Sympatric ecological divergence with coevolution of niche preference

Pavel Payne[1,2], Jitka Polechová[3]

[1] Charles University in Prague, Department of Zoology
[2] Institute of Vertebrate Biology, Czech Academy of Sciences
[3] University of Vienna, Faculty of Mathematics

Keywords

Sympatric speciation, maintenance of polymorphism, polygenic trait, Levene model, epistasis, trade-off

Abstract

Reinforcement, the increase of assortative mating driven by selection against unfit hybrids, is conditional on pre-existing divergence. Yet, for ecological divergence to precede the evolution of assortment, strict symmetries between fitnesses in niches must hold, and/or there must be low gene flow between the nascent species. It has thus been argued that conditions favouring sympatric speciation are rarely met in nature. Indeed, we show that under disruptive selection, violating symmetries in niche sizes and increasing strength of the trade-off in selection between the niches quickly leads to loss of genetic variation, instead of evolution of specialists. The region of the parameter space where polymorphism is maintained further narrows with increasing number of loci encoding the diverging trait and the rate of recombination between them. Yet, evolvable assortment and pre-existing assortment both substantially broaden the parameter space within which polymorphism is maintained. Notably, pre-existing niche preference speeds up further increase of assortment, thus facilitating reinforcement in the later phases of speciation. We conclude that in order for sympatric ecological divergence to occur, niche preference must co-evolve throughout the divergence process. Even if populations come into secondary contact, having

diverged in isolation, niche preference substantially broadens the conditions for coexistence in sympatry and completion of the speciation process.

Introduction

Darwin observed that many closely related species occupied the same habitat. However, he considered the sympatric origin of species by ecological divergence due to an advantage of specialists over generalists rather unlikely [1]. Since then, the contribution of sympatric speciation to observed biodiversity has remained controversial [2–5]. Speciation in a well-mixed, panmictic population is difficult for two main reasons. First, gene flow and recombination oppose divergence in polygenic traits as well as preventing reinforcement (the build-up of associations between the loci contributing to pre- and post-zygotic isolation; [6,7]. Second, the diverging types must be able to coexist. It is often thought that ecological divergence must precede the evolution of mating assortment [8], as assortment is reinforced by selection against maladapted hybrids. However, this order of events is not necessary: Assortment can precede divergence, or co-evolve with ecological loci. It is the combined effect of assortment and ecological divergence which contributes to the isolation of the nascent species [9–11]. Yet, in contrast to ecological selection, the loss of fitness due to missed heterospecific matings is typically compensated by an increase in conspecific matings. This stabilises polymorphism at the loci determining assortment and can protect polymorphism at other loci diverging under natural selection.

Habitat choice is an important source of assortment when mating occurs locally. Host-specific mating is prevalent in nature, as in phytophagous insects (reviewed in [12]). Habitat fidelity based on a preference for the hatching site is common for birds [13], another group rich in sympatric species. In general, habitat choice can be based on learned characteristics of the hatching site, on specific preference loci, or on an association with another (ecological) phenotype [14]. The first two drove the classic sympatric speciation process in the experiment by Rice and Salt [15], where fruit flies learned to choose based on phototaxis and chemotaxis. In this experiment, 60 mated females from each of two extreme habitats founded the next generation. Severe disruptive selection on multiple traits in a fairly large population, coupled with independent regulation within equally sized niches, led to nearly

complete reproductive isolation over 30 generations. Effectively, behavioural allopatry evolved (c.f. [8]). Further divergence in other traits, both selective and neutral, could have followed. While the plausibility of sympatric speciation is undisputed, how common can it really be? Selection is rarely that severe in nature, and even if niches are independently regulated, they will very rarely be perfectly symmetric.

A recognition and preference for the correct habitat is a common example of assortment by association with another phenotype. While preference for food source is omnipresent, strong [16] trade-offs in specialization of the sympatric-species-rich phytophagous insects have been hard to find, triggering a substantial controversy [17,18]. Yet, strong trade-offs must be reasonably common – otherwise, all habitats would be colonised by one generalist species. Indeed, such trade-offs have often been recovered when life-time fitness and/or performance of F1 hybrids on both hosts is assessed [19–23]. Moreover, a preference-performance correlation seems to be common in herbivorous insects [24,25].

Here we use mathematical modelling to investigate how genetically encoded preferences for the most favourable habitat facilitate sympatric speciation by ecological divergence. We assume that there is a strong (convex) trade-off in a polygenic trait, leading to disruptive selection such that generalists have a lower mean fitness than specialists. Assortment arises via preference for the correct niche. Because regulation is independent within niches (soft selection, [26]), the frequency of a type well adapted to its niche rises faster when this type is rare than when it is common. This protects the polymorphism in the ecological loci (coexistence of the specialists) via negative frequency-dependence [26].

We consider a polygenic trait under ecological divergence, and focus on assortment via preference for the niche the individual is best adapted to. The assortment allele can therefore be seen as a type of a 'modifier-allele' [8], realized via niche choice. The population-wide matching of the habitat thus increases with the frequency of this modifier allele. The modifier represents an 'one-allele model' [7], where recombination cannot break down the allele-specific association with the choice-locus. Thereby, the constraints on symmetries, which render sympatric speciation implausible, are relaxed (c.f. [7,27]). To our knowledge, despite recent theoretical development (reviewed in [16]), evolution under

divergent selection (strong trade-off) and evolving niche recognition has not been analysed beyond early 2-locus models [8,28]. Note that in the studies of Ravigné et al. [16,29], which focus on the difference between local regulation (Levene model [30]) and global regulation (Dempster model [31]), 'matching' habitat choice does not evolve. Here we emphasise how the conditions for speciation tighten as the number of loci increases, and how the parameter range widens again if there is some pre-existing niche recognition ('live where you thrive') and initial divergence in the ecological loci, as would be expected under secondary contact.

