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

Mariana P Braga¹, Sabrina BL Araujo², Salvatore Agosta³, Daniel Brooks⁴, Eric Hoberg⁵, Sören Nylin¹, Niklas Janz¹ and Walter Boeger⁴

¹ Zoology Department, Stockholm University, 10691 Stockholm, Sweden

² Departamento de Física, Universidade Federal do Paraná, Caixa Postal 19044, Curitiba, PR 81531-980, Brazil

³ Center for Environmental Studies and Department of Biology, Virginia Commonwealth University, Richmond, VA, USA 23284

⁴ Laboratório de Ecologia Molecular e Parasitologia Evolutiva, Universidade Federal do Paraná, Caixa Postal 19073, Curitiba, PR 81531-980, Brazil

⁵ US National Parasite Collection, US Department of Agriculture, Agricultural Research Service, BARC East No. 1180, Beltsville, MD, USA 20705

Key-words: host range, individual-based model, resource heterogeneity, species richness

19 **Abstract**

20 Hosts are thought to play an important role on parasite diversification, yet little
21 consensus has been achieved about the macroevolutionary consequences of changes in host
22 use. Part of the problem might be related to the classification of host use properties, such as
23 host range, on discrete states, which precedes most analytical methods. Our goal is to offer a
24 mechanistic basis for the origins of macroevolutionary patterns of parasite diversity by
25 simulating lineages evolved in silica, where character states are an outcome of the system
26 dynamics. Here we describe an individual-based model in which (i) parasites undergo sexual
27 reproduction limited by genetic proximity, (ii) hosts are uniformly distributed along a one-
28 dimensional resource gradient, and (iii) host use is determined by the interaction between
29 the phenotype of the parasite and a heterogeneous fitness landscape. We found two main
30 effects of host use on the evolution of a parasite lineage. First, the colonization of a novel host
31 allows parasites to explore new areas of the resource space, increasing phenotypic and
32 genotypic variation. Second, hosts produce heterogeneity in the parasite fitness landscape,
33 resulting in speciation. By separating the effects of the number of hosts used by a parasite and
34 the diversity of resources they encompass, we found that resource heterogeneity, rather than
35 host range *per se*, is the main driver of parasite species richness. We see this result as an
36 important step forward in our understanding of host-driven diversification processes.

37

39 **Introduction**

40 All organisms that spend most of their lifetime on a host without killing it could be called
41 parasites (Price 1980), regardless of the trophic level where the interaction occurs (e.g.
42 phytophagous insects, ecto and endoparasites with direct or indirect life cycles). This intimate
43 ecological interaction may persist over long time spans, coupling the evolutionary history of
44 the interacting lineages and producing broad patterns of host use across parasite taxa, such as
45 conservatism in host use (Dethier 1954; Ehrlich & Raven 1964; Janz & Nylin 1998; Brooks &
46 McLennan 1993, 2002). Nonetheless, the level of intimacy or specialization of the parasites to
47 their hosts varies greatly between and within lineages. Among butterflies, for example, most
48 species use plants from a single family as hosts, but there are several species with a host
49 range that includes from two to 36 plant families (Forister et al. 2015). When placing this
50 variation in a phylogenetic context, most studies have found that host use and host range are
51 labile across time and space (Nosil 2002; Braga, Araújo & Boeger 2014; Nylin, Slove & Janz
52 2014; Calatayud et al. 2016), which shows that the two (apparently contradictory) paths of
53 fine-tuning to a host or exploring the range of potential hosts can be taken without preventing
54 the other (Agosta & Klemens 2008; Araújo et al. 2015). Adaptation and colonization processes
55 are partly independent, but jointly shape host use patterns (Ellis et al. 2015), producing a
56 complex history of diversification. This complexity is then escalated by biogeographical
57 processes (Hoberg & Brooks 2008).

58 A recent effort to integrate biogeography, ecology, and evolution is emerging as a new
59 synthesis for the origin and persistence of biodiverse systems (Brooks, Hoberg & Boeger
60 2015; Hoberg et al. 2015; Hoberg & Brooks 2015). A key part of this synthesis is the

61 recognition of ecological perturbations and the formation of new ecological interactions (e.g.
62 host colonization) via ecological fitting (Janzen 1985) as drivers of diversification and
63 community assembly. In a previous paper, we modeled the formation of new host-parasite
64 interactions and found that, satisfying a very simple set of model rules, successful colonization
65 of a new host depends mainly on opportunity and compatibility (Araújo et al. 2015).
66 According to our minimalistic model, phenotypic variation increases the range of compatible
67 hosts for parasites, but even parasites with very low phenotypic variation ('highly
68 specialized') are able to ecologically fit to some new hosts, given opportunity (Araújo et al.
69 2015). This finding, which is supported by empirical data (Agosta 2006; Gompert et al. 2015),
70 is an important step towards the understanding of the process of host range expansion, which
71 is central to the new synthesis.

