doi:10.1111/evo.13557

Host use dynamics in a heterogeneous fitness landscape generates oscillations in host range and diversification

Mariana P. Braga,[1,2] Sabrina B. L. Araujo,[3] Salvatore Agosta,[4] Daniel Brooks,[5] Eric Hoberg,[6] Sören Nylin,[1] Niklas Janz,[1] and Walter A. Boeger[7]

[1]*Department of Zoology, Stockholm University 10691, Stockholm, Sweden*

[2]*E-mail: mariana.braga@zoologi.su.se*

[3]*Departamento de Física, Universidade Federal do Paraná, Curitiba, PR 81531, Brazil*

[4]*Center for Environmental Studies and Department of Biology, Virginia Commonwealth University, Richmond, Virginia 23284*

[5]*Institute for Advanced Studies, Kőszeg, Europe House, Kőszeg Chernel st. 14 H-9730, Hungary*

[6]*US National Parasite Collection, US Department of Agriculture, Agricultural Research Service, Beltsville, Maryland 20705*

[7]*Laboratory of Molecular Ecology and Evolutionary Parasitology, Departamento de Zoologia, Universidade Federal do Paraná, Curitiba, PR 81531, Brazil*

Received March 28, 2017
Accepted July 8, 2018

Colonization of novel hosts is thought to play an important role in parasite diversification, yet little consensus has been achieved about the macroevolutionary consequences of changes in host use. Here, we offer a mechanistic basis for the origins of parasite diversity by simulating lineages evolved in silico. We describe an individual-based model in which (i) parasites undergo sexual reproduction limited by genetic proximity, (ii) hosts are uniformly distributed along a one-dimensional resource gradient, and (iii) host use is determined by the interaction between the phenotype of the parasite and a heterogeneous fitness landscape. We found two main effects of host use on the evolution of a parasite lineage. First, the colonization of a novel host allowed parasites to explore new areas of the resource space, increasing phenotypic and genotypic variation. Second, hosts produced heterogeneity in the parasite fitness landscape, which led to reproductive isolation and therefore, speciation. As a validation of the model, we analyzed empirical data from Nymphalidae butterflies and their host plants. We then assessed the number of hosts used by parasite lineages and the diversity of resources they encompass. In both simulated and empirical systems, host diversity emerged as the main predictor of parasite species richness.

KEY WORDS: **Host range, individual-based model, parasite diversity, phenotypic amplitude.**

Parasitism, broadly defined, is a ubiquitous kind of ecological interaction that includes organisms in various trophic levels (e.g., phytophagous insects, ecto- and endoparasites with direct or indirect life cycles). Such intimate ecological interactions may persist over long-time spans, coupling the evolutionary history of the interacting lineages and producing broad patterns of association between host and parasite taxa, such as conservatism in host use (Dethier 1954; Ehrlich and Raven 1964; Janz and Nylin 1998; Brooks and McLennan 1993, 2002). Nonetheless, the level of specialization of parasites to their hosts varies greatly between and within lineages. Among butterflies, for example, most species use plants from a single family as hosts, but there are several species able to use a repertoire of hosts that includes between two and 36 plant families (Forister et al. 2015). When placing this variation in a phylogenetic context, most studies have found that host repertoire is labile across time and space (Nosil 2002; Braga et al.

2014; Nylin et al. 2014; Calatayud et al. 2016), which shows that the two apparently contradictory paths of fine-tuning to a host or exploring the range of potential hosts can be taken without preventing the other (Agosta and Klemens 2008; Araujo et al. 2015; Nylin et al. 2018).

Variation in host specialization has been shown to impact several ecological and evolutionary processes, including parasite diversification (Janz et al. 2006; Dennis et al. 2011; Hardy and Otto 2014). According to the Oscillation Hypothesis, host range expansion and contraction over macroevolutionary time is one of the main drivers of diversification of plant-feeding insects (Janz and Nylin 2008). Host range expansion, coupled with the expansion of overall niche breadth and geographic range (Slove and Janz 2011; Dennis et al. 2011), is expected to create opportunities for divergence by both adaptive and neutral processes (Janz et al. 2016). These dynamics then produce the observed pattern of correlation between diversity of host use and species richness in a clade (Janz et al. 2006; Wang et al. 2017).

While the importance of host range expansions for the evolution of host use have gained support following the proposal of the Oscillation Hypothesis, its role in diversification remains controversial (Hardy and Otto 2014; Nylin et al. 2014; Hamm and Fordyce 2015; Janz et al. 2016; Hardy 2017; Wang et al. 2017). This controversy is perhaps anticipated given the number of factors that potentially drive speciation in parasite lineages. Ultimately, any process that reduces gene flow between populations generates patterns of genetic differentiation (Slatkin 1993; Bolnick and Otto 2013). Decades of research into the process of isolation by distance (Wright 1943) have shown that gene flow is often geographically restricted. However, geography is not the only landscape component that affects population connectivity (Kool et al. 2013). Environmental heterogeneity can also structure genetic variation, producing the pattern of isolation by environment (Wang and Summers 2010). This pattern can be generated by a variety of ecological factors, ranging from a temperature cline to more complex interactions between continuous and discrete variables representing abiotic and biotic factors (Wang and Bradburd 2014). Although, the local environment of parasitic organisms encompasses as many dimensions as that of free-living organisms (Agosta et al. 2010), genetic differentiation within and between parasite species is likely mediated by their hosts (Ellis et al. 2015), increasing the discreteness of the fitness landscape.

Here, we seek to advance our understanding of host-associated diversification processes by investigating the potential for isolation by environment in a simple but heterogeneous fitness landscape. The aim of the present paper is to offer a mechanistic basis for the origins of macroevolutionary patterns of parasite diversity and host range, by studying lineages evolved in silico where species-level properties, such as host range, are an outcome of the system dynamics. Our simulations consist of parasite individuals evolving in a fixed fitness landscape where reproductive isolation is allowed. Since geographic space is not explicitly modeled here, isolation can only happen as a result of ecologically based divergent selection, that is isolation by environment (Wang and Bradburd 2014), leading to ecological speciation (Rundle and Nosil 2005). This is a simplification and can be seen as the "worst-case scenario" for diversification, as adaptation to a given host should decrease genetic and phenotypic variation, and therefore evolvability. While the Oscillation Hypothesis is agnostic about modes of speciation, this island-like scenario can give us insight about the processes in the natural world that allow parasites to colonize new hosts even when the selective pressure imposed by hosts is high, and how the host use dynamics affects parasite diversification.