How do two nascent species evolve towards coexistence? Coexistence during divergence in sympatry is often overlooked in speciation models, side-lined by focus on growth of associations between and within polygenic traits, assuming symmetries are protected [11]. The importance of pre-existing ecological divergence and/or of pre-existing assortment for further diversification in sympatry has not been much studied. Hence, it is not clear whether divergence and coexistence in sympatry is considerably more difficult than partial divergence in allopatry or broad parapatry, followed by sympatry as assortment increases. When neutral divergence in allopatry or broad parapatry does not lead to a sufficient niche separation then (similarly to instant assortment by chromosomal variations; [32]) coexistence will not be possible. Yet, coexistence in sympatry is essential: without it, every speciation in sympatry or parapatry would lead to a subdivision of the existing range – and through time, species' ranges would become ever smaller.

Methods and results

Model

Since we focus on the effects of disruptive selection on sympatric divergence, we use a haploid biallelic version of Levene's [30] model with mating within niches, which is more favourable to speciation [33]. In the lifecycle of the model, individuals with non-overlapping generations are – in the basic model without assortment – randomly distributed into two niches, where selection acts. Alleles of the first type (capital letters) are beneficial in one niche and alleles of the second type (small letters) in the other niche. The individuals who survive mate within the niches and then each niche contributes to a common pool

proportionally to its size. From the common pool, individuals are again randomly distributed into the two niches.

For further analysis, it is convenient to define fitness of the genotypes to be between zero and one. The presence of an allele which is deleterious in one of the niches reduces the fitness of its carrier by selection coefficient s, which is equal for all loci and both niches. The epistatic coefficient ϵ determines how much the intermediate generalist genotypes (e.g., *aB* and *Ab*) are disadvantaged relative to the specialist genotypes (e.g., *AB* and *ab* for $\epsilon < 0$) and is defined as the deviation from additivity of deleterious alleles in each niche if more than one of these alleles is present. Here we are assuming a special kind of symmetric epistasis, which acts among alleles deleterious in one niche and also those in the other niche. Fitnesses of all genotypes considered in this study are defined in detail further in the text and Tables 1 – 4.

Analytical solutions for polymorphic equilibria of our model are generally impossible to find. Instead, we have analysed the instability of the two monomorphic equilibria for allele frequencies in ecological and niche preference loci at 1 and 0. We assume that an equilibrium polymorphic at the ecological loci exists between them. This approach was used by Levene [30], Prout [34] and Hoekstra et al. [35]. Gliddon and Strobeck [36] analysed the haploid multilocus analogue of Levene's [30] model and proved that conditions for instability of the monomorphic equilibria are also necessary and sufficient conditions for stability of the unique polymorphic equilibrium. Although this conclusion may not be valid in the presence of epistasis and is not in general valid when assortment is present, it still allows us to estimate the stability of the polymorphic equilibrium for the ecological loci in the full model for most of the parameter range. The stability is then tested numerically. We discuss a small parameter region where the local equilibrium also depends on the initial conditions. Depending on the initial allele frequencies, the system evolves either to a polymorphic equilibrium in the ecological loci, or to another equilibrium where ecological loci are fixed and the locus for assortment is polymorphic, or where all polymorphism is lost.

We complement the 2-3 locus analytical and numerical results by an individual-based haploid simulation, with up to 16 ecological loci under disruptive selection (strong trade-off),

and 10 loci which gradually change the bias in the niche choice from 0 to 1. Within niches, each individual is selected at random (with replacement) to pair up, with a probability proportional to its fitness. The recombination rate r is uniform across loci, as is the mutation, set to $\mu = 10^{-3}$. Each of the N_i pairs has exactly one offspring, the size of niche *i* is thus kept constant at N_i. The juveniles are then redistributed between habitats, according to their preference, which reflects linearly their number of matching alleles. For example, an A-B-c-M-M-m-[..]-m type will add to its preference $\mathcal{A}(\mathrm{z}) = {}^{2}/_{3}\,(\alpha_1 + \alpha_2)$, and thus choose niche 1 with a probability of $\frac{c(1+\mathcal{A}(\mathrm{z}))}{1+(2c-1)\mathcal{A}(\mathrm{z})}$, where c is the relative size of niche 1. Generations are discrete and non-overlapping.

Ecological divergence at a polygenic trait

Ecological divergence is often driven by disruptive selection acting on polygenic traits. Therefore, we first set out to analyse how increasing the number of loci influences the ability of the model to maintain polymorphism under disruptive selection. We study two to three loci analytically and for up to 16 using individual-based simulations. In particular, we focus on violation of symmetries in niche proportions, which are defined as $c = c_{\mathrm{I}} = 1 - c_{\mathrm{II}}$ for niche I and II, respectively ($\mathrm{c} \in (0,1)$). In order to be able to compare the two- ($n_{eco} = 2$) and three- ($n_{eco} = 3$) locus models, we normalise the strength of selection (s) and epistasis (ϵ) in the models with more than two ecological loci such that the mean per-trait selection and epistasis remain the same ($s^*{}_{n_{eco}} = \frac{2\,s_{2\mathrm{loci}}}{n_{eco}}$ and $\epsilon^*{}_{n_{eco}} = \frac{1}{\binom{n_{eco}}{2}}\epsilon_{2\mathrm{loci}}$). This normalisation assures that fitnesses of the specialists are always equal to $1 - 2s - \epsilon$, whilst preserving the scaling of selection independently of epistasis. This leads to a slight weakening of the trade-off as the number of loci increases (see also Fig. S1).

For simplicity, we assume all pairwise epistatic effects have the same value, and neglect higher-order interactions between alleles. Fitnesses of the individual genotypes for the two- and three-locus model are defined in Table 1A and Table 1B, respectively.

Graphical representation of the trade-offs between fitnesses of individual genotypes in niche I and niche II is shown discussed in Fig. S1 and the corresponding supplementary text.