72 According to one important part of the new synthesis, namely the Oscillation Hypothesis,
73 variation in host range over evolutionary time is one of the main drivers of diversification of
74 plant-feeding insects (Janz & Nylin 2008). Host range expansion, coupled with the expansion
75 of overall niche breadth and geographic range (Slove & Janz 2011; Dennis et al. 2011), is
76 expected to create opportunities for divergence by both adaptive and neutral processes (Janz
77 et al. 2016). These dynamics then produce the observed pattern of correlation between
78 diversity of host use (i.e. host range at clade level) and species richness (Janz, Nylin &
79 Wahlberg 2006). Evolutionary ecologists have traditionally focused on the causes and
80 consequences of ecological specialization, yet little consensus has been achieved about the
81 macroevolutionary consequences of specialization to diversification (as reviewed by Forister
82 et al. 2012). However, there is an increased realization that changes in host range can happen
83 both via specialization and generalization, but the importance of generalization processes for

84 diversification of phytophagous insects is still under debate (Hamm & Fordyce 2015; 2016;
85 Janz et al. 2016).

86 Novel statistical approaches to studying diversification have been developed recently
87 (Maddison, Midford & Otto 2007; FitzJohn 2012), but so far have produced divergent results
88 and, consequently, different explanations for the changes in diversification rates across
89 parasite phylogenies (Hardy & Otto 2014; Hamm & Fordyce 2015; Hardy, Peterson &
90 Normark 2016). Part of this problem may lie in the classification of host range into two
91 opposing states (or multiple states), which precedes most analytical methods. Strictly
92 speaking, host range is not an independently evolving trait in its own right, but an emergent
93 property of the underlying dynamics of gaining and losing host plants. The aim of the present
94 paper is to offer a mechanistic basis for the origins of macroevolutionary patterns of parasite
95 diversity and host range, by studying lineages evolved *in silico* where character states are an
96 outcome of the system dynamics. In order to do that, we developed an individual-based model
97 so that key species-level properties, such as host range, can emerge from the system dynamics
98 rather than being set as parameters (Rossberg et al. 2006). In short, our simulations consist of
99 parasite individuals evolving in a fixed fitness landscape where speciation is allowed. We are
100 particularly interested in the role of changes in host range on parasite evolution and
101 diversification. Specifically, how the colonization of new hosts affects the parasite phenotypic
102 distribution, its ability to use novel resources, and the likelihood of speciation.

103

104

105 **Methods**

106 **Model description**

107 We simulated host-use dynamics by modeling individuals of a parasite population that
108 evolves under the combined effects of sexual reproduction, mutation, natural selection, and
109 dispersal. In our model, parasite characterization derives mainly from the agent-based model
110 described by (de Aguiar et al. 2009). Parasite individuals, which are identified by a haploid
111 genotype, are hermaphrodites and undergo sexual reproduction limited by genetic
112 differentiation. Genotypes consist of binary strings of length L , and the number of mismatches
113 along two genotypes determines the genetic distance between two parasites. At each time
114 step, there is a reproduction phase when all individuals have a chance to find a compatible
115 mate. A random mate is selected for each individual, but reproduction only occurs if they have
116 a maximum genetic differentiation of g percent, which is the threshold for mate recognition.
117 This genetic restriction on mating can produce groups of individuals reproductively isolated
118 from all others, which we identify as a parasite species. Species unity is maintained by mate
119 recognition, without requiring that all species members are able to mate with all others.
120 When mate pairing is successful, the offspring replaces the parents in the population and their
121 genotype is the result of genetic inheritance from both parents, with recombination and
122 mutation. Each reproduction event has genetic crossover, where the genetic contribution of
123 each parent is randomly chosen, and mutation occurs with probability m .

124 The number of individuals generated by the population is equal to bM , where b is the
125 average number of offspring generated by an individual and M is the population size before
126 reproduction. However, population growth is limited by a carrying capacity that is
127 determined by the number of hosts used by the parasites, since each host supports a

128 maximum of K individuals. The parasite population can use N_h available hosts uniformly
 129 distributed along a one-dimensional resource gradient. The number of hosts and the distance
 130 between them define the resource space for the parasites. The location of a parasite in the
 131 resource space is determined by its phenotype, which combines, for simplicity, all parasite
 132 traits related to host use. The phenotype of parasite i is the sum of all L loci of its genotype:

$$p_i = \frac{1}{10} \sum_{n=1}^L X_{i,n}$$

133 where $X_{i,n} \in \{0,1\}$, the two possible alleles. The term $1/10$ limits the range of possible
 134 phenotypes to avoid edge effects.