Methods

MODEL DESCRIPTION

We simulated host use dynamics by modeling individuals of a parasite population that evolves under the combined effects of sexual reproduction, mutation, natural selection, and host probing (i.e., attempts to colonize a new host). In our model, characterization of parasites derives mainly from the agent-based model described by de Aguiar et al. (2009). The model description follows the ODD protocol for describing individual-based models (Grimm et al. 2006, 2010) and consists of seven elements. The first three elements provide an overview, the fourth element explains general concepts underlying the model's design, and the remaining three elements provide details.

Purpose

In this study, we focus on how the colonization of new hosts affects the parasite phenotypic distribution, its ability to use novel resources, and the likelihood of speciation. The novelty in the study does not come from model elements, but from the use of these well-known representations of biological systems to address long-standing questions in host-parasite evolution.

Entities, state variables, and scales

The model comprises three hierarchical levels: individual, species, and fitness landscape. Individuals are characterized by three state variables: host, genotype, and species identity. The first variable is simply the identity of the host being used by the parasite individual at a given time step, and the genotype is a binary string of length L. The number of mismatches along two genotypes determines the genetic distance between two parasite individuals.

A species is a group of individuals connected by reproduction, which only occurs between individuals with a genetic distance smaller than the mate recognition threshold, g. Note that the genetic distance between two individuals from the same species

can be larger than g, as long as they are connected by individuals with intermediary genotypes.

The fitness landscape is composed of fitness peaks, which represent hosts for the parasites. The phenotype of a parasite is defined by the number of 1s in the genotype (see eq. 1). Hosts are distributed in the resource space in the beginning of each simulation from the center to the peripheries with distance P from each other (i.e., fitness-peak interval). The position of a given host in the resource space represents the optimum phenotype to use that given host (the parasite phenotype that yields maximum survival). Survival decreases with increasing difference between the host optimum and the parasite phenotype (see eq. 2). As the fitness landscape is fixed (hosts do not evolve), it can also be interpreted as the result of the combination of environmental and host-related factors that allow parasite survival.

Space is not explicitly modeled here and the time scale is a parasite life cycle, so that each time step in the model is equivalent to one parasite generation.

Process overview and scheduling

At each time step, three events happen in the following order: reproduction, host probing, and selection. Each event is detailed in the section *Submodels*. The parasite population in a given host at a given time step is composed of the offspring of the parasites from the previous time step that did not attempt to colonize a new host and survived the selective pressure, in addition to the parasites that successfully colonized this given host.

Design concepts

Emergence: Host use dynamics and parasite diversification emerge from the behavior of the individuals, but the individual's behavior is determined by, for example, survival and host probing probabilities.

Collectives: Individuals are grouped by host use. The distribution of individuals on the available hosts emerges from the successful colonization of new hosts.

Fitness: The survival probability of an individual depends on how well adapted the individual is to the host being used. Survival is maximized when the phenotype of the individual matches the phenotype favored by the host, and decreases with the distance from it (eq. 2).

Interaction: Individuals interact indirectly via use of shared resources, as reproduction is truncated by the carrying capacity, which determines the maximum number of parasite individuals each host can support regardless of parasite species identity.

Stochasticity: All events are probabilistic. The model compares the input probabilities for each life-history event with a randomly generated number from a uniform distribution between 0 and 1. If the number is smaller than the input, the event is carried out. During *reproduction*, mate choice is random—regardless of species identity and host use. The genetic contribution of each parent during recombination is randomly chosen and mutation can happen in any locus. During *host probing*, each parasite individual has the same chance to probe a new host, which is also randomly chosen. During *selection*, parasite survival probability varies with the phenotypic match with the host (see eq. 2).

Observation: Parasite phenotype, species identity, and utilized host were recorded for every individual over time.

Table 1. Overview of processes, parameters, and values for the simulations reported.

Parameter	Value
Reproduction	
Genome size	$L = \mathbf{200}$
Mate recognition threshold	$g = \{\mathbf{1, 5, 10}\}$
Mutation rate	$m = \{\mathbf{10^{-5}, 10^{-4}, 10^{-3}}, 10^{-2}\}$
Maximum number of offspring per event	$b = \{1, \mathbf{2, 4, 8}\}$
Carrying capacity per host	$K = \{\mathbf{50}, 100, 200\}$
Host probing	
Host probing probability	$h = \{0.005, 0.01, 0.05, 0.1, 0.2, \mathbf{0.3}, 0.4, 0.5\}$
Total number of hosts in the resource space	$N_h = \mathbf{11}$
Fitness-peak interval	$P = \{\mathbf{0.2}, 0.3, \mathbf{0.4}, 0.5, \mathbf{0.6}, 0.7, \mathbf{0.8}, 0.9, \mathbf{1}\}$
Selection	
Standard deviation for survival probability	$\sigma_r = 0.5$

Values in bold are the ones used in the main analysis.

Initialization

Every simulation starts with three parasite individuals utilizing one host at the center of the resource space. Genetic and phenotypic variation increases over time due to mutation and recombination, as long as the offspring is able to survive the selective pressure imposed by the host.

Input

The particular data used to parameterize the model will depend on the particular system to which it is applied. Table 1 lists the basic set of parameters, which would be required for any system.

Submodels

Reproduction: All individuals have a chance to find a compatible mate. A random mate is selected for each individual, but reproduction only occurs if the genetic distance is smaller than the mate recognition threshold, g. When mate pairing is successful, the genotype of the offspring is the result of genetic inheritance from

both parents, with recombination and mutation. Each reproduction event has genetic crossover, where the genetic contribution of each parent is randomly chosen, and mutation occurs with probability m (following de Aguiar et al. 2009). The number of offspring generated by each reproduction event is equal to b, as long as the number of parasites on the focal host (where the original parasite was) is smaller than the carrying capacity K.

Host probing: After reproduction, the offspring can probe new hosts with probability h, so that, on average, h percent of the population tries to colonize a new host during each generation. Successful host colonization happens when colonizers survive the selective pressure imposed by the new host. Note that there are no explicit trade-offs modeled between hosts or costs for species with a wide host range, that is using a large number of hosts.