A

genotype	niche I	niche II
AB	1	$1-2s-\epsilon$
aB, Ab	$1-s$	$1-s$
ab	$1-2s-\epsilon$	1

B

genotype	niche I	niche II
ABC	1	$1-2s-\epsilon$
ABc, AbC, aBC	$1-\frac{2}{3}s$	$1-\frac{4}{3}s-\frac{\epsilon}{3}$
Abc, aBc, abC	$1-\frac{4}{3}s-\frac{\epsilon}{3}$	$1-\frac{2}{3}s$
abc	$1-2s-\epsilon$	1

***Table 1**. **Fitnesses of individual genotypes in the model without niche preference.** A. Two-locus model. B. Three-locus model. Selection coefficient s determines how the deleterious alleles in (a, b, c) niche I and the deleterious alleles niche II (A, B, C) reduce the fitness relative to the well adapted genotype. The epistatic coefficient ϵ is defined as the deviation from additivity of the two deleterious alleles present in a genotype together. We only consider $\epsilon < 0$ as it generates a convex trade-off and hence disruptive selection. In the three-locus model the selection and epistasis are normalised as described in the main text.*

First, we show that the region of parameter space where polymorphism is maintained is highly sensitive to violation of symmetry in niche proportions, c, even if loci are completely linked. Also, increasing convexity of the trade-off (i.e., more negative ϵ, more disruptive selection) further reduces the parameter space with maintained polymorphism (Fig. 1A). Both when the loci are completely linked (r=0) and when they are freely recombining (r=0.5), there is no difference between two and three loci. The regions of parameter space with maintained polymorphism coincide and shift towards stronger selection as epistasis increases (Fig. 1C). When recombination between loci is low (r=0.01, i.e., 1cM), the regions of parameter space with maintained polymorphism shift closer towards those of the free recombination regime in the three-locus model than in the two-locus model, as in the three-locus case the per-locus strength of selection is lower (Fig. 1B). Interestingly, there is a threshold for recombination rate, above which the parameter space is independent of the

number of loci. Below we provide the analytical expressions for the boundaries of the stable regions, and the recombination threshold. The conditions are for protected polymorphism in the ecological loci, obtained by assessing local instability of the monomorphic equilibrium. Note that as we have disruptive selection with symmetric selection coefficients are, all ecological loci evolve to the same allele frequencies.

In both models with normalized selection and $\epsilon = 0$ a polymorphic equilibrium is stable if

$$\frac{1-2s}{2-2s} < c < \frac{1}{2-2s}.$$

In the case with no recombination (Fig. 1A) and $\epsilon < 0$, polymorphism is maintained if

$$\frac{2s+\epsilon-1}{2s+\epsilon-2} < c < \frac{1}{2-\epsilon-2s},$$

and with free recombination (Fig. 1C) if

$$\frac{s(1-2s-\epsilon)}{(1-s)(2s+\epsilon)} < c < \frac{s+\epsilon}{(1-s)(2s+\epsilon)}.$$

These conditions hold for the two-locus model for recombination rates $r > \frac{-\epsilon}{1-s-\epsilon}$. If recombination rate is low ($0 < r < \frac{-\epsilon}{1-s-\epsilon}$; Fig. 1B), the conditions above do not hold anymore and maintenance of polymorphism is determined by

$$\frac{(1-2s-\epsilon)(2s+\epsilon(1-r)+r(1-2s))}{(1-r)(4(1-s)s+2\epsilon(1-2s)-\epsilon^2)} < c < \frac{2s+\epsilon-r}{(1-r)(4(1-s)s+2\epsilon(1-2s)-\epsilon^2)}.$$

Since the equations describing the conditions for the three-locus model with low recombination and all other models presented from this point on exceed the width of a page, we confine them to the Supplementary material, where we also give more details on the stability analysis.

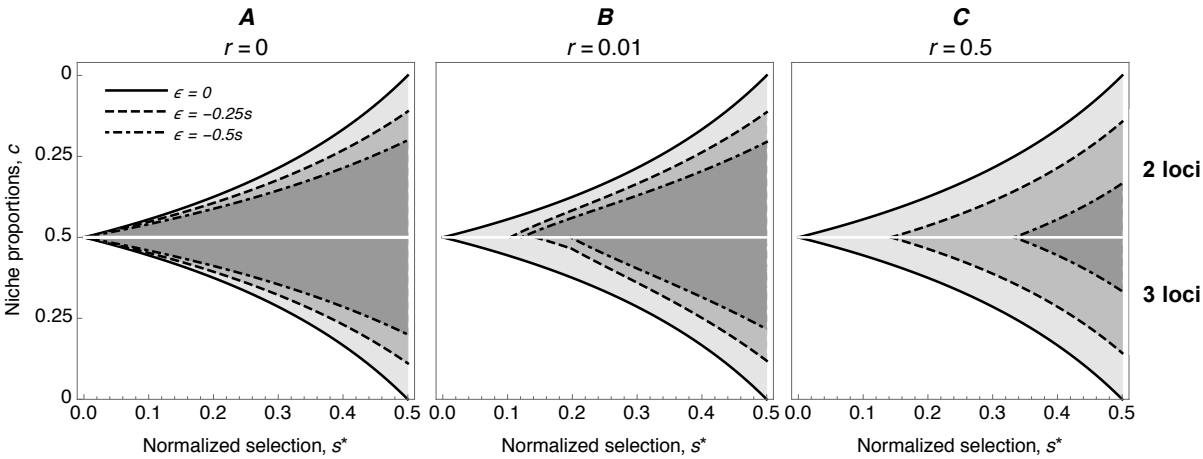

Figure 1. **Range of niche proportions where polymorphism is maintained: where ecological divergence towards two specialists is stable.** The upper half of the graphs show conditions for the two-locus model, the lower half for the three-locus model. On the x-axis is the normalized strength of selection (symmetric across loci) and on the y-axis are niche proportions. A polymorphic equilibrium is achieved for the parameter combinations of s^* and c between the black curves and the white axis at $c = 0.5$ (shaded). Note that we are showing only one half of the parameter space for each model as the conditions are symmetrical. Therefore, every condition at a value of c has its symmetric counterpart at $1 - c$. The outer solid curves represent linear trade-offs ($\epsilon = 0$), the dashed curves are for ($\epsilon = -0.25s$), and the dash-dotted curves for ($\epsilon = -0.5s$). A) No recombination; B) Low recombination ($\mathrm{r} = 0.01$), and C) Free recombination ($\mathrm{r} = 0.5$). It is noteworthy that in the regime with low recombination (B), the region of parameter space where polymorphism is maintained decreases with increasing number of loci if selection and epistais is normalized as described in the main text.