135 Every simulation starts with three parasite individuals in one host at the center of the
 136 resource space. Genetic and phenotypic variation increases over time due to mutation and
 137 recombination, as long as the offspring is able to survive the selective pressure imposed by
 138 the host. Each host is modeled as a fitness peak in the fitness landscape of the parasites, which
 139 means that there is a unique phenotype value that yields maximum survival for each host (as
 140 done by (Araújo et al. 2015)). Nonetheless, hosts are not only used by perfectly adapted
 141 parasites, instead, survival decreases with the distance from the optimum, following a normal
 142 distribution centralized on $|p_{ij} - q_j|$:

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

143 where q_j is the optimal phenotype to use host j , p_{ij} is the phenotype of parasite i using host j ,
 144 and σ_r is the parameter that controls the intensity of the selective pressure. Therefore, the
 145 fitness of a given parasite depends on the host it is using, since each individual can only use
 146 one host per generation. Dispersal in the resource space occurs with probability d in every
 147 generation, so that, in average, d percent of the population tries to colonize a new host. A

148 successful host colonization happens when dispersers survive the selective pressure imposed
149 by the new host. There are no explicit trade-offs between hosts or costs for species with a
150 wide host range, i.e. using a large number of hosts.

151 **Analysis**

152 According to tests of parameter variation, the distance between hosts in the resource
153 space and the dispersal probability (d) are the main determinants of host use dynamics.
154 Variation in the remaining parameters only affects the speed of the dynamics, therefore, we
155 used a given set of parameters (see table 1) for the simulations reported herein. Pairwise
156 combinations of the distance between hosts and the dispersal probability were investigated in
157 sets of 10 simulations, each one iterated for 10000 generations. The following summary
158 statistics were calculated and averaged across the 10 replicates of each parameter
159 combination: (i) maximum host range throughout the simulation, (ii) total number of hosts
160 used (final host range), (iii) portion of the resource space covered in the final host range
161 (resource heterogeneity), and (iv) number of parasite species (species richness). Of these, (i)
162 was averaged across parasite species, while (ii)-(iv) were measured across species, at the end
163 of the simulation. Data analysis was conducted with the *PLSPM* package (Sanchez et al. 2015)
164 of R (R Core Team 2016), where we used path analysis to estimate the relationships between
165 parameters and summary statistics, with parasite species richness as the endogenous
166 variable. Exogenous variables were combined in four blocks: dispersal, distance, host range
167 (the combination of maximum host range and final host range), and resource heterogeneity.
168 Significance of path coefficients was assessed by bootstrap validation with 1000 resamples.

169

170

171 Table 1. Short description of model parameters with values used in the simulations reported.

Parameter	Short meaning
$L = 200$	Genome size
$g = 5 \%$	Maximum genetic differentiation for mate recognition
$m = 0.0001$	Mutation rate
$b = 4$	Number of offsprings
$K = 50$	Carrying capacity
$N_h = 11$	Total number of hosts in the resource space
$\sigma_r = 0.5$	Standard deviation of the survival probability function

172

173 The development of a new method to assess diet breadth (i.e., resource heterogeneity)

174 (Fordyce et al. 2016) allowed us to compare the results of our simulations with empirical

175 data. Using the *ordiBreadth* (Fordyce 2015) package of R and path analysis, we assessed the

176 direct and indirect effects of host range and resource heterogeneity on species richness for a

177 well-studied group of nymphalid butterflies. Host-plant families were distributed in the

178 ordination space (analogous to the resource space in the model) based on their association

179 with butterfly species, according to the dataset assembled by Schäpers et al (unpublished).

180 Species richness was measured at the tribe level (according to the website of Savelle 2016),

181 therefore, we calculated the ordinated diet breadth (i.e., resource heterogeneity) for each of

182 the 15 butterfly tribes present in the dataset. The host range of each tribe was calculated as a

183 combination of the taxonomic host range - i.e., total number of plant families used by species

184 of the tribe - and the phylogenetic diversity of this group of host plant families (herein

185 referred to as phylogenetic host range). Phylogenetic diversity was calculated based on the

186 phylogenetic relationships between plant families proposed by the Angiosperm Phylogeny

187 Website (Stevens 2001). As in the simulations, species richness (log transformed) was the

188 endogenous variable in the path analysis and the exogenous variables were resource

189 heterogeneity and host range. Bootstrap validation was also conducted. Since spurious
190 correlations between variables could arise from the phylogenetic relationship between
191 butterfly tribes, we tested and confirmed that none of the variables in the path model showed
192 significant phylogenetic signal.