Selection: The genotype determines the phenotype of the individual, which combines all parasite traits related to host use, for simplicity. The phenotype of parasite i is proportional to the sum of all L loci of its genotype:

$$z_i = \frac{1}{10}\sum_{n=1}^{L} X_{i,n} \quad (1)$$

where $X_{i,n} \in \{0,1\}$, the two possible alleles. The term 1/10 limits the range of possible phenotypes to $0 \leq z_i \leq L/10$ with discrete intervals of 0.1. The parasite population can use N_h available hosts uniformly distributed along a one-dimensional resource gradient. The number of hosts and the distance between them in this gradient define the resource space for the parasites. Each host is modeled as a fitness peak in the fitness landscape of the parasites (Fig. 1), which means that there is a unique phenotype value that yields maximum survival for each host, as in Araujo et al. (2015). Nonetheless, this does not mean that hosts are only used by perfectly adapted parasites; instead, parasite survival decreases with the distance of the phenotype from the host optimum, following a normal distribution centralized on $|z_{ij} - q_j|$:

$$P\left(q_j, z_{ij}\right) = exp\left[\frac{(z_{ij} - q_j)^2}{2\sigma_r^2}\right], \quad (2)$$

where q_j is the optimal phenotype to use host j, z_{ij} is the phenotype of parasite i using host j, and σ_r is the parameter that controls the intensity of the selective pressure. Therefore, the fitness of a given parasite depends on the host it is using (Fig. 1), and the likelihood of successfully colonizing a new host varies with the distance between hosts, herein referred to as peak interval. For simplicity, we assume that selection intensity is the same for all hosts. It is also important to highlight that although genotypes translate deterministically into phenotypes, mate recognition is determined by allele position and phenotype by sum of alleles. This relaxes the link between reproductive isolation and host adaptation, allowing different species to use the same host.

Analysis

Analyses were conducted in the R statistical environment v. 3.3.3 (R Core Team 2017). For preliminary analysis, we conducted pairwise combinations of parameters in sets of ten simulations, each one iterated for 10,000 generations. Because we were mainly interested in the role of the fitness landscape on parasite diversification, we varied the fitness-peak interval, P, and one other parameter at a time. Table 1 lists all parameter values tested. At the end of each simulation, we recorded the total number of hosts used by all parasites (host range), the difference between maximum and minimum parasite phenotypes (phenotypic amplitude), and the number of isolated reproductive units (species richness). This preliminary analysis showed that carrying capacity and host-probing probability did not have a significant effect on species richness (Table S1), therefore they were removed from subsequent analyses. Also, parasites went extinct when only one offspring was produced in each reproduction event ($b = 1$) or when mutation rate was set to 0.01, so we removed these parameter values from subsequent analyses.

We then performed simulations crossing values of fitness-peak interval, birth rate, mate recognition threshold, and mutation rate (values in bold in Table 1), which resulted in 135 parameter combinations. Each combination was iterated for 1000 generations and replicated three times. Then we estimated the effect of each variable on parasite species richness using a Poisson regression, and on phenotypic amplitude and host range using Gaussian regressions. The proportion of deviance explained (D^2) by each variable was calculated as the improvement from the model without the focal variable (variables were added by decreasing contribution to deviance explained), using the function *Dsquared* of the R package *modEvA* (Guisan and Zimmermann 2000; Barbosa et al. 2016). Finally, to estimate the relationships between model outcomes we used partial correlations between species richness, phenotypic amplitude, and host range.

VALIDATION WITH EMPIRICAL DATA

The development of a new method to assess the diet breadth of herbivorous insects (Fordyce et al. 2016) allowed us to compare the results of our simulations with empirical data from butterfly-plant interactions. This method uses information on how often different host plant taxa are utilized by the same butterfly taxon as a proxy for resource similarity. Using the R package *ordiBreadth* (Fordyce 2015) and the host use dataset assembled by Nylin et al. 2014, host-plant orders were distributed in a multidimensional ordination space (analogous to the resource space in the model) based on their interactions with nymphalid butterfly genera. The data set from Nylin et al. (2014) included reliable records, for which phylogenetic information is available, of 566 interactions between 295 Nymphalidae genera from 43 tribes and 43 plant

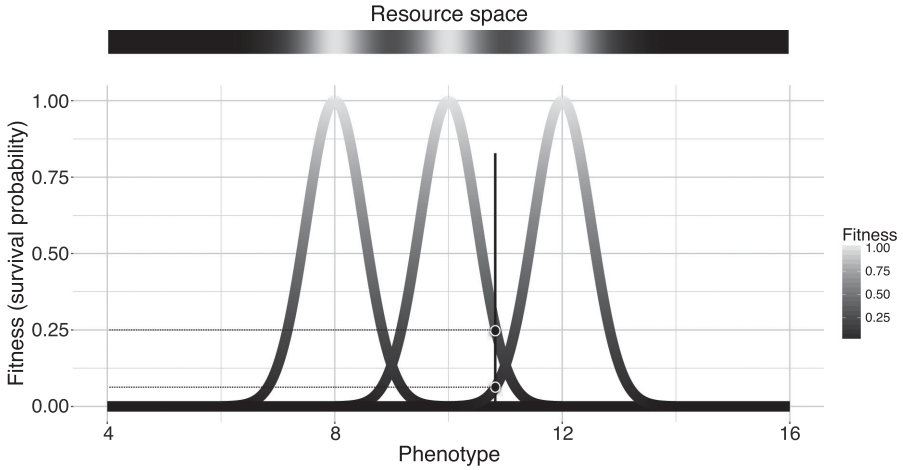

Figure 1. Hypothetical fitness landscape with three hosts (peaks) and the resulting 1-dimensional resource space, where white patches represent phenotypes with positive fitness for each host. The solid vertical line is a projection of a given phenotype to show that the fitness of a parasite on a given host depends on where the line crosses the fitness curve (dots), and that this varies between hosts (note that this individual has a positive fitness in two hosts).

orders. Once the ordination is done, each butterfly taxon can be placed at the multivariate centroid of its host plants and the ordinated diet breadth (ODB) is calculated based on the distances between each host and the butterfly centroid (Fordyce et al. 2016). Therefore, the ODB of a butterfly taxon increases with the number of plants it uses as hosts (taxonomic host range) and the distance between these hosts in the ordination space (which is analogous to the fitness-peak interval in the model). Since the direct assessment of phenotypic amplitude is much harder in natural systems, a measure of host diversity (such as ODB) is a good proxy for parasite host-use abilities.