Ecological divergence with niche choice

Next, we analyse how the presence of a niche preference allele influences the sensitivity of polymorphism to violation of symmetry in niche proportions. We assume that habitat choice depends both on the presence of a choosy (modifier) allele *M* and the phenotype under natural divergent selection. For the individuals with wildtype modifier allele *m*, there is no bias in selecting niches. When the modifier allele is *M*, choice depends on the selected loci in the following manner: intermediate generalist genotypes (*Ab*, *aB*) disperse to both niches with equal probability, and specialist genotypes (*AB, ab*) choose with probability $1 + \alpha$ the niche which they are adapted to, and with probability $1 - \alpha$ the niche where they are maladapted. Since in our model mating occurs within niches, this translates into a direct modification of fitnesses of the individual genotypes, as defined in Table 2.

genotype	niche I	niche II
ABM	$1+\alpha$	$(1-2s-\epsilon)(1-\alpha)$
ABm	1	$1-2s-\epsilon$
Abm, AbM, aBm, aBM	$1-s$	$1-s$
abm	$1-2s-\epsilon$	1
abM	$(1-2s-\epsilon)(1-\alpha)$	$1+\alpha$

Table 2. *Fitnesses of individual genotypes modified by niche choice parameter α. Selection and epistasis modify fitnesses of individual genotypes as in table 1A. Genotype ABM goes to niche I with a probability increased by the factor α and to niche II decreased by α. Similarly, genotype abM preferentially goes to niche II instead of niche I.*

In the presence of niche preference, the parameter space where polymorphism is maintained broadens (Fig. 2). As allele *M* is beneficial for both specialist genotypes (*AB* and *ab*), it quickly goes to fixation when polymorphisms for the other two loci are maintained – hence, two specialists coexist. This holds in the case of complete linkage (Fig. 2A) and low recombination (Fig. 2B). In the scenario with free recombination, maintenance of polymorphism at the ecological loci depends on the initial frequencies of alleles *A, B* and *M* in the population. With weaker selection against the maladapted genotypes, s^*, the model is more sensitive to initial conditions, as indicated by the fading grey colour in the lower left part of Fig. 2C. Within this region, depending on the initial allele frequencies, either the preference allele M goes to fixation and protects polymorphism in the other loci, or the ecological loci go to fixation and allele M remains polymorphic and converges to its equilibrium frequency $\frac{|1-2c|}{\alpha}$. Figure 2D shows which combinations of initial frequencies of the ecological ($p_1 = p_2$) and preference (p_3) alleles lead to which equilibria for four points (a, b, c, d) in the parameter space in Fig. 2C.

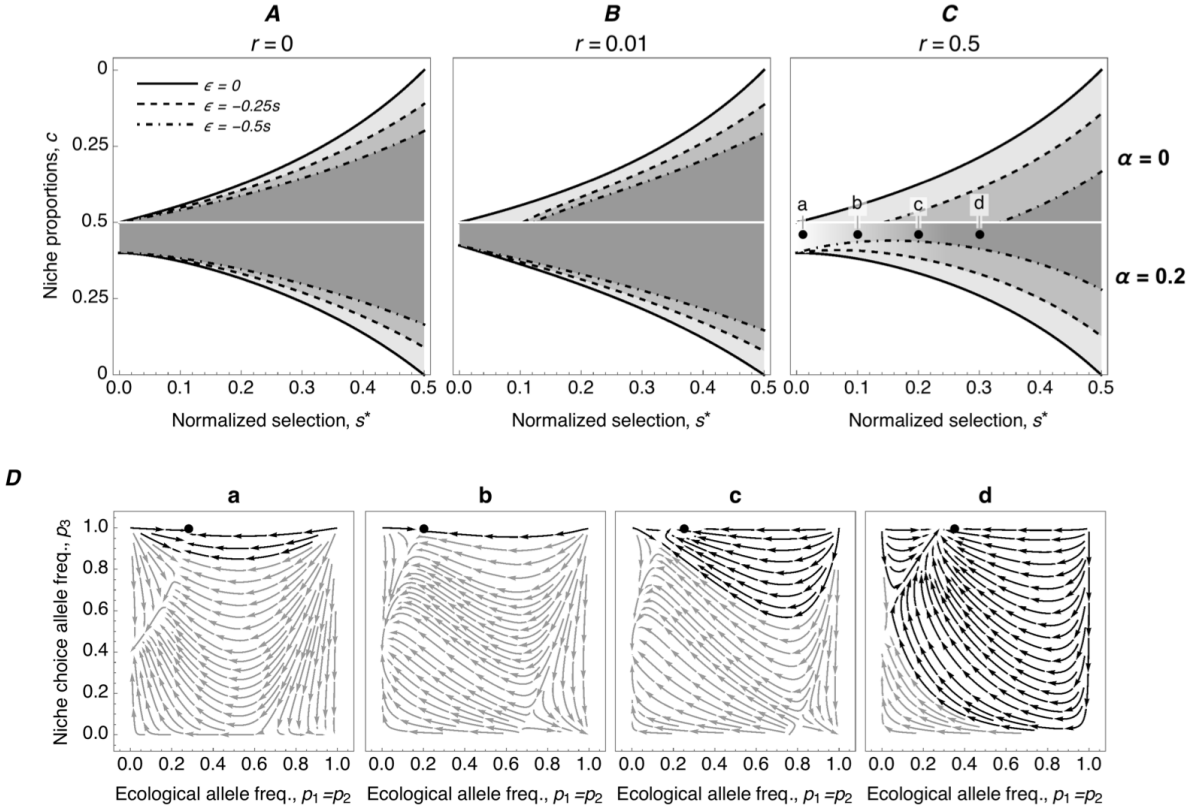

Figure 2. Ranges of niche proportions where ecological polymorphism is maintained (i.e., two specialists evolve and coexist), under coevolution of niche preference with ecological divergence.

A-C: Plots showing how niche recognition allele *M* with niche preference $\alpha = 0.2$ broadens the parameter space where polymorphism at two ecological loci is maintained, relative to the model without preference (upper half of the plot; axes and all other parameters are identical to Fig 1).

D: Streamline plots showing how maintenance of polymorphism depends on initial allele frequencies at ecological and niche preference loci for four points in Fig 2C (**a**, s=0.01; **b**, s=0.1; **c**, s=0.2; **d**, s=0.3; r=0.5, c=0.46). Black streamlines represent combinations of allele frequencies where ecological polymorphism is maintained, grey streamlines where either only ecological or all polymorphisms are lost. Since the plots were generated assuming linkage disequilibrium between ecological loci goes to 0 (i.e. loss of ecological polymorphism), the true equilibrium ecological allele frequencies, which are slightly offset with the black streamline attractors, are represented by black dots at the top of plots (**D.a-D.d**).

