots was sampled exhaustively for caterpillars, regardless of size. We refer to this expanded data set as the 'Wanang' data set, and refer to interactions found only in the Wanang data set as additional, e.g. involving a host or insect not sampled in the Madang data set. While the Madang data set represents the most commonly available data format [32], comprising a phylogenetically stratified selection of plant species sampled with a uniform sampling effort, the advantage of the Wanang data set is that it captures ‘all’ local interactions; it is a truly quantitative food-web. We used the Madang data set to make (and test) blind predictions of the host plant associations of insects from the Wanang felled plot data set. Overall, our comparison provides an excellent test of our ability to scale up our models, calibrated using selective and standardised collections, to continuous areas of natural forest. However, our approach has the advantage of using the most commonly available type of data set, represented by the Madang data, to make predictions for the Wanang data, that most accurately represent local food-webs.

Estimating Phylogenies

We generated multigene molecular phylogenies for both hosts and caterpillar herbivores. For hosts we used sequence data generated in previous studies [14,30] . We estimated caterpillar phylogenies by integrating existing DNA barcode (COI) data and newly collected data from four nuclear genes (CAD, wingless, RpS5 and DDC) with extensive sequence data from published studies (Appendix 2). We used published primers and protocols [33] to sequence individuals with existing barcode sequences in BOLD to ensure that only individuals actually sampled from plant hosts were included in our study. Detailed phylogenetic methods are given in Appendix 1.

Quantifying Oxidative Activity and Protein Precipitation Capacity

Plant tissue was collected in the field over 251 days between 2013 and 2014, and we sampled leaf discs of 2.4 cm in diameter from ten young but fully expanded leaves per individual tree for between three to six individuals per species. Collections were made throughout the year to capture seasonal variation. We examined the influence of time between collections on variation around the mean, assessed the contribution of intra-and inter-specific variation using linear mixed models, and analysed replicate samples of highly variable species with UPLC-QqQ-MS/MS [34] (Appendix 1). All leaf discs were weighed fresh and stored in UPLC grade acetone at -20˚C before quantitative extraction and analysis at the University of Turku in Finland (Appendix 1). We quantified the portion of total phenolics that is easily auto-oxidized at the alkaline pH especially common to the midgut of Lepidopteran larvae. This oxidative activity was measured in both mg/g dry weight and in % of total phenolics derived from the Salminen & Karonen assay calibrated with gallic acid [27] for all 88 species in the Madang data set. To quantify the protein precipitation capacity of each species we used the radial diffusion assay [35], with BSA as the protein and calibrated with pentagalloyl glucose as the tannin. We acknowledge that while polyphenols are likely to play an important role in plant defence, they represent only one major group of plant secondary metabolites. Many plant families (particularly Moraceae, Rubiaceae and Euphorbiaceae) contain other types of secondary metabolites (e.g. alkaloids and terpenes) that can be toxic to insect herbivores or sequestered as chemical defences (Volf et al., in revision).

Statistics: Foliar Chemistry and Unipartite Phylogenies

We calculated the phylogenetic signal (Pagel's Lambda) of both chemical variables across the 88 plant species surveyed in the Madang data set. The power of the chemical variables to predict host associations was tested with binary logistic regression, using generalised linear phylogenetic models [36] to account for phylogenetic non-independence. Explanatory variables included both chemical activity measures. We dissected the effects of traits and phylogeny further, using phylogenetic eigenvector regression (PVR). Our main aim was to capture the higher-level bifurcating structure of the host phylogeny. While PVR has flaws [37], it is informative when used alongside model based approaches. We decomposed a patristic distance matrix of the host phylogeny using principal coordinates analysis (PCoA); all eigenvectors were positive so no correction was applied. We used the first 10 eigenvectors to explain over 80% of the variance in phylogenetic structure. These eigenvectors were first included in quasibinomial regressions, without chemical variables. We selected significant explanatory vectors according to quasi-AICc, specifying the dispersion parameter for each model [38]. We subsequently ran simplified models including the chemical variables. We used the predicted probabilities of occurrence to calculate the accuracy of our models using a decision boundary of 0.5 and plotted the true positive rate against the false positive rate to generate a receiver operating characteristic (ROC) curve; the area under the curve (AUC) was used to assess the predictive power of our models.

Statistics: Modelling Host Use Using Bipartite Phylogenies

Our analytical workflow is presented graphically in Figure S1. We used the Phylogenetic Bipartite Linear models (PBLM) of Ives & Godfray [39] as implemented in the R package 'Picante' [40] to assess the phylogenetic signal through each level of our food-webs and predict trophic interactions between hosts and insects in our standardised food-webs. Bipartite phylogenies refer to matching pairs of phylogenies from two trophic levels, which can be combined with a matrix of host use data. The PBLM models of Ives & Godfray [39] allow the inclusion of covariates associated with one or both interactants. We included both host oxidative activity and protein precipitation capacity as covariates and ran additional PBLM models for the Madang data, assessing model fit using the reduction in mean square error (MSE). In all models we excluded singleton interactions, e.g. any matrix entry of one, and square-root transformed all abundance data. We ran PBLMs with phylogenetic correlation for both lineages of caterpillars separately to estimate the phylogenetic signal through the host (d_h) and herbivore (d_p) matrices. We used 1,000 iterations for the 'optim' procedure and 100 bootstrap replicates.