193

194 **Results**

195 **Host use and speciation dynamics**

196 Despite the small initial population size, host range expansion and speciation happened
197 under a broad range of parameters. As the population grows, mutation and recombination
198 produce genetic and phenotypic variation, which can be seen in the first 250 generations of
199 the example simulation shown in fig. 1A and 2A. Until this point in time (which varies
200 between simulations) all parasites belong to the same species, but speciation happens in the
201 following generation, meaning that the population was divided in two reproductive units due
202 to the genetic differentiation threshold for mating. Both species continue to accumulate new
203 phenotypes until speciation happens again in one of the lineages, producing a third species.
204 Figure 1 shows the overall dynamics during the first 1000 generations of the example
205 simulation, and fig. 2 shows changes in phenotype frequency around the first speciation event
206 of the same example. All panels in fig. 1 and 2 show the phenotypic distribution of parasites
207 through time, but fig. 1B and 2B also show which host is used by each parasite individual,
208 allowing us to infer the role played by the hosts on parasite diversification.

209

210

211 Figure 1 - Parasite phenotypic distribution through time in the example simulation, where *distance*
 212 *between hosts* = 0.75 and *dispersal probability* = 0.2. Each dot represents a parasite individual and
 213 species are identified by color. In A all parasites are shown together while in B each panel shows the
 214 parasites using each given host. Numbers at the right end of the panels show the optimum phenotype
 215 to use each host.

216 Although parasites are initially perfectly adapted to the original host ($p_1 = q_1 = 10$), they
217 colonized the two closest hosts within the first 50 generations in the example shown in fig. 1B
218 and 2B. New hosts have two main effects in this system. First, phenotypic variation increases
219 when a new host is added to the repertoire because, with time, part of the population adapts
220 to the new host, extending the phenotypic distribution across both fitness peaks (e.g.,
221 generations 60-250). Second, the inclusion of a second fitness peak in the fitness landscape of
222 the parasite promotes speciation by divergent selection. Once gene flow between parasites on
223 different hosts is interrupted, each species evolves towards the optimum phenotype for each
224 host (or hosts, as in the case of species 1 in fig. 1B and 2B). It is important to notice, however,
225 that although each species specializes on their main host(s), they continue to use other hosts,
226 including ancestral ones (e.g., green dots in the central panel of fig. 1B). The continuous ability
227 to explore the resource space by using alternative hosts (hosts that yield a lower fitness than
228 the main host) coupled with adaptation to a subset of the host range due to lack of gene-flow
229 produces variation in host range through time.

230

Figure 2 - Phenotype frequency through time in the example simulation. Each panel shows the frequency for a given set of generations. Individuals are identified by species (A) and by host use (B).

4.2. Drivers of parasite species richness

Variation in distance between hosts and dispersal probability clearly affect species richness, resource heterogeneity, and host range (fig. 3). Species richness and resource heterogeneity (i.e., the greatest distance between two hosts in the parasite repertoire) are enhanced at intermediate parameter values. Host range has a similar response but with a threshold for distance between hosts, under which all hosts are colonized (fig. 3C).

Figure 3 - Variation in species richness (A), resource heterogeneity (B), and host range (C) across parameter combinations. Each square shows the average of 10 replicates for a given combination.

The path analysis shows that the interaction between these variables is complex (fig. 4A). Dispersal probability had a negligible effect on species richness, with a total effect statistically not different from zero. All other variables have a direct and/or indirect effect on parasite richness, but all positive effects go through resource heterogeneity. Increasing distance between hosts results in smaller host range; but given the same host range, the furthest away hosts are, the larger the heterogeneity in resources used by the parasite. Additionally, resource heterogeneity is positively associated with host range. Finally, host range has opposing direct and indirect effects on parasite richness: host range expansions that increase resource heterogeneity produce higher richness; conversely, adding hosts without increasing resource heterogeneity lowers parasite richness. Overall, the path model fits well to the data (GoF = 0.88), explaining most of the variation in species richness ($R^2 = 0.95$).

257

258 Figure 4 - Summary of path analyses for simulated host-parasite systems (A) and nymphaline
 259 butterflies-plant interactions (B). Blue arrows show positive effects and red arrows, negative effects.
 260 Arrow thickness is scaled approximately with the absolute path coefficients, which are shown for each
 261 path. For each path model, the overall goodness of fit and R^2 for species richness is also shown.

262

263 The data set from Shäpers et al. (unpublished) includes reliable records of 753
 264 interactions between 233 nymphaline species from 15 tribes and 238 plant genera from 42
 265 families. The ordination of diet breadth produced a high-dimensional resource space, where
 266 the first three principal coordinates explain respectively, 10.4%, 6.5%, and 5.8% of the
 267 variation (fig. S1). Species richness varies greatly across nymphaline tribes (7 - 257 species),
 268 as does taxonomic host range (1 - 17 plant families). Resource heterogeneity (or ordinated
 269 diet breadth *sensu* Fordyce 2016) of nymphaline tribes is so correlated to taxonomic host
 270 range that we can not tease apart their effects on parasite richness. We therefore added
 271 phylogenetic diversity as a component of host range, producing a path model with high

272 prediction power (GoF = 0.88) that explains almost 60% of the variation in species richness
273 (fig. 4B). Interactions between host range, resource heterogeneity, and species richness show
274 the same pattern with data from nymphaline butterflies and from our model simulations.
275 Species richness is higher in tribes that use heterogeneous resources, which is often
276 associated with a larger host range. On the other hand, tribes that use hosts that are very
277 close in the resource space are not species rich. The same analysis was done at plant genus
278 level and the results are qualitatively the same (fig. S2).