303 **Increase of assortment**

304 Once such a niche preference modifier allele gets fixed in a population, it not only inflates
305 the parameter space where polymorphism is maintained but it also always favours fixation
306 of another modifier allele, which then reinforces the divergent process. In order to analyse
307 such an increase of assortment, we again redefined fitnesses of the genotypes as shown in
308 Table 3. The fitnesses of the specialist genotypes (*AB* and *ab*) are now defined such that the
309 niche preference allele *M* from the previous model is fixed in the population. Therefore,
310 both specialist genotypes have their probability of going to the right niche increased by α_1.
311 The third locus can now be polymorphic for another modifier allele, which further increases
312 the probability of going to the right niche by α_2. In Fig. 3A-C we show how the parameter
313 space where polymorphism is maintained further broadens in the presence of another
314 modifier allele, increasing the assortment.

315

316 Furthermore, we numerically tested how different levels of pre-existing niche preference
317 affect time to fixation of a new niche preference allele. In Fig. 3D-G we show that with
318 increasing initial niche preference (α_1) the time to fixation of a new niche preference allele
319 ($\alpha_2 = 0.2$) decreases. In particular, this effect is most profound when selection (s^*) is weak
320 and recombination (r) is low (Fig 3D, E).

321

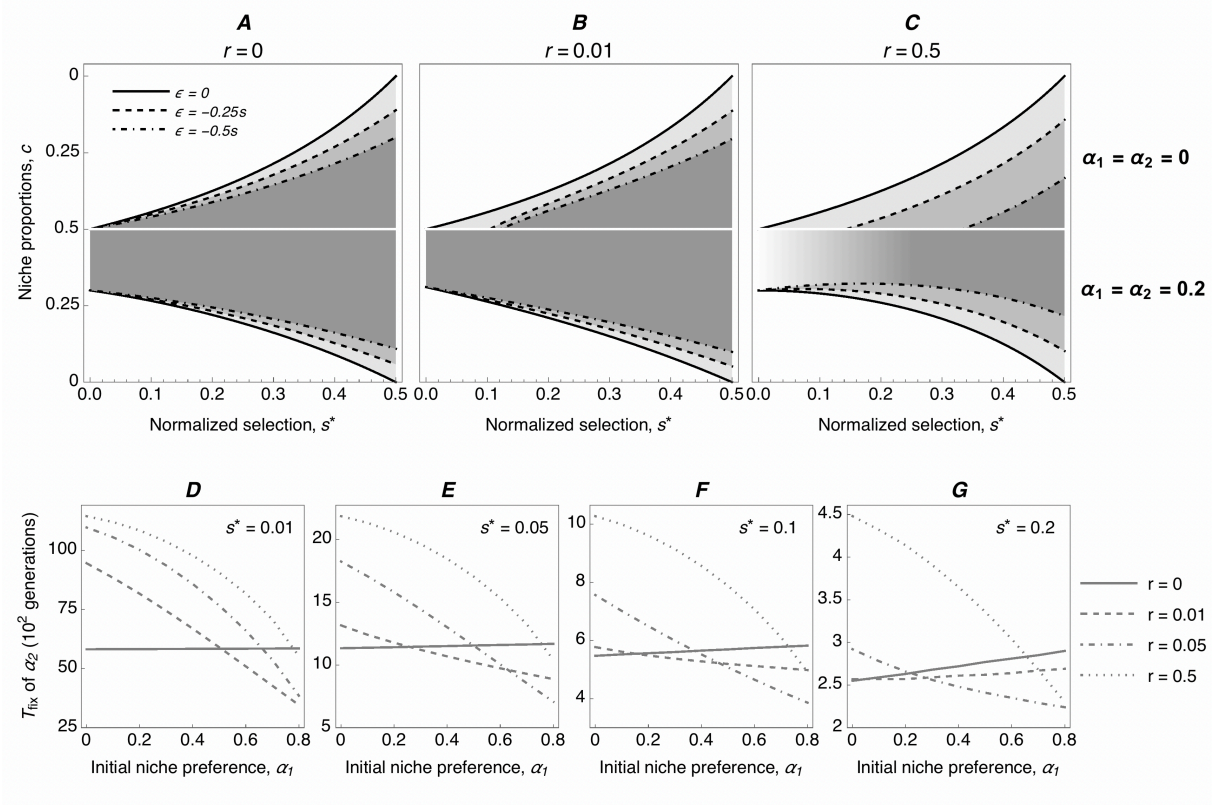

Figure 3. Range of niche proportions where polymorphism is maintained when assortment further increases. **A-C**: Plots showing how the parameter space where polymorphism at two ecological loci is maintained with already fixed niche preference $\alpha_1 = 0.2$, and an additional allele *M₂* with niche preference $\alpha_2 = 0.2$. In Figure 3C, which depicts the parameter space for free recombination ($r = 0.5$), the region where maintenance of ecological polymorphism depends on initial allele frequencies is indicated by a shading grey area (as in Fig. 2C). **D-G**: Time to fixation (y-axis), of the additional niche preference allele *M₂* with initial frequency $p_3 = 0.01$ for various combination of initial niche preference, α_1 (x-axis), strengths of selection, s^*, and recombination rates, r. All plots are for the same disruptiveness of selection, $\epsilon = -0.5s$. The time to fixation is reduced most when selection is weak ($s^* = 0.01, 0.05$), and recombination rate is low ($r = 0.01, 0.05$) (**D, E**). In contrast, with moderate or strong selection ($s^* = 0.1, 0.2$) and low recombination ($r = 0.01$) the time to fixation of the niche preference allele *M₂* either decreases very little or even slightly increases with increasing initial niche preference, α_1.

genotype	niche I	niche II
ABM_2	$1+\alpha_1+\alpha_2$	$(1-2s-\epsilon)(1-\alpha_1-\alpha_2)$
ABm_2	$1+\alpha_1$	$(1-2s-\epsilon)(1-\alpha_1)$
$AbM_2, aBM_2, Abm_2, aBm_2$	$1-s$	$1-s$
abm_2	$(1-2s-\epsilon)(1-\alpha_1)$	$1+\alpha_1$
abM_2	$(1-2s-\epsilon)(1-\alpha_1-\alpha_2)$	$1+\alpha_1+\alpha_2$

Table 3. Fitnesses of individual genotypes modified by a fixed niche choice parameter α_1 and a polymorphic allele M_2, which increases niche preference for the better niche by α_2. The genotype ABM_2 goes to niche I with a probability increased by the factor $\alpha_1+\alpha_2$ and to niche II decreased by $\alpha_1+\alpha_2$. Similarly, genotype ABM_2 preferentially goes to niche II instead of niche I. Preferences of other genotypes are identical as in Table 2.