We then validated our models by assessing the strength of the correlations between quantitative observed and predicted values following Ives & Godfray [39]. Furthermore, we directly compared observed and predicted networks based on their matrix fill. We filled the predicted network with the highest estimated interaction strengths for each association derived from our models (the rank probability of that interaction occurring), keeping the row and column sums equal to the observed network. Both observed and predicted networks were transformed into binary (presence-absence) networks. This essentially retains only the most probable interactions in accordance to the observed matrix sums. To compare the structure of these networks, we generated matching distance matrices based on Euclidean distances and correlated these distance matrices using a Mantel test with Pearson's product moment correlation.

We also calculated the proportion of ‘direct hits’ (exact predictions) for moth groups and compared this value to a series of increasingly constrained null distributions generated through 1,000 randomisations using the C0 randomisation in the R package ‘Vegan’ [41]. The first null distribution (total randomisation) was generated from the observed data by keeping only column totals (moth species) constant, the second respected the rows (host species) as grouped into major plant clades (monophyletic groups of similar ages, Figure S2) but allowed column fill to be randomised, the third constrained interactions at the plant family level and the fourth constrained interactions at the plant genus level. We compared the observed proportion of exact hits and the mean as generated under total randomisation and the mean arising from the most similar taxonomic randomisation. We also assessed how our models worked at various taxonomic levels by calculating the proportion of interactions they predicted at the clade (Figure S2) and family levels. The significance of any difference between the modelled ‘direct hit’ values and the mean values obtained from randomised distributions (hereafter ‘random mean’) was tested using two-tailed tests.

Finally, we tested the ability of our models to predict the host plant use of twenty locally abundant caterpillars in the Wanang felled plots, representing the ten most abundant geometrids and ten most abundant pyraloids. This was done by using the covariance matrices and phylogenetically corrected means of association strength generated in our PBLMs after adding the additional hosts sampled for caterpillars in Wanang to our phylogeny (using rbcl sequence data and constraints). It was then possible to extrapolate host use data onto our expanded host set in a second round of predictive models [39].

Results

Phylogenies

We obtained mitochondrial COI barcode data from all specimens included and nuclear DNA sequence data for 80% of non-singleton taxa (Appendix 2). Our alignments included up to 2.9kb of mitochondrial and nuclear sequence data. After the removal of singleton matrix entries and missing data we retained 43 geometrids and 63 pyraloids in our Madang data set. Please refer to Figures S2, S3 and S4 for labelled molecular phylogenies of both plants and insects. Justifications of taxonomic changes indicated by molecular and morphological data are found in Appendix 3, including the recognition of *Syllepte planeflava* Hampson as a new synonym of *Eusabena paraphragma* (Meyrick) (Crambidae).

Using Trait Data for Predictions

Neither of the chemical traits that we measured showed significant increases in variation around the mean with time between collection while there was more variation between than within species (Appendix 1; Figures S5 and S6). We found moderate phylogenetic signal in trait values (oxidative activity: Lambda=0.698, p=0.001, protein precipitation capacity: Lambda=0.703. p=0.001).

The probability of geometrid occurrence was predicted well by both oxidative activity (21% of the deviance) and host phylogeny (22% of the deviance) whilst protein-precipitation explained a smaller non-significant proportion of the deviance (<1%) (Table 1 and Figure S6). When included in separate PVR models alongside phylogeny, oxidative activity had a significantly steeper slope (0.237) than protein precipitation capacity (0.044) (p=0.027). The model including oxidative activity and phylogeny had an accuracy of 81% and an AUC of 89%. There was a strong significant positive relationship between geometrid occurrence and oxidative activity. While moderate oxidative activity predicted geometrid occurrence, phylogeny predicted occurrence on low activity hosts (Figure S7). The significant phylogenetic axes separated i) magnoliids and monocots from higher angiosperms, ii) Malpighiales from Rosales and Fabales, and iii) Malvids from the rest. Pyraloid occurrence was not significantly related to either oxidative activity (2% of the deviance) or protein precipitation capacity (<1% of the deviance) (Table 1).