279

280 **Discussion**

281 Our model shows that host colonization allows parasites to explore new areas of the
282 resource space, with increasing phenotypic and genotypic variation (fig. 1 and 2). The
283 incorporation of a novel host starts by ecological fitting: some parasite individuals have
284 positive fitness on the novel host even if they are better adapted to the original host. Because
285 the novel host represents a new region in the fitness landscape, it allows more parasite
286 variants to (co)exist, increasing variation in the population (c.f. Agosta & Klemens 2008).
287 Greater phenotypic amplitude (which translates to greater resource heterogeneity) increases
288 the probability of successful host colonizations (Araújo et al. 2015), closing a positive
289 feedback loop between host range expansion and increasing resource heterogeneity. This
290 loop is then broken by speciation.

291 In a system where individuals undergo sexual reproduction limited by genetic proximity,
292 speciation depends on clustering of genetic variants (de Aguiar et al. 2009). In our model,
293 genetic clusters are formed because the fitness landscape is heterogeneous, even though the
294 resource space and the associated fitness landscape modeled here are very simplified. In

295 nature, hosts differ in various traits - e.g., nutrients, defenses, mutualists - producing more
296 heterogeneity in the resource space and potentially in the fitness landscape of parasites
297 (Nyman 2010). As predicted by Nyman (2010), speciation was maximized when hosts were at
298 an intermediate distance in resource space (fig. 3A), balancing the probabilities of
299 colonization and divergent selection. Colonization happens quickly when hosts are very
300 similar (fig. 3C) but divergent selection is stronger when hosts are distant. On the other hand,
301 dispersal probability had little effect on the outcome of the simulations, with negligible effect
302 on species richness. Dispersal can have a positive effect on host colonization - and therefore,
303 on parasite species richness - but high levels of dispersal lead to higher mortality when
304 dispersers encounter incompatible hosts, and also promote gene-flow between genetic
305 clusters, which prevents differentiation.

306 According to the path analysis (fig. 4A), distance between hosts along the resource space
307 affects both the host range and the heterogeneity among resources used by parasites: host
308 range is narrow when hosts are too different, but if parasites eventually colonize distant
309 hosts, resource heterogeneity quickly increases. Therefore, resource heterogeneity is
310 maximized when a parasite uses a wide range of distant hosts. In turn, resource heterogeneity
311 emerged as the main driver of parasite species richness. Regardless of the other variables,
312 species richness increases with resource heterogeneity, which is a result of the phenotypic
313 variation in the parasite lineage. There is no upper limit for phenotypic/genotypic variation
314 within a species in the model; that is, genetic clusters that differ more than the threshold for
315 mate recognition still belong to the same species as long as individuals with intermediate
316 genotypes connect them. However, the greater the variation within a species, the higher the
317 likelihood of fragmentation. Interestingly, in some cases (when host range does not affect
318 resource heterogeneity), a wide host range can actually prevent fragmentation. The key to

319 this perhaps counterintuitive negative association between host range and species richness is
320 heterogeneity in the fitness landscape. Adding more hosts without increasing resource
321 heterogeneity leads to a more homogeneous fitness landscape. Therefore, in this case,
322 parasite lineages with more hosts will be less species-rich because parasite clusters
323 associated to hosts at the edges of the resource space will be connected by parasites on
324 intermediate hosts. In the extreme case where the distance between hosts is minimal, the
325 fitness landscape is homogeneous and the unique parasite species is uniformly distributed
326 along a narrow resource space.

327 In nature, host-parasite systems are more complex than modeled here: parasite
328 individuals have flexible phenotypes (West-Eberhard 2003; Nylin & Janz 2009), hosts are
329 geographically distributed (Calatayud et al. 2016), the distance between hosts in the resource
330 space is not uniform (Nyman 2010), and the resource space is multidimensional (Harrison et
331 al. 2016; Fordyce et al. 2016). Still, we recovered the same pattern of interactions between
332 host range, resource heterogeneity, and species richness for modeled parasites and
333 nymphaline butterflies (fig. 4). Nymphaline tribes with wider aggregated host range
334 (taxonomic + phylogenetic) are more species-rich when the wide host range translates into
335 high resource heterogeneity (fig. 4B). That is the case of the speciose tribes Melitaeini,
336 Nymphalini, and Junoniini, which contain 257, 102, and 95 species respectively. But when
337 comparing tribes with similar resource heterogeneity, as these three, host range has a
338 negative effect on species richness (fig. 4B), driven mainly by the phylogenetic component of
339 host range. Nymphalini and Junoniini are the tribes with wider host ranges (both taxonomic
340 and phylogenetic), while Melitaeini - the most speciose tribe - has a comparatively low
341 phylogenetic host range. At least part of Melitaeini species richness seems to be explained by

342 the use of a group of plant families that represent a higher resource heterogeneity than
343 expected by its phylogenetic diversity (fig. S3).