We calculated ODB for the repertoire of host plant orders utilized by each of the 43 butterfly tribes represented in the dataset. We also calculated Faith's phylogenetic diversity of plant orders used by each tribe (phylogenetic host range), based on the phylogenetic relationships among plant orders proposed by the Angiosperm Phylogeny Website (Stevens 2011). Finally, species richness of butterfly tribes was retrieved from Savela (2014).

We assessed the relationships between ODB, taxonomic host range, phylogenetic host range, and species richness using phylogenetic path analysis (Hardenberg and Gonzalez Voyer 2013), which controls for nonindependence due to shared ancestry among butterfly tribes. For that, we used the phylogenetic tree for the tribes within Nymphalidae assembled by Nylin et al. 2014. However, strong correlation between ODB and taxonomic host range in our dataset prevented us from disentangling their effects on species richness. To resolve that, we calculated the average distance between hosts and the multivariate centroid, that is ODB/taxonomic host range, for each butterfly tribe with nonzero ODB. This measure can be interpreted as how much the niche expands when one plant order is added to the host repertoire of a butterfly tribe, and is herein referred to as adjusted ODB (following Fordyce et al. 2016). Using the R package *phylopath* version 1.0.0 (van der Bijl 2018), we fitted path models with different causal structures between the four variables. Model selection was performed using the *C* statistic Information Criterion, CICc (Cardon et al. 2011; Shipley 2013) and bootstrap confidence intervals were computed for each path coefficient.

Results

HOST USE AND SPECIATION DYNAMICS

Despite the small initial population size, host range expansion, and speciation happened under a broad range of parameters (Fig. S1). As the population grew, mutation, and recombination produced genetic and phenotypic variation, as seen in the first 260 generations of the example simulation shown in Figures 2A and 3A. Up to that point in time (which varies between simulations) all parasites belonged to the same species, but reproductive isolation happened in the following generations, when the minimum genetic distance between the two reproductive units became larger than the threshold, g. Both descendant species continued to accumulate new phenotypes until reproductive isolation occurred again in either lineage, producing new species. Figure 2 shows the overall dynamics during the first 1000 generations of the example simulation, and Figure 3 shows changes in phenotype frequency around the first speciation event of the same example. All panels in Figures 2 and 3 show the phenotypic distribution of parasites through time, but Figures 2B and 3B also show which host is used by each parasite individual.

Although parasites were initially perfectly adapted to the original host ($z_1 = q_1 = 10$), they colonized the two closest

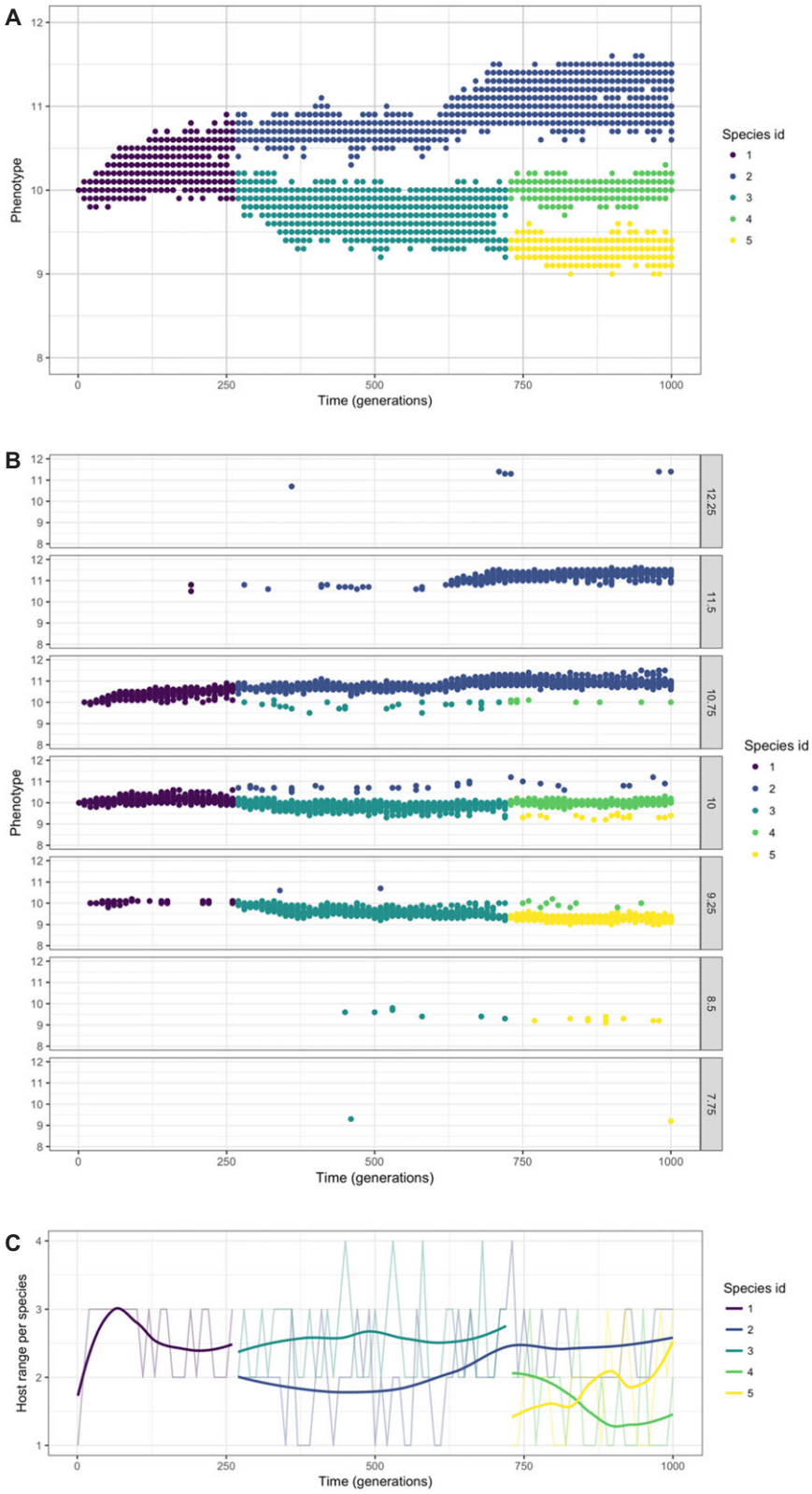

Figure 2. Parasite phenotypic distribution through time in the example simulation, where peak interval = 0.75 and probing probability = 0.2. Each dot represents a parasite individual (overlapping points can occur) and species are identified by color. In A, all parasites are shown together while in B, each panel shows the parasites using each given host. Numbers at the right end of the panels show the optimum phenotype to use each host. Variation in host range per species through time is shown in C. Because colonizations and local extinctions happen frequently, a loess curve was added for each species to highlight the overall direction of change.