Phylogenetic Linear Bipartite Models

Madang Predictions

For both geometrids and pyraloids PBLM model fit was best when phylogeny was included. The mean square error (MSE) was lower for the models incorporating our phylogenetic estimates than those that used star phylogenies (total polytomies) for both trophic levels (geometrids: MSE_{Data}=1.23, MSE_{Star}=1.65, $MSE_{Brownian}$=4.85, pyraloids: MSE_{Data}=1.52, MSE_{Star}=1.77, $MSE_{Brownian}$=3.59), but including oxidative activity as a covariate improved model predictions only for geometrids. For both families of caterpillar, we also estimated the strength of phylogenetic signal through both the host (d_h) and herbivore (d_p) phylogenies. Significantly non-zero values of d_h indicate related host-plant species being eaten by the same herbivores (although the herbivores themselves need not be related). While high values of d_p indicate related herbivore species eating the same hosts. We found contrasting results for each family, with phylogenetic signal being clearly non-zero [42] and stronger through the host level (d_h=0.15, 95% CI: 0.01-0.44) than the herbivore level (d_p<0.01, 0.001-0.26) in geometrids, and weaker through the host level (d_h=0.05, <0.001-0.27) than the herbivore level in pyraloids (d_p=0.20, 0.03-0.45). This indicated that host plant phylogeny best predicted geometrid host use, while insect phylogeny best predicted pyraloid host use. Closely related geometrids can use distantly related hosts (which are clustered in several 'islands' across the host phylogeny). Host use between closely related insects is more conserved in pyraloids, with clades of insects utilising the same subset of plant hosts. Several pyraloid clades are even restricted to related plant species. The interaction networks between hosts and herbivores can be visualized as co-phylogenetic plots (Figure 1). Phylogenetic signal of host use is not always deep, but clearly relevant for many clades.

The mean correlation between observed and predicted values was 0.24 for pyraloids and 0.32 for geometrids. For both geometrids (r=0.72, p=0.001), and pyraloids (r=0.94, p=0.001) there was a significant positive correlation between observed and predicted matrix structure (Figure S9). Our phylogenetic models predicted 24% of all 80 interactions between geometrids and their hosts exactly (random mean=11%, p<0.001) and 45% of all 191 interactions between pyraloids and their hosts (random mean=12%, p<0.001), in both cases the predictive power was significantly better than random. We predicted the correct plant clade for 50% of geometrid interactions and 71% of pyraloid interactions (Figure 2).

Wanang Predictions

We predicted host use in the ten most abundant pyraloid and geometrid caterpillar species surveyed in the exhaustively sampled Wanang felled plots. These 10 species represented 3,122 (60%) of the 5,199 pyraloid caterpillars and 620 (48%) geometrid caterpillars sampled at Wanang (excluding singleton interactions in both cases). Of these 20 species, 18 were recorded in our phylogenetically standardised survey; one geometrid (*Idiochlora celataria*) and one pyraloid (*Paraphomia disjuncta*) were added into the phylogenies for the Wanang analysis. The Wanang caterpillar host use data are deposited in Dryad. In Wanang, we sampled these pyraloids from 19 additional plant species that were not sampled in our standardised survey (an expanded set of hosts). Our models predicted 54% of additional interactions with this expanded set of host plants (mean 'direct hits' under total randomisation of the matrix=20%, p<0.001). For geometrids, we sampled 27 additional hosts in Wanang and predicted 40% of additional interactions (random mean=21%, p<0.001). It is worth noting that our predictive models performed well in terms of predicting the major interactions in the data set (e.g. the strongest links in the network). Host use across the expanded set of hosts was correctly predicted for 79% of all pyraloid individuals (1,901 out of 2,430 interactions across 19 additional plants with 619 individuals maintaining the same host species as in the standardised data set) and 53% of all geometrid individuals (192 out of 360 interactions across 27 additional plants with 130 individuals maintaining the same host species as in the standardised data set).

Discussion

We explored the evolution of insect herbivore diversity and the pervasive nature of phylogenetic constraints and/or plant traits in host use by using predictive models to explore food-webs. This was done by studying two lineages of caterpillars (Geometridae and Pyraloidea) across 122 plant species in a lowland tropical rainforest. We sought to understand cases of predictive power and breakdown in the context of existing hypotheses aimed at explaining evolutionary diversification. For geometrids, we demonstrated that both host phylogeny and foliar polyphenol chemistry were reasonable predictors of host use, acting in a complementary manner to predict suitable hosts, suggesting an evolutionary history of host shifting. In contrast pyraloids generally responded less strongly to oxidative activity, and neither group responded strongly to protein precipitation capacity. Pyraloid phylogeny itself was a good predictor of host use, indicating phylogenetic constraints at the herbivore level and an evolutionary history potentially more dominated by limited host shifts and/or co-diversification. Including additional measures of chemical diversity and activity is key to fully understanding how extant community structure is related to hypotheses of diversification. We argue that a community approach can complement more focused macroevolutionary studies.