344 Resource heterogeneity across nymphaline butterflies is not completely explained by
345 taxonomic and/or phylogenetic host range, giving the opportunity to disentangle their effects
346 on diversification. Our ability to find the mechanisms by which host use affects diversification
347 of parasite lineages depends on our knowledge about how latin binomials translate into
348 resources for parasites (Janzen 1973; Brooks & McLennan 2002; Harrison et al. 2016).
349 Although there is much yet to learn, we view the results of the present paper as an important
350 step forward in the development of a theoretical framework for the study of host-parasite
351 associations, such as the patterns described and predicted by the Oscillation Hypothesis (Janz
352 & Nylin 2008). On theoretical and empirical grounds, this study highlights the importance of
353 the differentiation between host range and resource heterogeneity, with the latter having the
354 main direct effect on diversification. As suggested by the Oscillation Hypothesis, clades with
355 wider aggregated host ranges are in general more species-rich, but here we show that this
356 interaction is likely to be mediated by resource heterogeneity. Host range expansions lead to
357 diversification, as long as they increase heterogeneity in the resource space, and
358 consequently, in the fitness landscape. Moreover, our results might be a bridge between
359 contrasting processes of diversification of phytophagous insects, and maybe parasites in
360 general. In the nymphaline butterflies, resource heterogeneity is mainly driven by the use of
361 unrelated host plants, hence the indirect positive interaction between host range and species
362 richness. However, in other groups, resource heterogeneity might be high even if hosts are
363 closely related, as long as the host plant clade is heterogeneous regarding at least one
364 resource axis, such as habitat, growth form, geographic distribution, or phenology. The grass-
365 feeding Satyrini is one such example, where colonization of a phylogenetically restricted plant

366 group that is instead very diverse in terms of e.g. habitat types, has resulted in an impressive
367 diversification. One important aspect to investigate in future studies, we believe, is how
368 different kinds of resource heterogeneity relate to phylogenetic host range and how that
369 affects diversification.

370

371 **References**

- 372 Agosta, S.J. (2006) On ecological fitting, plant-insect associations, herbivore host shifts, and
373 host plant selection. *Oikos*, 114, 556–565.
- 374 Agosta, S.J. & Klemens, J.A. (2008) Ecological fitting by phenotypically flexible genotypes:
375 implications for species associations, community assembly and evolution. *Ecology Letters*,
376 11, 1123–1134.
- 377 Araújo, S.B.L., Braga, M.P., Brooks, D.R., Agosta, S.J., Hoberg, E.P., Hartenthal, von, F.W. &
378 Boeger, W.A. (2015) Understanding Host-Switching by Ecological Fitting. (ed UG
379 Munderloh). *PLoS ONE*, 10, e0139225.
- 380 Braga, M.P., Araújo, S.B.L. & Boeger, W.A. (2014) Patterns of interaction between Neotropical
381 freshwater fishes and their gill Monogenoidea (Platyhelminthes). *Parasitology Research*,
382 113, 481–490.
- 383 Brooks, D.R., Hoberg, E.P. & Boeger, W.A. (2015) In the Eye of the Cyclops: The Classic Case of
384 Cospeciation and Why Paradigms are Important. *Comparative Parasitology*, 82, 1–8.
- 385 Brooks, D.R. & McLennan, D.A. (1993) *Parascript: Parasites and the Language of Evolution*.
386 Smithsonian Institution Press, Washington, DC.
- 387 Brooks, D.R. & McLennan, D.A. (2002) *The Nature of Diversity: An Evolutionary Voyage of*
388 *Discovery*. University of Chicago Press, Chicago.
- 389 Calatayud, J., Hórreo, J.L., Madrigal-González, J., Migeon, A., Rodríguez, M.Á., Magalhães, S. &
390 Hortal, J. (2016) Geography and major host evolutionary transitions shape the resource
391 use of plant parasites. *Proceedings of the National Academy of Sciences*, 113, 201608381–
392 9845.
- 393 de Aguiar, M.A.M., Baranger, M., Baptestini, E.M., Kaufman, L. & Bar-Yam, Y. (2009) Global
394 patterns of speciation and diversity. *Nature*, 460, 384–387.
- 395 Dennis, R.L.H., Dapporto, L., Fattorini, S. & Cook, L.M. (2011) The generalism-specialism
396 debate: the role of generalists in the life and death of species. *Biological Journal of the*
397 *Linnean Society*, 104, 725–737.
- 398 Dethier, V.G. (1954) Evolution of Feeding Preferences in Phytophagous Insects. *Evolution*, 8,
399 33.
- 400 Ehrlich, P.R. & Raven, P.H. (1964) Butterflies and plants: a study in coevolution. *Evolution*, 18,
401 586.
- 402 Ellis, V.A., Collins, M.D., Medeiros, M.C.I., Sari, E.H.R., Coffey, E.D., Dickerson, R.C., Lugarini, C.,
403 Stratford, J.A., Henry, D.R., Merrill, L., Matthews, A.E., Hanson, A.A., Roberts, J.R., Joyce, M.,
404 Kunkel, M.R. & Ricklefs, R.E. (2015) Local host specialization, host-switching, and
405 dispersal shape the regional distributions of avian haemosporidian parasites. *Proceedings*
406 *of the National Academy of Sciences*, 112, 11294–11299.