Secondary sympatry

It should be noted that all conditions presented above are independent of the initial linkage disequilibrium between the ecological loci. Therefore, for maintenance of genetic polymorphism in the ecological loci it is irrelevant whether the population starts in a Hardy-Weinberg linkage equilibrium or if it consists of only two specialist genotypes as it would during a secondary contact after divergence in allopatry.

Cost of niche choice

It has been suggested by de Meeûs et al. [37] that in a model with soft selection and habitat preference, the conditions under which polymorphism is maintained appear to be slightly broadened but only if there is no cost to habitat selection. We implemented a cost of niche choice, γ, into our model, which is defined as a fraction of the selection coefficient (fitnesses are described in Table 4).

As analytical solutions are rather difficult to obtain, we numerically assessed how various values of γ affect maintenance of polymorphism in the ecological loci. In Fig. S2 we show that for low recombination (around $r = 0.01$), certain cost is robustly tolerated in a large region of the parameter space. E.g., in an example with $s = 0.1, c = 0.46$ and $\epsilon = -0.5s$, cost $\gamma = 0.05s$ is tolerated for $\alpha_2 \geq 0.1$, cost $\gamma = 0.1s$ for $\alpha_2 \geq 0.2$, and cost $\gamma = 0.2s$ for $\alpha_2 \geq 0.3$.

genotype	niche I	niche II
ABM_2	$1+\alpha_1+\alpha_2-\gamma$	$(1-2s-\epsilon)(1-\alpha_1-\alpha_2-\gamma)$
ABm_2	$1+\alpha_1$	$(1-2s-\epsilon)(1-\alpha_1)$
AbM_2, aBM_2	$1-s-\gamma$	$1-s-\gamma$
Abm_2, aBm_2	$1-s$	$1-s$
abm_2	$(1-2s-\epsilon)(1-\alpha_1)$	$1+\alpha_1$
abM_2	$(1-2s-\epsilon)(1-\alpha_1-\alpha_2-\gamma)$	$1+\alpha_1+\alpha_2-\gamma$

Table 4. Fitnesses of individual genotypes modified by a fixed niche choice parameter α_1 a polymorphic allele M_2, which increases niche preference for the better niche by α_2, and a cost of this allele, γ.

Individual-based simulations: polygenic traits with more than 3 loci

We use individual-based simulations to illustrate that initial degree of assortment is essential to allow for the evolution of coupling between ecological loci. When assortment has to evolve from low frequencies, even with strong recombination (r = 0.01), selection has to be strong for the coupling to evolve (c.f. Fig. 3 for 2 ecological loci loci, approx. s > 0.1). Week pre-existing niche preference (one locus out of ten with α = 0.1 is near fixation) allows for evolution of two coexisting specialist if linkage is strong and the number of loci is low (Fig 4 A,B – top, n_{eco} = 3 and 6). The number of loci that would evolve coupling increases readily with increasing initial niche-recognition (Fig. 4 vs 5). For example, with two loci out of ten with α = 0.1 are near fixation, variation is already maintained for a moderate number of loci (Fig 5), though coupling between the ecological loci (in colour) stays weak when recombination is high (r=0.1) (Fig 5B, bottom rows). As initial assortment increases further, conditions for coexistence broaden considerably (see Fig S4) – two nascent specialist can then diverge and coexist even for a polygenic trait under a strong trade-off.

For more ecological loci - assuming variation is not lost faster - we see a slow build-up of coupling between ecological loci (Fig 5C, second row). This can be seen as an abrupt phase change (tipping point, c.f. **[38]**). For our scenario, however, the crucial question is whether enough assortment can build up before variation is lost. Simply because the increase of allele frequencies grows with its heterozygosity, we see a slow increase in niche recognition

initially, when its allele frequencies are low (black). It is to be noted that the niche-preference alleles are growing from low variation, not from zero – this would need different time scales (even with high mutation of $\mu = 10^{-3}$), and the waiting time does further narrow the parameter range for speciation towards coexistence. Here we focus on the later stages, assuming some variation is present.

To help the curious reader to understand the patterns in relations to other changes in the parameters, in the supplement we also include results in the absence of epistasis (when variation is maintained more easily), stronger selection (*ditto*), and a lower population size (converse). Curiously, even with total population size of N=10 000 and selection per locus s_i > 0.01 (16 loci), genetic drift is still important. Unfortunately, detailed exploration of these patterns is beyond the scope of this paper. We provide the source code (for Mathematica, Wolfram) and are happy to deliver further specific reference-simulations upon request.

Figure 4 and 5: As initial niche recognition increases (Figure 4 *vs.* 5), variation in the ecological loci is maintained more easily – for increasing number of loci (columns) and increasing recombination (rows). First row shows the change in allele frequencies for ecological loci (in colour) and niche preference loci (black) under strong linkage (r = 0.01), second row shows the dynamics of the pairwise linkage disequilibria LD between ecological loci (colour), niche preference loci (black) and between ecological and niche preference loci (grey). Third and fourth depicts allele frequencies and LDs when loci are only weakly linked, r = 0.1. Positive linkage disequilibrium evolves between the ecological loci establishes as niche preference evolves. Note that both selection and epistasis are normalized with respect to the baseline two-locus case. This leads to a slight flattening of the trade-off curve as the number of loci increases (see Model and Figure S1). Parameters: s = 0.1, ε = - 0.05, slightly asymmetric niche sizes with c = 0.46, population size across niches N = 10 000. 10 niche preference loci with alpha = 0.1 (average initial niche preference for the specialist is $\mathcal{A}$ = 0.135 for 1 niche-preference locus near fixation (Fig 4) and 0.22 for 2 (Fig. 5). Selection per locus is 0.1, 0.067, 0.033, 0.017, 0.0125 and pairwise epistasis is 0.05, 0.017, 0.0033, 0.00075, 0.00042 for 2,3,6 and 12 loci, respectively (2 loci are the reference value).