Variation in host range between herbivorous insects has been subject to intensive study from both evolutionary and ecological perspectives and at multiple taxonomic scales, from families to populations [18,43]. Of particular interest are clades of herbivores that exhibit variation in host specificity [44] and/or phylogenetic lability of host use [45] as these may include radiating lineages. These lineages play a key role in most evolutionary hypotheses regarding insect diversification. Here we argue that baseline phylogenetic expectations are needed to formulate further hypotheses about the proximate and ultimate reasons for such variation in host use patterns. Detailed data on plant defensive traits (e.g. secondary metabolites) and ecological interactions (e.g. between parasitoids and bacterial gut symbionts) might improve the explanatory power of such models considerably, although the key covariates (and levels of host specificity) may vary between herbivore guilds. A systematic approach is required to fully understand the evolution of herbivore host use.

When we consider broad phylogenetic patterns of host use we can see that, overall, incorrect predictions are generated more frequently for geometrids than pyraloids. This is largely because of lower phylogenetic signal through the geometrid phylogeny. However, combining traits and phylogeny can improve predictions substantially when considering the lower trophic level for geometrid interactions, and mismatches in observations and predictions can help shape our understanding of insect herbivore host use. Network mismatches in geometrid host use often involve missed or incorrect assignment of moth species to members of the plant families Euphorbiaceae and Myrtaceae. Our community sample for geometrids comprised five subfamilies, and almost all of them utilized Myrtaceae along with a wide selection of other plant families. Members of the subfamily Geometrinae commonly used Euphorbiaceae and/or Phyllanthcaeae and Myrtaceae as core hosts, while several species of Ennominae used a phylogenetically diverse set of hosts. Many species can make long distance host shifts that are harder to predict based on phylogeny alone. Members of the Ennominae tribe Boarmiini often use a phylogenetically broad set of hosts and Robinson *et al.* [46] list records from 28 families for *Ectropis bhurmitra* (Appendix 4). Within Larentiinae there are also many polyphagous species [47]. This suggests that even ‘super generalists’ may have preferred (core) hosts that represent islands from which host expansion can proceed, perhaps these are ancestral hosts. It also suggests that exposure to several hosts may build up the necessary metabolic mechanisms that broaden host range [48] so that host shifts beget host shifts. Overall, there is limited evidence here for clades of geometrids radiating across plant lineages. Indeed, a pattern of convergence onto key plant nodes and phylogenetic over-dispersion is more in line with diversification processes involving multiple long distance radiations [8] as predicted by the ‘oscillating radiation’ hypothesis. These results concur with recent findings in Lymantriinae [23]. The power of convergent plant defences as predictors is also in line with a relaxed interpretation of the ‘escape and radiate’ hypothesis, but there is a distinct disparity in phylogenetic signal between networks. It is also possible that particularly polyphagous species occur within Geometridae as phylogenetically apical taxa capable of colonizing new hosts [22], requiring consideration of the micro-taxonomic scale.

Some moths consume tannin-rich foliage. It has been shown that C-glycosidic ellagitannins – with the highest oxidative activity of all polyphenol classes – can be found in both Myrtales and Fagales [49]. Similarly, oligomeric ellagitannins – with slightly lower oxidative activity – are found in Myrtales, Fagales and Rosales. We show here that the leaves of Euphorbiaceae and Myrtaceae have relatively high oxidative activity, and that there is a positive relationship between geometrid occurrence and oxidative activity. Many plant families found to have high oxidative activity in this study and elsewhere [49] are major geometrid hosts (Appendix 4). In fact, many outbreak geometrid species are associated with plant genera rich in ellagitannins that are primarily responsible for oxidative activity [50], e.g. *Betula* [51] and *Quercus* [52,53]. In contrast to the proanthocyanidins (syn. condensed tannins) that accumulate in mature leaf tissue and are less actively oxidized at high pH (as found in caterpillar guts), ellagitannins are often richer in the younger foliage utilised by geometrid moths [52,53]. Nevertheless, it has been shown that some moths are well adapted to consume tannin-rich foliage and, even though oxidative damage may occur in the midgut, it is not necessarily sufficient to increase the resistance of trees to tannin-tolerant caterpillars [54,55]. It is even possible that a net benefit exists, if ellagitannins and their oxidative activity contribute to increased resistance of larvae against pathogens and parasites for example [56,57].

In contrast to geometrids, host family level associations are more predictable for clades of pyraloids with closely related moths using the same clades of plants. Many pyraloids are from the subfamily Spilomelinae, which may represent a more recent radiation that has been through fewer 'oscillating radiations' (e.g. fewer host shifts overall) in the terms of Janz & Nylin [8] and retained more phylogenetic signal in terms of host use. Certainly, the relative age of the Spilomelinae clade that we sampled (node height=0.256) is considerably younger than for Geometridae (node height=0.507). Host use in pyraloids does not relate directly to host phylogenetic distance but instead, insect phylogeny (specialisation and radiation of insect clades), suggesting a role for co-adaptation rather than phylogenetic resource similarity. Perhaps short distance host shifts are more important in accessing new resources for pyraloids, in line with the ‘musical chairs’ hypothesis. Intriguingly, semi-concealed pyraloids do not appear to accumulate across plants with either low or high polyphenolic activity at this phylogenetic scale.