- 407 FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R.
408 *Methods in Ecology and Evolution*, 3, 1084–1092.
- 409 Fordyce J (2015). ordiBreadth: Ordinated Diet Breadth. R package version 1.0.
410 <https://CRAN.R-project.org/package=ordiBreadth>
- 411 Fordyce, J.A., Nice, C.C., Hamm, C.A. & Forister, M.L. (2016) Quantifying diet breadth through
412 ordination of host association. *Ecology*, 97, 842–849.
- 413 Forister, M.L., Dyer, L.A., Singer, M.S., Stireman, J.O.I. & Lill, J.T. (2012) Revisiting the evolution
414 of ecological specialization, with emphasis on insect-plant interactions. *Ecology*, 93, 981–
415 991.
- 416 Forister, M.L., Novotny, V., Panorska, A.K., Baje, L., Basset, Y., Butterill, P.T., Cizek, L., Coley,
417 P.D., Dem, F., Diniz, I.R., Drozd, P., Fox, M., Glassmire, A.E., Hazen, R., Hrcek, J., Jahner, J.P.,
418 Kaman, O., Kozubowski, T.J., Kursar, T.A., Lewis, O.T., Lill, J., Marquis, R.J., Miller, S.E.,
419 Morais, H.C., Murakami, M., Nickel, H., Pardikes, N.A., Ricklefs, R.E., Singer, M.S., Smilanich,
420 A.M., Stireman, J.O., Villamarín-Cortez, S., Vodka, S., Volf, M., Wagner, D.L., Walla, T.,
421 Weiblen, G.D. & Dyer, L.A. (2015) The global distribution of diet breadth in insect
422 herbivores. *Proceedings of the National Academy of Sciences*, 112, 442–447.
- 423 Gompert, Z., Jahner, J.P., Scholl, C.F., Wilson, J.S., Lucas, L.K., Soria-Carrasco, V., Fordyce, J.A.,
424 Nice, C.C., Buerkle, C.A. & Forister, M.L. (2015) The evolution of novel host use is unlikely
425 to be constrained by trade-offs or a lack of genetic variation. *Molecular Ecology*, 24, 2777–
426 2793.
- 427 Hamm, C.A. & Fordyce, J.A. (2015) Patterns of host plant utilization and diversification in the
428 brush-footed butterflies. *Evolution*, 69, 589–601.
- 429 Hamm, C.A. & Fordyce, J.A. (2016) Greater host breadth still not associated with increased
430 diversification rate in the Nymphalidae—A response to Janz et al. *Evolution*, 70, 1156–
431 1160.
- 432 Hardy, N.B. & Otto, S.P. (2014) Specialization and generalization in the diversification of
433 phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proceedings*
434 *of the Royal Society B: Biological Sciences*, 281, 20132960–20132960.
- 435 Hardy, N.B., Peterson, D.A. & Normark, B.B. (2016) Nonadaptive radiation: Pervasive diet
436 specialization by drift in scale insects? *Evolution*, 70, 2421–2428.
- 437 Harrison, J.G., Gompert, Z., Fordyce, J.A., Buerkle, C.A., Grinstead, R., Jahner, J.P., Mikel, S., Nice,
438 C.C., Santamaria, A. & Forister, M.L. (2016) The Many Dimensions of Diet Breadth:
439 Phytochemical, Genetic, Behavioral, and Physiological Perspectives on the Interaction
440 between a Native Herbivore and an Exotic Host (ed WJ Etges). *PLoS ONE*, 11, e0147971.
- 441 Hoberg, E.P. & Brooks, D.R. (2008) A macroevolutionary mosaic: episodic host-switching,
442 geographical colonization and diversification in complex host-parasite systems. *Journal of*
443 *Biogeography*, 35, 1533–1550.
- 444 Hoberg, E.P. & Brooks, D.R. (2015) Evolution in action: climate change, biodiversity dynamics

445 and emerging infectious disease. *Philosophical transactions of the Royal Society of London.*
446 *Series B, Biological sciences*, 370, –20130553.