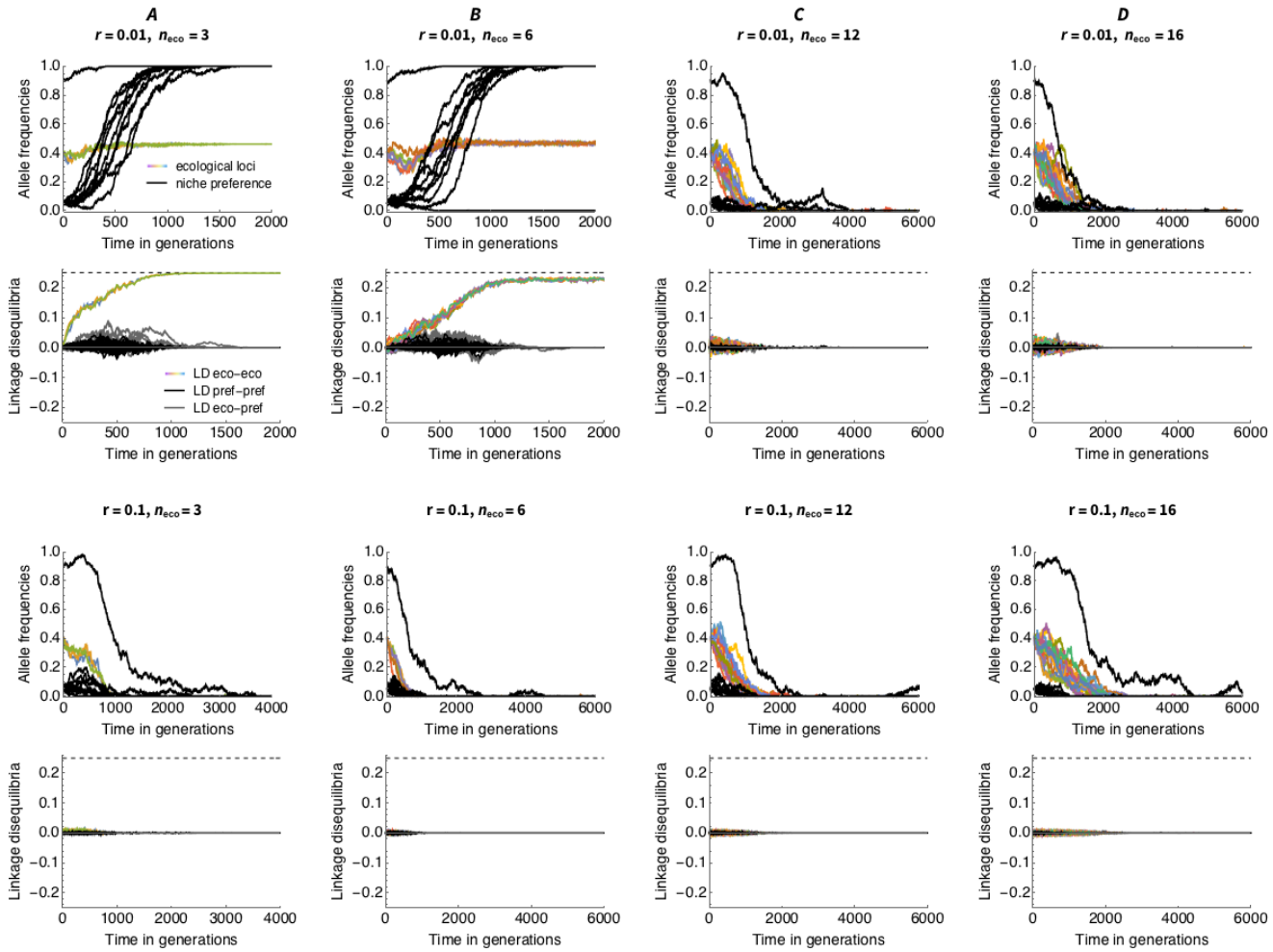
A
r = 0.01, n_eco = 3
B
r = 0.01, n_eco = 6
C
r = 0.01, n_eco = 12
D
r = 0.01, n_eco = 16
r = 0.1, n_eco = 3
r = 0.1, n_eco = 6
r = 0.1, n_eco = 12
r = 0.1, n_eco = 16
Allele frequencies
Linkage disequilibria
Time in generations
ecological loci
niche preference
LD eco–eco
LD pref–pref
LD eco–pref

Figure 4.