Our host records largely overlap with documented host use records (Appendix 4), making our results relevant to both wider lepidopteran evolutionary ecology and pest species. Although the generic classification of some moths is difficult and not always available in the pest control literature [58], a combination of molecular phylogenetic data and natural host use data may help predict novel hosts with economic importance. This is especially true in Pyraloidea where host preference seems more phylogenetically conserved, and where more pest species are found. For example, we predicted that *Parotis* sp. AAC8820 (near *marginata*) fed on the previously unsampled Wanang hosts *Tabernaemontana andacaqui*, *Alstonia scholaris* (Apocynaceae) and *Uncaria appendiculata* (Rubiaceae), while we accurately predicted that *Glyphodes* sp. AAD1816 (near *stolalis*) fed on *Ficus* (Moraceae). Indeed, *Glyphodes stolalis* is known to infest *Ficus* trees while *Parotis marginata* is a pest of *Alstonia scholaris*. Studying the host use of entire communities of insects in their natural habitats can provide insights into their potential as pests, and is therefore useful for applied branches of biology.

We sampled 88 species of woody plant from a local species pool of around 600 [59]. We used bipartite models and host records to predict 40-54% of interactions across an expanded set of 34 locally dominant plant species sampled according to abundance (53-79% of all trophic links at the individual level). In terms of predicting the strongest links our models performed well, but sometimes missed weak links. In tropical regions it is unlikely that entire floras will ever be sampled completely for insects, making predictive models of host use important for estimating diversity. In temperate regions it has been possible to achieve higher predictive power, but it was necessary to sample thousands of hosts, millions of caterpillars and almost two thirds of the local vascular plant flora [11].

Our model framework allowed us to compare phylogenetic conservatism between two ecologically dominant Lepidoptera lineages. Our results give insights into the evolutionary hypotheses of host use, the contribution of conserved and labile traits and the evolution of polyphagy both between and within [43] species. It is unlikely that any of the evolutionary process put forward can explain all insect herbivore diversity, with inter-dependence between partners and constraints to host shifts being highly variable between clades. We suggest that detailed studies of proximate mechanisms would also give extra insights into how host use has evolved in herbivores, but suggest that this is best done in a bipartite phylogenetic context. Finally, we suggest that predictive models of trophic interactions represent an efficient way of testing our hypotheses.

Data availability

Sequence data available from GenBank and EMBL: accession numbers KY370871-KY370926 and LT674168-LT674424. BOLD dataset doi: 10.5883/DS-SEGAR16. All other data and code used are available from Dryad: doi:10.5061/dryad.8f5f3.

Competing financial interests

We have no competing interests.

Author contributions

STS conceived the study, collected the caterpillar sequence data and leaf tissue, performed the statistical analyses and wrote the first draft of the manuscript, MV collected the sequence data and helped write the manuscript, BI and MS helped identify host species and collect tissue, CMR and BG led the morphotyping of the Wanang specimens, MER managed the specimen collections and barcode data base, KM, CD and YB led the collection of caterpillar specimens in the field, JDH identified and clarified the morphological species of geometrid specimens and conducted the literature review, SEM led the barcoding and species delimitation of Lepidoptera, GW collected the plant sequence data and contributed to phylogeny estimation and VN helped conceive the study and led many aspects of the field work. All authors commented on a first draft of the manuscript and contributed substantially to the text.

Acknowledgments

We thank the assistants and taxonomists acknowledged in [14] and the staff of the New Guinea Binatang Research Centre in PNG. We thank Tom Fayle for analytical assistance. Many people helped with the chemical analyses, we especially thank Atte Tuominen, Jorma Kim, Suvi Vanhakylä, Annika Lehtimäki, Saara Nenonen and Marianna Manninen. STS and VN acknowledge funding from and a Univ. of South Bohemia Postdoc project (reg.no. CZ.1.07/2.3.00/30.0006) and the Grant Agency of the Czech Republic (grant numbers 14-36098G and 15-24571S). VN acknowledges support from the ERC grant 669609. This material is based upon work supported by the U.S. National Science Foundation under grants DEB 9707928, 0211591 and 0515678, 0816749, and 0841885). Geometrid taxonomy was partially supported by US National Institutes of Health through ICBG 5UO1TW006671. DNA barcoding was provided by the Biodiversity Institute of Ontario, University of Guelph, with funding from Genome Canada and the Ontario Genomics Institute to the International Barcode of Life Project. JPS acknowledges funding from the Academy of Finland (grant no 258992). Access to computing and storage facilities provided by the National Grid Infrastructure MetaCentrum (CESNET LM2015042) is greatly appreciated.