447 Hoberg, E.P., Agosta, S.J., Boeger, W.A. & Brooks, D.R. (2015) An integrated parasitology:
448 revealing the elephant through tradition and invention. *Trends in Parasitology*, 31, 128–
449 133.

450 Janz, N. & Nylin, S. (1998) Butterflies and plants: a phylogenetic study. *Evolution*, 52, 486.

451 Janz, N. & Nylin, S. (2008) The oscillation hypothesis of host-plant range and speciation. pp
452 203–215 in K.J. Tilmon, ed. *Specialization, speciation, and radiation: the evolutionary*
453 *biology of herbivorous insects*. Univ. of California Press, California.

454 Janz, N., Braga, M.P., Wahlberg, N. & Nylin, S. (2016) On oscillations and flutterings—A reply to
455 Hamm and Fordyce. *Evolution*, 70, 1150–1155.

456 Janz, N., Nylin, S. & Wahlberg, N. (2006) Diversity begets diversity: host expansions and the
457 diversification of plant-feeding insects. *BMC Evolutionary Biology*, 6, 4–10.

458 Janzen, D.H. (1973) Comments on host-specificity of tropical herbivores and its relevance to
459 species richness. *Taxonomy and Ecology* (ed. by V. H. Heywood), pp. 201–211. Academic
460 Press, London.

461 Janzen, D.H. (1985) On ecological fitting. *Oikos*, 45, 308–310.

462 Maddison, W.P., Midford, P.E. & Otto, S.P. (2007) Estimating a Binary Character's Effect on
463 Speciation and Extinction. *Systematic Biology*, 56, 701–710.

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

466 Nylin, S. & Janz, N. (2009) Butterfly host plant range: an example of plasticity as a promoter of
467 speciation? *Evolutionary Ecology*, 23, 137–146.

468 Nylin, S., Slove, J. & Janz, N. (2014) Host plant utilization, host range oscillations and
469 diversification in nymphalid butterflies: a phylogenetic investigation. *Evolution*, 68, 105–
470 124.

471 Nyman, T. (2010) To speciate, or not to speciate? Resource heterogeneity, the subjectivity of
472 similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding
473 insects. *Biological Reviews*, 85, 393–411.

474 Price, P.W. (1980). *Evolutionary biology of parasites*, Princeton, N.J. : Princeton University
475 Press.

476 R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation
477 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

478 Rossberg, A., H. Matsuda, T. Amemiya, and K. Itoh. (2006). Food webs: experts consuming
479 families of experts. *Journal of Theoretical Biology* 241:552–563.

480 Sanchez G, L. Trinchera & G. Russolillo (2015). plspm: Tools for Partial Least Squares Path
 481 Modeling (PLS-PM). R package version 0.4.7. <https://CRAN.R-project.org/package=plspm>
 482 Savela, M. (2016) Lepidoptera and some other life forms. Available at
 483 <http://www.nic.funet.fi/pub/sci/bio/life/intro.html>. Accessed October 14, 2016.
 484 Slove, J. & Janz, N. (2011) The Relationship between Diet Breadth and Geographic Range Size
 485 in the Butterfly Subfamily Nymphalinae – A Study of Global Scale (ed M Briffa). *PLoS ONE*,
 486 6, e16057.
 487 Stevens, P. F. (2001) Angiosperm phylogeny website. Available at
 488 <http://www.mobot.org/MOBOT/research/APweb/>. Accessed April, 2016.
 489 West-Eberhard, M.J. (2003) *Developmental Plasticity and Evolution*. Oxford U Press, New York.
 490

Supplementary figures

Figure S1 - Scree plot of PCoA on nymphaline butterflies-plant families associations showing the eigenvalues for each axis of the ordination.

$$\text{GoF} = 0.87; R^2_{\text{rich}} = 0.59$$

Figure S2 - Summary of path analysis for interactions between nymphaline tribes and plant genera. Blue arrows show positive effects and red arrows, negative effects. Arrow thickness is scaled approximately with the absolute path coefficients, which are shown for each path. The overall goodness of fit and R^2 for species richness is also shown.

Figure S3 - Linear relationship between resource heterogeneity and phylogenetic host range. We highlighted the most speciose nymphaline tribes. Note that Melitaeini is above the line, which means that it feeds on hosts that represent higher resource heterogeneity than expected by the phylogenetic host range.