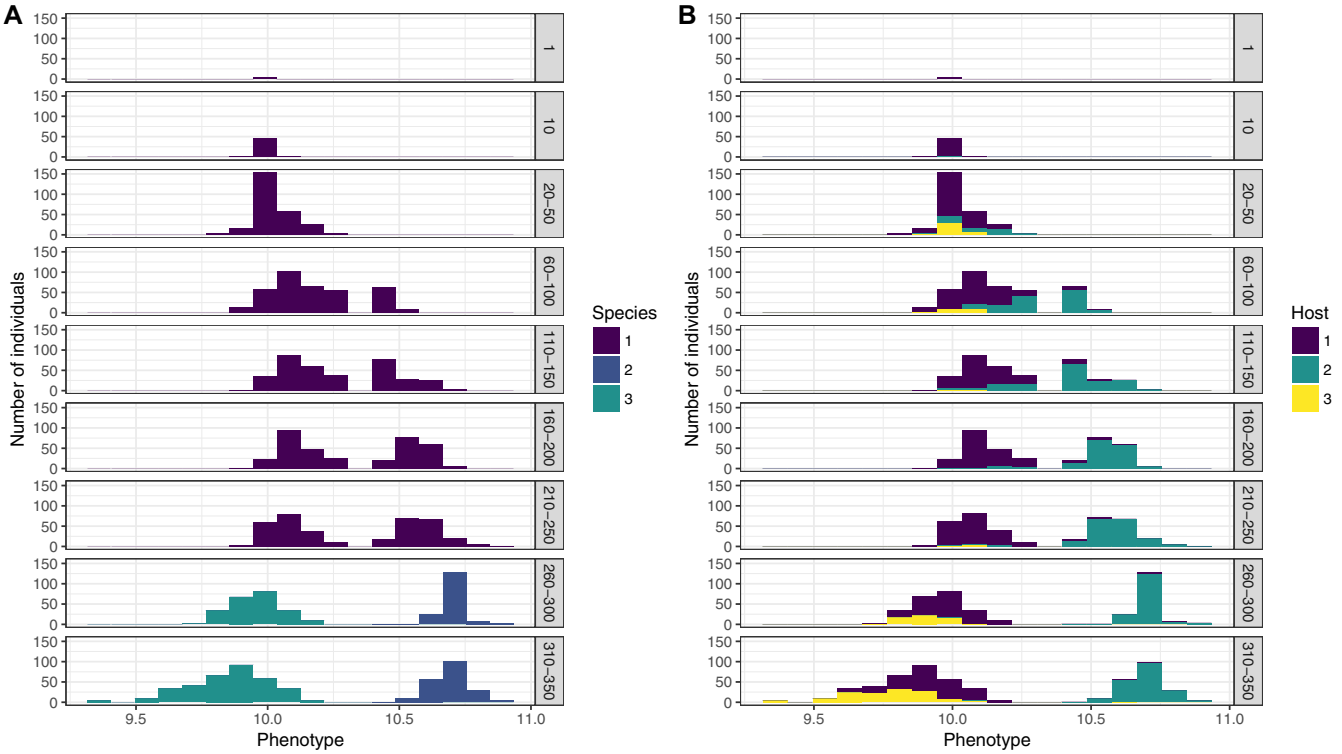

Figure 3. Parasite phenotype frequency through time in the example simulation. Each panel shows the frequency for a given time interval (which are shown at the right end of the panels). In A, individuals are colored by species to show phenotypic expansion until reproductive isolation occurs. In B, individuals are colored by host to show host use dynamics and its effect on speciation.

hosts within the first 50 generations in the example shown in Figures 2B and 3B. New hosts had two main effects in this system. First, phenotypic variation increased when a new host was added to the repertoire because, with time, part of the population adapted to the new host, extending the phenotypic distribution across both fitness peaks (e.g., generations 60–250 in the example). Second, the inclusion of a second fitness peak in the fitness landscape of the parasite allowed divergent selection, which could lead to speciation. Once gene flow between parasites on different hosts was interrupted, each species evolved toward the optimum phenotype for each host (or hosts, as in the case of species 3 in Figs. 2B and 3B). It is important to note that although each species specialized on its main host(s), they continued to colonize other hosts, including ancestral ones (e.g., blue and yellow dots in the central panel of Fig. 2B). The continuous ability to explore the resource space by using alternative hosts (hosts that yield a lower fitness than the main host), coupled with adaptation to a subset of the host range due to lack of gene-flow produced oscillations in host range through time (Fig. 2C).

DRIVERS OF PARASITE SPECIES RICHNESS

All four parameters tested had some effect on simulation outcomes, with parasite species richness and phenotypic amplitude mostly affected by mutation rate, and host range mostly affected

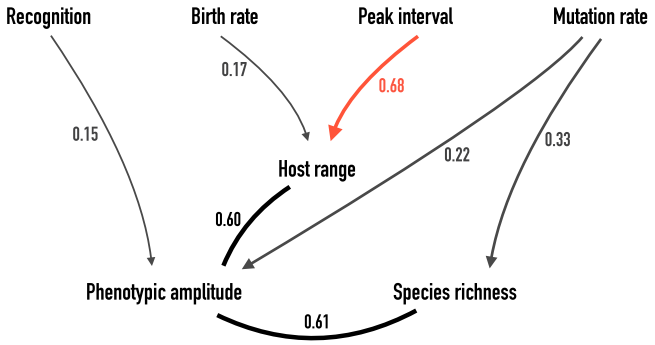

Figure 4. Relationships between model parameters and model outcomes. Arrows represent the effect of parameters on outcomes and associated numbers show the deviance explained (D^2) by each parameter. Gray arrows show positive effects and the red arrow shows a negative effect. Relationships with $D^2 < 0.15$ were omitted. Lines connecting model outcomes indicate partial correlations between the three variables.

by fitness-peak interval (Table 2, Fig. 4). Across all parameter combinations, phenotypic amplitude was positively correlated with species richness, even when controlling for the effect of host range (Spearman partial $r = 0.61$, $n = 405$, $P < 0.001$). On the other hand, host range has no effect on species richness when controlling for phenotypic amplitude (Spearman partial

Table 2. Effect of parameters on simulation outcomes.