References

1. Price PW. 2002 Resource-driven terrestrial interaction webs. *Ecol. Res.* **17**, 241–247.

2. Ehrlich PR, Raven PH. 1964 Butterflies and plants: a study in coevolution. *Evolution* **18**, 586–608.

3. Janz N. 2011 Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies. *Annu. Rev. Ecol. Syst.* **42**, 71–89.

4. Suchan T, Alvarez N. 2015 Fifty years after Ehrlich and Raven, is there support for plant-insect coevolution as a major driver of species diversification? *Entomol. Exp. Appl.* **157**, 98–112. (doi:10.1111/eea.12348)

5. Agrawal AA, Salminen J-P, Fishbein M. 2009 Phylogenetic Trends in Phenolic Metabolism of Milkweeds (*Asclepias*): Evidence for Escalation. *Evolution* **63**, 663–673. (doi:10.1111/j.1558-5646.2008.00573.x)

6. Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007 The genetic basis of a plant–insect coevolutionary key innovation. *Proc. Natl. Acad. Sci. USA* **104**, 20427–20431.

7. Edger PP *et al.* 2015 The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. USA* **112**, 8362–8366. (doi:10.1073/pnas.1503926112)

8. Janz N, Nylin S. 2008 The Oscillation Hypothesis of Host-Plant Range and Speciation. In *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects* (ed KJ Tilmon), pp. 203–215. Los Angeles, CA: University of California Press.

9. Endara M-J, Coley PD, Ghabash G, Nicholls JA, Dexter KG, Donoso DA, Stone GN, Pennington RT, Kursar TA. 2017 Coevolutionary arms race versus host defense chase in a tropical herbivore–plant system. *Proc. Natl. Acad. Sci. USA* **114**, E7499–E7505. (doi:10.1073/pnas.1707727114)

10. Lewinsohn TM, Novotny V, Basset Y. 2005 Insects on plants: diversity of herbivore assemblages revisited. *Annu. Rev. Ecol. Evol. S.* **36**, 597–620. (doi:10.1146/annurev.ecolsys.36.091704.175520)

11. Pearse IS, Altermatt F. 2013 Predicting novel trophic interactions in a non-native world. *Ecol. Lett.* **16**, 1088–1094. (doi:10.1111/ele.12143)

12. Schoonhoven LM, van Loon JJA, Dicke M. 2005 *Insect-Plant Biology*. Oxford: Oxford University Press.

13. Winkler IS, Mitter C. 2008 The phylogenetic dimension of insect/plant interactions: a summary of recent evidence. In *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects.* (ed K Tilmon), pp. 240–263. Berkeley, California: University of California Press.

14. Novotny V *et al.* 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.

15. Forister ML, Dyer LA, Singer MS, Stireman III JO, Lill JT. 2012 Revisiting the evolution of ecological specialization, with emphasis on insect-plant interactions. *Ecology* **93**, 981–991.

16. Weiblen GD, Webb CO, Novotny V, Basset Y, Miller SE. 2006 Phylogenetic dispersion of host use in a tropical insect herbivore community. *Ecology* **87**, S62–S75.

17. Joy JB, Crespi BJ. 2012 Island phytophagy: explaining the remarkable diversity of plant-feeding insects. *Proc. Roy. Soc. B Biol. Sci.* **279**, 3250–3255.

18. Bernays EA, Chapman RF. 1994 *Host plant Selection by Phytophagous Insects*. London: Chapman and Hall.

19. Novotny V, Miller SE, Basset Y, Cizek L, Drozd P, Darrow K, Leps J. 2002 Predictably simple: assemblages of caterpillars (Lepidoptera) feeding on rainforest trees in Papua New Guinea. *Proc. R. Soc. Lond. B* **269**, 2337–2344.

20. Marquis RJ, Salazar D, Baer C, Reinhardt J, Priest G, Barnett K. 2016 Ode to Ehrlich and Raven or how herbivorous insects might drive plant speciation. *Ecology* **97**, 2939–2951. (doi:10.1002/ecy.1534)

21. Althoff DM, Segraves KA, Johnson MTJ. 2014 Testing for coevolutionary diversification: linking pattern with process. *Trends. Ecol. Evol.* **29**, 82–89. (doi:10.1016/j.tree.2013.11.003)

22. Nylin S, Slove J, Janz N. 2014 Host plant utilization, host range oscillations and diversification in nymphalid butterflies: a phylogenetic investigation: host range oscillations in butterflies. *Evolution* **68**, 105–124. (doi:10.1111/evo.12227)

23. Wang H, Holloway JD, Janz N, Braga MP, Wahlberg N, Wang M, Nylin S. In. Press. Polyphagy and diversification in tussock moths: support for the oscillation hypothesis from extreme generalists. *Ecol. Evol.*

24. Hamm CA, Fordyce JA. 2015 Patterns of host plant utilization and diversification in the brush-footed butterflies: butterfly diversification and host use. *Evolution* **69**, 589–601. (doi:10.1111/evo.12593)

25. Janz N, Braga MP, Wahlberg N, Nylin S. 2016 On oscillations and flutterings-A reply to Hamm and Fordyce: TECHNICAL COMMENT. *Evolution* **70**, 1150–1155. (doi:10.1111/evo.12927)

26. Barbehenn RV, Constabel CP. 2011 Tannins in plant–herbivore interactions. *Phytochem.* **72**, 1551–1565. (doi:http://dx.doi.org/10.1016/j.phytochem.2011.01.040)

27. Salminen J-P, Karonen M. 2011 Chemical ecology of tannins and other phenolics: we need a change in approach: Chemical ecology of tannins. *Funct. Ecol.* **25**, 325–338. (doi:10.1111/j.1365-2435.2010.01826.x)

28. Miller SE, Novotny V, Basset Y. 2003 Studies on New Guinea moths. 1. Introduction (Lepidoptera). *Proc. Ent. Soc. Wash.* **105**, 1035–1043.

29. Novotny V *et al.* 2007 Low beta diversity of herbivorous insects in tropical forests. *Nature* **448**, 692–695. (doi:10.1038/nature06021)

30. Whitfeld TJS, Novotny V, Miller SE, Hrcek J, Klimes P, Weiblen GD. 2012 Predicting tropical insect herbivore abundance from host plant traits and phylogeny. *Ecology* **93**, 211–222.

31. Miller SE, Hrcek J, Novotny V, Weiblen GD, Hebert PDN. 2013 DNA barcodes of caterpillars (Lepidoptera) from Papua New Guinea. *Proc. Ent. Soc. Wash.* **115**, 107–109.

32. Novotny V, Basset Y. 2005 Host specificity of insect herbivores in tropical forests. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1083–1090. (doi:10.1098/rspb.2004.3023)

33. Wahlberg N, Wheat CW. 2008 Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Syst. Biol.* **57**, 231–242.

34. Engstrom MT, Palijarvi M, Salminen J-P. 2015 Rapid Fingerprint Analysis of Plant Extracts for Ellagitannins, Gallic Acid, and Quinic Acid Derivatives and Quercetin-, Kaempferol- and Myricetin-Based Flavonol Glycosides by UPLC-QqQ-MS/MS. *Journal of Agricultural and Food Chemistry* **63**, 4068–4079. (doi:10.1021/acs.jafc.5b00595)

35. Hagerman AE. 1987 Radial diffusion method for determining tannin in plant extracts. *Journ. Chem. Ecol* **13**, 437–449. (doi:10.1007/BF01880091)

36. Ives AR, Garland T. 2014 Phylogenetic Regression for Binary Dependent Variables. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice* (ed LZ Garamszegi), pp. 231–261. Berlin, Heidelberg: Springer Berlin Heidelberg. (doi:10.1007/978-3-662-43550-2_9)

37. Freckleton RP, Cooper N, Jetz W. 2011 Comparative Methods as a Statistical Fix: The Dangers of Ignoring an Evolutionary Model. *Am. Nat.* **178**, E10–E17. (doi:10.1086/660272)

38. Barton K. 2016 *MuMIn: Multi-Model Inference*. See http://CRAN.R-project.org/package=MuMIn.

39. Ives AR, Godfray HCJ. 2006 Phylogenetic analysis of trophic associations. *Am. Nat.* **168**, 1–14.

40. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010 PICANTE: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463.

41. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, Solymos P, Stevens HH, Wagner H. 2010 *Vegan: Community Ecology Package. R package version 1.17-3.* See http://CRAN.R-project.org/package=vegan.

42. Leppänen SA, Altenhofer E, Liston AD, Nyman T. 2012 Ecological Versus Phylogenetic Determinants of Trophic Associations in a Plant-Leafminer-Parasitoid Food Web. *Evolution* , 1492–1503. (doi:10.1111/evo.12028)

43. Singer MS. 2008 In *Evolutionary ecology of polyphagy*, pp. 29–42. Univ. California Press, Berkeley.

44. Nosil P. 2002 Transition rates between specialization and generalization in phytophagous insects. *Evolution* **56**, 1701–1706.

45. Lopez-Vaamonde C, Godfray HC., Cook JM. 2003 Evolutionary dynamics of host-plant use in a genus of leaf-mining moths. *Evolution* **57**, 1804–1821.

46. Robinson GS, Ackery PR, Kitching I, Beccaloni G, Hernández L. 2001 *Hostplants of the moth and butterfly caterpillars of the Oriental Region.* Natural History Museum.