	β	β se	*z*-score	*P*	Res. dev.	Res. df	D^2
Model 1—Species richness							
intercept	3.825	0.172	22.196	<0.001	1191.05	404	
mutation rate (log)	0.325	0.017	18.585	<0.001	796.81	403	0.33
birth rate	0.112	0.011	9.850	<0.001	699.97	402	0.12
peak interval	–0.883	0.104	–8.498	<0.001	626.28	401	0.11
mate recognition threshold	0.004	0.008	0.449	0.654	626.08	400	–
Model 2—Phenotypic amplitude							
intercept	4.080	0.337	12.117	<0.001	985.74	404	
mutation rate (log)	0.390	0.030	12.907	<0.001	767.76	403	0.22
birth rate	0.202	0.023	8.876	<0.001	664.68	402	0.13
mate recognition threshold	0.136	0.015	8.785	<0.001	563.71	401	0.15
peak interval	–1.116	0.201	–5.554	<0.001	523.35	400	0.07
Model 3—Host range							
intercept	12.535	0.517	24.246	<0.001	5689.5	404	
peak interval	–10.907	0.309	–35.353	<0.001	1834.9	403	0.68
birth rate	0.351	0.035	10.02	<0.001	1525.2	402	0.17
mate recognition threshold	0.176	0.024	7.429	<0.001	1355	401	0.11
mutation rate (log)	0.291	0.046	6.272	<0.001	1233.7	400	0.09

Residual deviance and residual degrees of freedom are given for each model and for each predictor variable. The slope (β), β standard error, *z*-score, *P*-value, and D^2 are also given for each variable.

$r = –0.05$, $n = 405$, $P = 0.28$). Finally, host range and phenotypic amplitude are correlated regardless of species richness (Spearman partial $r = 0.60$, $n = 405$, $P < 0.001$).

VALIDATION WITH EMPIRICAL DATA

The ordination of diet breadth produced a high-dimensional resource space, where the first three principal coordinates explain respectively, 12.3, 8.3, and 6.5% of the variation (Fig. S2). Species richness varied greatly across nymphalid tribes (4–1676 species), as did taxonomic host range (1–22 plant orders).

We tested eight different path models for the relationships between taxonomic host range, phylogenetic host range, adjusted ODB, and log-transformed species richness (Fig. S3). Of these, one model was selected based on ΔCICc <2 criterion (Fig. S4, Table S2). In this model (Fig. 5), taxonomic host range had a direct effect on phylogenetic host range (standardized path coefficient = 0.95 ± 0.056 SE) and on ODB_{adj} (0.80 ± 0.109 SE), but species richness was only influenced by ODB_{adj} (0.58 ± 0.148 SE). Species richness was higher in tribes that use more diverse resources, which was often associated with a larger host range (both taxonomic and phylogenetic). On the other hand, tribes that use hosts that were very close in the resource space are not as species rich (Fig. S5). Importantly, models with the inverse direction of causality (i.e., species richness determines taxonomic host range or ODB) received the least support from the data.

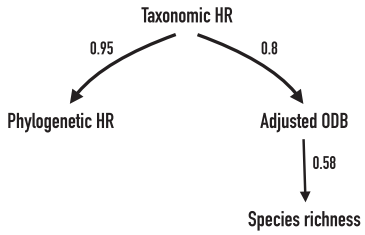

Figure 5. Summary of phylogenetic path analysis for butterfly-plant interactions showing that the effect of taxonomic host range on parasite species diversity is mediated by resource heterogeneity in Nymphalidae tribes. Arrow thickness is scaled approximately with the standardized path coefficients, which are shown for each path.

Discussion

Our model shows that host colonization allows parasites to explore new areas of the resource space, with increasing phenotypic and genotypic variation (Figs. 2 and 3). The incorporation of a novel host starts by ecological fitting: some parasite individuals have positive fitness on the novel host even when they are better adapted to the original host (Nylin et al. 2018). Because the novel host represents a new region in the fitness landscape, it allows the accumulation of more parasite variants, increasing variation in the population (c.f. Agosta and Klemens 2008). Greater phenotypic amplitude (which translates to greater host diversity) increases the probability of successful host colonization (Araujo et al. 2015),

closing a positive feedback loop between host range expansion and increasing phenotypic amplitude. Speciation breaks this loop.

In a system where individuals undergo sexual reproduction limited by genetic proximity, speciation depends on clustering of similar genetic variants (de Aguiar et al. 2009). In our model, such genetic clusters are formed because the fitness landscape is heterogeneous, despite the simplicity of the resource space and the associated fitness landscape modeled here. In nature, hosts differ in various traits—for example nutrients, defense mechanisms, mutualists, geographic distribution—producing more heterogeneity in the resource space and potentially in the fitness landscape of parasites (Nyman 2010), which would further increase opportunity for speciation. As predicted by (Nyman 2010), speciation was maximized when hosts were at an intermediate distance in resource space (Fig. S1), balancing the probabilities of colonization and divergent selection. Colonization happens quickly when hosts are very similar but divergent selection is stronger when hosts are distant.

Phenotypic amplitude emerged as the main driver of parasite species richness in the model. There is no upper limit for phenotypic/genotypic variation within a species in the model; that is, genetic clusters that differ more than the threshold for mate recognition still belong to the same species as long as individuals with intermediate genotypes connect them. However, the greater the variation within a species, the higher the likelihood of fragmentation. The distance between hosts along the resource space affects both the host range and the similarity between hosts used by parasites: host range is narrow when hosts are too different, but if parasites eventually colonize distant hosts, phenotypic amplitude quickly increases. Therefore, host diversity is maximized when a parasite uses a wide range of distant hosts. It is important to remember, however, that we assume that all hosts coexist, therefore, we cannot assess the role of geographic co-occurrence of hosts and parasites. Similarly, coevolution cannot be addressed because hosts do not evolve in our model.