47. Schmidt O. 2016 Larval food plants of Australian Larentiinae (Lepidoptera: Geometridae) - a review of available data. *Biodiv. Data Journ.* **4**, e7938. (doi:10.3897/BDJ.4.e7938)

48. Celorio-Mancera M de la P, Wheat CW, Huss M, Vezzi F, Neethiraj R, Reimegård J, Nylin S, Janz N. 2016 Evolutionary history of host use, rather than plant phylogeny, determines gene expression in a generalist butterfly. *BMC. Evol. Biol.* **16**. (doi:10.1186/s12862-016-0627-y)

49. Moilanen J, Koskinen P, Salminen J-P. 2015 Distribution and content of ellagitannins in Finnish plant species. *Z. Naturforsch. C.* **116**, 188–197. (doi:10.1016/j.phytochem.2015.03.002)

50. Barbehenn RV, Jones CP, Hagerman AE, Karonen M, Salminen J-P. 2006 Ellagitannins have greater oxidative activities than condensed tannins and galloylglucoses at high pH: potential impact on caterpillars. *J. Chem. Ecol.* **32**, 2253–2267.

51. Ruuhola T, Salminen P, Salminen J-P, Ossipov V. 2013 Ellagitannins: defences of Betula nana against *Epirrita autumnata* folivory? *Agr. Forest Entomol.* **15**, 187–196.

52. Roslin T, Salminen J-P. 2008 Specialization pays off: contrasting effects of two types of tannins on oak specialist and generalist moth species. *Oikos* **117**, 1560–1568. (doi:10.1111/j.0030-1299.2008.16725.x)

53. Salminen J-P, Roslin T, Karonen M, Sinkkonen J, Pihlaja K, Pulkkinen P. 2004 Seasonal Variation in the Content of Hydrolyzable Tannins, Flavonoid Glycosides, and Proanthocyanidins in Oak Leaves. *Journ. Chem. Ecol* **30**, 1693–1711. (doi:10.1023/B:JOEC.0000042396.40756.b7)

54. Barbehenn RV, Maben RE, Knoester JJ. 2008 Linking phenolic oxidation in the midgut lumen with oxidative stress in the midgut tissues of a tree-feeding caterpillar *Malacosoma disstria* (Lepidoptera: Lasiocampidae). *Env. Entomol.* **37**, 1113–1118.

55. Barbehenn RV, Jaros A, Lee G, Mozola C, Weir Q, Salminen J-P. 2009 Hydrolyzable tannins as 'quantitative defenses': limited impact against *Lymantria dispar* caterpillars on hybrid poplar. *Journ. Insect Physiol.* **55**, 297–304.

56. Quideau S, Feldman KS, Appel HM. 1995 Chemistry of gallotannin-derived o-quinones: Reactivity toward nucleophiles. *J. Org. Chem* **60**, 4982–4983.

57. Feldman KS, Sambandam A, Bowers KE, Appel HM. 1999 Probing the Role of Polyphenol Oxidation in Mediating Insect−Pathogen Interactions. Galloyl-Derived Electrophilic Traps for the *Lymantria d ispar* Nuclear Polyhedrosis Virus Matrix Protein Polyhedrin. *J. Org. Chem* **64**, 5794–5803. (doi:10.1021/jo982477n)

58. Holloway JD, Kibby G, Peggie D. 2001 *The families of Malesian moths and butterflies*. Leiden [Netherlands] ; Boston : Brill.

59. Anderson-Teixeira KJ *et al.* 2015 CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Chang. Biol.* **21**, 528–549. (doi:10.1111/gcb.12712)

Figure and table captions

Table 1. Model coefficients and significance for phylogenetic logistic regressions between moth occurrence and foliar chemistry for the Madang data set.

Response	Explanatory	Estimate	SE	z	p
Geometridae	Oxidative Activity	0.317	0.109	2.902	0.004
Geometridae	Protein Precipitation Capacity	-0.038	0.029	-1.290	0.197
Pyraloidea	Oxidative Activity	0.030	0.054	0.563	0.573
Pyraloidea	Protein Precipitation Capacity	0.015	0.021	0.700	0.484

Figure 1. Co-phylogeny between i) geometrid and ii) pyraloid caterpillars included in the Madang data. Heat maps superimposed on the geometrid host phylogeny show observed abundance (left) and the predicted probability of occurrence (right).

Figure 2. The distributions and predicted proportion of hits under different randomisations. Expectations under totally random assignment of host use and constrained at the clade level are given in blue for geometrids and red for pyraloids. Actual predictions are given in filled lines of the same colour for each moth lineage. Predictions under randomisation within families (white) and genera (grey) are given for comparison. Dashed lines show the number of interactions correctly predicted at the clade level for each moth lineage.