In nature, host-parasite systems are more complex than modeled here: parasite and host individuals have flexible phenotypes (West-Eberhard 2003; Nylin and Janz 2009; Mason 2015), hosts are geographically distributed (Calatayud et al. 2016), the distance between hosts in the resource space is not uniform (Nyman 2010), and the resource space is multidimensional (Harrison et al. 2016; Fordyce et al. 2016). Still, we recovered compatible relationships between host range, host diversity, and species richness for modeled parasites and nymphalid butterflies. As for simulated parasite clades, Nymphalidae tribes that use more host plant orders are more species-rich when the wide host range translates into high host diversity (Fig. S5).

Our ability to find the mechanisms by which host use affects diversification of parasite lineages depends on our knowledge about how Latin binomials translate into resources for parasites (Janzen 1973; Brooks and McLennan 2002; Harrison et al. 2016). Although there is much yet to learn, we view the results of the present article as an important step forward in the development of a theoretical framework for the study of host-parasite associations, such as the patterns described and predicted by the Oscillation Hypothesis (Janz and Nylin 2008). On theoretical and empirical grounds, this study highlights the importance of the differentiation between host range and host diversity, with the latter having the main direct effect on diversification.

As suggested by the Oscillation Hypothesis, clades with wider aggregated host ranges are in general more species-rich, but here we show that this interaction is likely to be mediated by host diversity. Host range expansions lead to diversification, as long as they increase heterogeneity in the resource space, and consequently, in the fitness landscape. Moreover, our results might be a bridge between contrasting processes of diversification of phytophagous insects, and maybe parasites in general. Among nymphalid butterflies, host diversity is mainly driven by the use of unrelated host plants, hence the indirect positive interaction between host range and species richness. However, in some groups, host diversity might be high even if hosts are closely related, as long as the host plant clade is diverse regarding at least one resource axis, such as habitat, growth form, geographic distribution, or phenology. The grass-feeding Satyrini is one such example (Fig. S5). Colonization of a phylogenetically restricted plant group (Poaceae) that is instead very diverse in terms of for example habitat types, has resulted in an impressive diversification (Peña and Wahlberg 2008; Peña et al. 2011). One important aspect to investigate in future studies, we believe, is how different kinds of resource diversity relate to taxonomic and phylogenetic host range and how that affects diversification.

To conclude, we have shown that a model where host range is an outcome of the system dynamics can offer a mechanistic basis for oscillations in host range over evolutionary time, as well as for the origins of (broad-sense) parasite diversity. We see this as an important step forward in our understanding of host-associated diversification processes.

AUTHOR CONTRIBUTIONS

MPB, SBLA, DRB and WAB designed the study with input from all other authors; SBLA wrote the code; MPB performed simulations and analyses; SN and NJ provided empirical data; MPB wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

ACKNOWLEDGMENTS

S.N. was supported by VR (2015-04218) and also received funding for the workshop "Changing species associations in a changing world: a Marcus Wallenberg symposium" (MWS 2015.0009).

DATA ARCHIVING

The doi for our data is https://doi.org/10.5061/dryad.rn72k04. The program designed to run the simulations is available at http://fisica.ufpr.br/araujosbl, with versions executable in Windows, Linux, and Mac; the source code is available at GitHub https://github.com/mpiresbr/host-parasite-ibm.

LITERATURE CITED

Agosta, S. J., and J. A. Klemens. 2008. Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. Ecol. Lett. 11:1123–1134.

Agosta, S. J., N. Janz, and D. R. Brooks. 2010. How specialists can be generalists: resolving the "parasite paradox" and implications for emerging infectious disease. Zoologia 27:151–162.

Araujo, S. B. L., M. P. Braga, D. R. Brooks, S. J. Agosta, E. P. Hoberg, F. W. von Hartenthal, and W. A. Boeger. 2015. Understanding host-switching by ecological fitting. PLoS ONE 10:e0139225.

Barbosa, A. M., J. A. Brown, A. Jimenez-Valverde, and R. Real. 2016. modEvA: model evaluation and analysis. R package version 1.3.2.

Bolnick, D. I., and S. P. Otto. 2013. The magnitude of local adaptation under genotype-dependent dispersal. Ecol. Evol. 3:4722–4735.

Braga, M. P., S. B. L. Araujo, and W. A. Boeger. 2014. Patterns of interaction between Neotropical freshwater fishes and their gill Monogenoidea (Platyhelminthes). Parasitol. Res. 113:481–490.

Brooks, D. R., and D. A. McLennan. 1993. Parascript: parasites and the language of evolution. Smithsonian Institution Press, Washington, DC.

———. 2002. The nature of diversity. Chicago Univ. Press, Chicago.

Calatayud, J., J. L. Hórreo, J. Madrigal-González, A. Migeon, M. Á. Rodríguez, S. Magalhães, and J. Hortal. 2016. Geography and major host evolutionary transitions shape the resource use of plant parasites. Proc. Natl. Acad. Sci. USA 113:9840–9845.

Cardon, M., G. Loot, G. Grenouillet, and S. Blanchet. 2011. Host characteristics and environmental factors differentially drive the burden and pathogenicity of an ectoparasite: a multilevel causal analysis. J. Anim. Ecol. 80:657–667.

de Aguiar, M. A. M., M. Baranger, E. M. Baptestini, L. Kaufman, and Y. Bar-Yam. 2009. Global patterns of speciation and diversity. Nature 460:384–387.

Dennis, R. L. H., L. Dapporto, S. Fattorini, and L. M. Cook. 2011. The generalism-specialism debate: the role of generalists in the life and death of species. Biol. J. Linnean Soc. 104:725–737.

Dethier, V. G. 1954. Evolution of feeding preferences in phytophagous insects. Evolution 8:33.

Ehrlich, P. R., and P. H. Raven. 1964. Butterflies and plants: a study in coevolution. Evolution 18:586.

Ellis, V. A., M. D. Collins, M. C. I. Medeiros, E. H. R. Sari, E. D. Coffey, R. C. Dickerson, C. Lugarini, J. A. Stratford, D. R. Henry, L. Merrill, et al. 2015. Local host specialization, host-switching, and dispersal shape the regional distributions of avian haemosporidian parasites. Proc. Natl. Acad. Sci. USA 112:11294–11299.

Fordyce, J. A. 2015. ordiBreadth: ordinated diet breadth. R package version 1.0. https://CRAN.R-project.org/package=ordiBreadth.

Fordyce, J. A., C. C. Nice, C. A. Hamm, and M. L. Forister. 2016. Quantifying diet breadth through ordination of host association. Ecology 97:842–849.

Forister, M. L., V. Novotny, A. K. Panorska, L. Baje, Y. Basset, P. T. Butterill, L. Cizek, P. D. Coley, F. Dem, I. R. Diniz, et al. 2015. The global distribution of diet breadth in insect herbivores. Proc. Natl. Acad. Sci. USA 112:442–447.

Grimm, V., U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback. 2010. The ODD protocol: a review and first update. Ecol. Model. 221:2760–2768.

Grimm, V., U. Berger, F. Bastiansen, and S. Eliassen. 2006. A standard protocol for describing individual-based and agent-based models. Ecol. Model. 198:115–126.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135:147–186.

Hamm, C. A., and J. A. Fordyce. 2015. Patterns of host plant utilization and diversification in the brush-footed butterflies. Evolution 69:589–601.

Hardenberg, von, A., and A. Gonzalez Voyer. 2013. Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. Evolution 67:378–387.

Hardy, N. B. 2017. Do plant-eating insect lineages pass through phases of host-use generalism during speciation and host switching? Phylogenetic evidence. Evolution 71:2100–2109.

Hardy, N. B., and S. P. Otto. 2014. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. Proc. R Soc. B Biol. Sci. 281:20132960–20132960.

Harrison, J. G., Z. Gompert, J. A. Fordyce, C. A. Buerkle, R. Grinstead, J. P. Jahner, S. Mikel, C. C. Nice, A. Santamaria, and M. L. Forister. 2016. The many dimensions of diet breadth: phytochemical, genetic, behavioral, and physiological perspectives on the interaction between a native herbivore and an exotic host. PLoS ONE 11:e0147971.

Janz, N., and S. Nylin. 1998. Butterflies and plants: a phylogenetic study. Evolution 52:486.

———. 2008. The oscillation hypothesis of host-plant range and speciation. Pp. 203–215 *in* K. J. Tilmon, ed. Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects. California Univ. Press, California.

Janz, N., M. P. Braga, N. Wahlberg, and S. Nylin. 2016. On oscillations and flutterings—a reply to Hamm and Fordyce. Evolution 70:1150–1155.

Janz, N., S. Nylin, and N. Wahlberg. 2006. Diversity begets diversity: host expansions and the diversification of plant-feeding insects. BMC Evol. Biol. 6:4–10.

Janzen, D. H. 1973. Comments on host-specificity of tropical herbivores and its relevance to species richness. Pp. 201–211 *in* V. H. Heywood, ed. Taxonomy and ecology. Academic Press, London.

Kool, J. T., A. Moilanen, and E. A. Treml. 2013. Population connectivity: recent advances and new perspectives. Landscape Ecol. 28:165–185.

Mason, P. A. 2015. On the role of host phenotypic plasticity in host shifting by parasites. Ecol. Lett. https://doi.org/10.1111/ele.12555.

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

Nylin, S., and N. Janz. 2009. Butterfly host plant range: an example of plasticity as a promoter of speciation? Evol. Ecol. 23:137–146.

Nylin, S., J. Slove, and N. Janz. 2014. Host plant utilization, host range oscillations and diversification in nymphalid butterflies: a phylogenetic investigation. Evolution 68:105–124.

Nylin, S., S. Agosta, S. Bensch, W. A. Boeger, M. P. Braga, D. R. Brooks, M. L. Forister, P. A. Hambäck, E. P. Hoberg, T. Nyman, et al. 2018. Embracing colonizations: a new paradigm for species association dynamics. Trends Ecol. Evol. 33:4–14.

Nyman, T. 2010. To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. Biol. Rev. 85:393–411.

Peña, C., and N. Wahlberg. 2008. Prehistorical climate change increased diversification of a group of butterflies. Biol. Lett. 4:274–278.

Peña, C., S. Nylin, and N. Wahlberg. 2011. The radiation of Satyrini butterflies (Nymphalidae: Satyrinae): a challenge for phylogenetic methods. Zool. J. Linnean Soc. 161:64–87.

R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecol. Lett. 8:336–352.

Savela, M. 2014. Lepidoptera and some other life forms. Available at http://www.nic.funet.fi/pub/sci/bio/life/intro.html. Accessed October 14, 2016.

Shipley, B. 2013. The AIC model selection method applied to path analytic models compared using a d-separation test. Ecology 94:560–564.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.

Slove, J., and N. Janz. 2011. The relationship between diet breadth and geographic range size in the butterfly subfamily Nymphalinae–a study of global scale. PLoS ONE 6:e16057.

Stevens, P. F. 2011. Angiosperm phylogeny website. Available at http://www.mobot.org/MOBOT/research/APweb/.

van der Bijl, W. 2018. phylopath: easy phylogenetic path analysis in R. PeerJ 6:e4718.

Wang, H., J. D. Holloway, N. Janz, M. P. Braga, N. Wahlberg, M. Wang, and S. Nylin. 2017. Polyphagy and diversification in tussock moths: support for the oscillation hypothesis from extreme generalists. Ecol. Evol. 7:7975–7986.

Wang, I. J., and G. S. Bradburd. 2014. Isolation by environment. Mol. Ecol. 23:5649–5662.

Wang, I. J., and K. Summers. 2010. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. Mol. Ecol. 19:447–458.

West-Eberhard, M. J. 2003. Developmental plasticity and evolution. Oxford Univ. Press, New York.

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

Associate Editor: I. Sanmartín
Handling Editor: Mohamed A. F. Noor

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Outcome of simulations after 1,000 generations.

Figure S2. Scree plot of PCoA on associations between nymphalid genera and plant orders showing the eigenvalues for each axis of the ordination performed by *ordiBreadth*.

Figure S3. Directed acyclic graphs of the tested hypothetical cause-effect models of the relationships among taxonomic host range (THR), phylogenetic host range (PHR), species richness (SP), and adjusted ordinated diet breadth (ODBadj).

Figure S4. Model selection. Weights and *p*-value for each path model tested.

Figure S5. Linear relationship between resource heterogeneity (ODBadj) and log-transformed species richness for Nymphalidae tribes, with 95% confidence interval shown as shaded area.

Table S1. Summary of preliminary analysis (Poisson regressions) on the effect of model parameters on species richness.

Table S2. Number of conditional independencies tested (*k*), *p* values of the C statistic, number of parameters estimated in each model (*q*), C statistic information criterion with correction for small sample sizes (CICc), ΔCICc, likelihoods (*li*), and CICc weights (ωi) for the eight path models depicted in Figure S3.