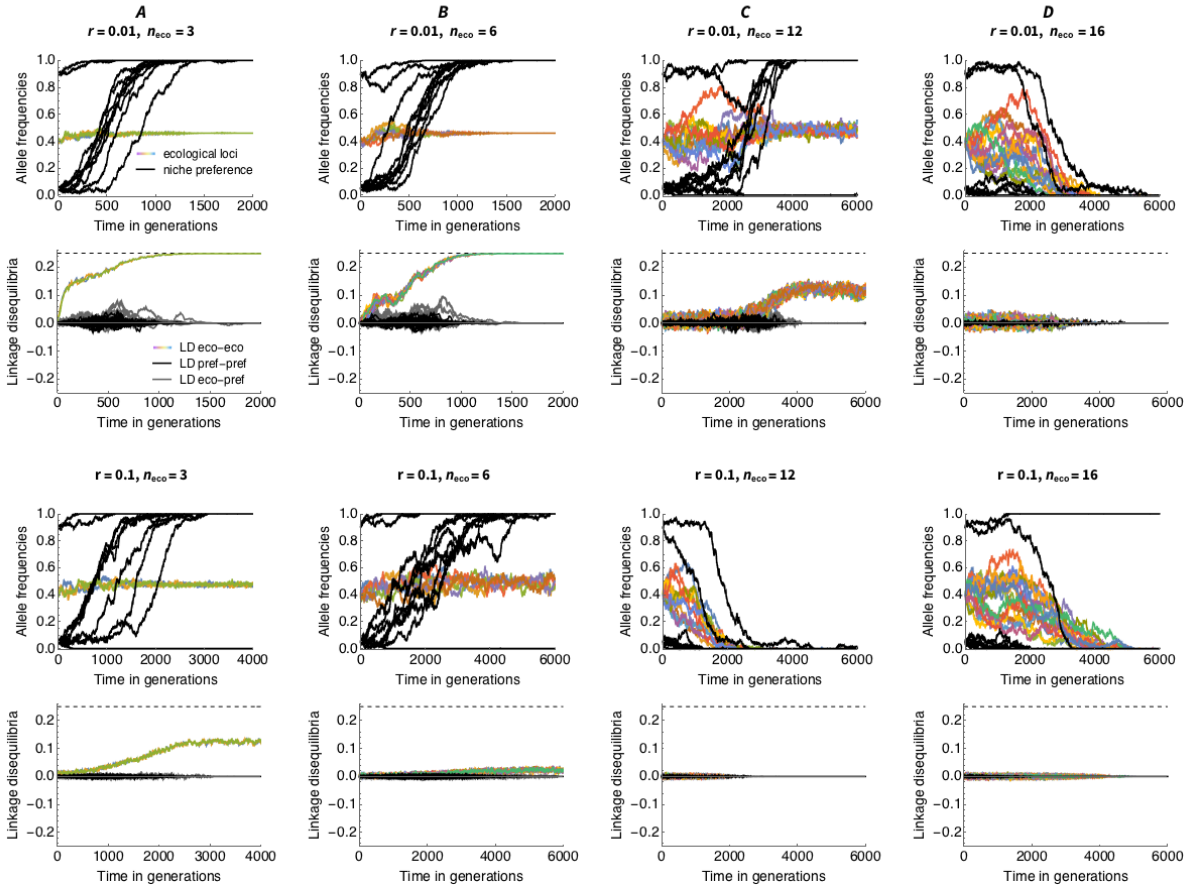
A
r = 0.01, n_eco = 3
B
r = 0.01, n_eco = 6
C
r = 0.01, n_eco = 12
D
r = 0.01, n_eco = 16
r = 0.1, n_eco = 3
r = 0.1, n_eco = 6
r = 0.1, n_eco = 12
r = 0.1, n_eco = 16
Allele frequencies
Linkage disequilibria
Time in generations
ecological loci
niche preference
LD eco–eco
LD pref–pref
LD eco–pref

Figure 5.

Discussion

Maintenance of polymorphism and the plausibility of sympatric speciation are some of the most persistent questions in evolutionary biology. An arguable assumption frequently made in many sympatric speciation models is their restriction to one or rarely several loci, whereas ecological adaptation of a population often involves a gradual change in a polygenic trait [39]. An obvious reason for this common restriction is the analytical difficulty of such models, especially if additional evolutionary forces such as epistasis and nonlinear trade-off between viabilities are included. Notably, the multilocus version of Levene's model – with a trade-off in adaptation to two different niches – has been analysed [40,41] but no epistasis was allowed (linear trade-off). An exception is Barton's [42] analyses of adaptation in a quantitative trait under a strong (convex) trade-off. He assumed the symmetries in allele frequencies were maintained under disruptive and negative frequency dependent selection [43]. Yet, when niches are not perfectly symmetric, independent regulation within niches does not always generate sufficient negative frequency-dependence to stabilise ecological polymorphism. Similarly, negative frequency-dependence is required for evolution of coupling of incompatibilities with underdominance [11].

In order for specialists to be favoured over generalists, trade-off in fitness between niches must be strong, such that a small change of the fitness in one habitat leads to a large change of fitness on the other (the fitness curve is convex). For a polygenic trait, this implies less-than-additive epistasis and/or heterozygote disadvantage. Yet, most early models of sympatric speciation focused on maintenance of polymorphism at a single diploid locus under heterozygote advantage, and on coupling of such polymorphism with a locus for assortment [8,28,35]. Felsenstein [7] analysed coupling of two loci under an ecological trade-off with an independent self-recognising locus for assortment concluded that assortment can only increase for linear or concave (weak) trade-offs. With negative epistasis (strong, convex trade-off), polymorphism was not maintained and assortment could not evolve ([7] , p. 130). It has been recognised that asymmetry in niche proportions significantly influences the ability of these models to maintain genetic polymorphism – and that symmetric selection and concave trade-offs in fitness make such maintenance easier (citations above). Curiously, Barton [42] showed that assortment can readily increase in a population under a

464 convex trade-off, provided the starting genetic variance is large enough. However, that
465 model assumed symmetric niches, and that symmetry in allelic frequencies is maintained by
466 frequency-dependent selection (hypergeometric model, [43]).
467
468 We show that specialists can readily evolve even when the trade-off in fitness between
469 niches is strong, provided that polymorphism is stabilised by assortment arising from a
470 preference for the ''correct'' niche (to which the individual is better adapted), and there is
471 some initial niche-recognition. In other words, some of the speciation and coexistence would
472 not have been attainable by gradual evolution from low frequencies (Fig. 2C,D; Fig 3; c.f. [44]
473 for coexistence under r/K trade-off, and [45,46] for evolution for a diploid locus under strong
474 trade-off and fixed matching habitat preference).
475
476 Pre-existing assortment both speeds up further evolution of niche preference and allows
477 ecological polymorphism to increase. As it stabilises polymorphism in divergent populations
478 under secondary contact, it is replacing the previous assortment due to geographic isolation.
479 Similar reduction of gene flow can arise via philopatry (learned habitat preference) – yet, as
480 the association with fitness is indirect, the effect is likely to be weaker [47,48]. Conversely, in
481 the absence of any form of assortment reflecting adaptation, the conditions for divergence
482 under disruptive selection are very restrictive, as they require specific and perfect
483 symmetries [4], and the parameter range even tightens when ecological traits are
484 determined by many loci.
485
486 We argue that niche recognition and preference is widespread across all life forms. Even
487 plants' roots have growth oriented towards nutrient-rich parts of the rhizosphere [49],
488 motile unicellular organisms, including even bacteria, can move towards their preferred food
489 source by chemotaxis [50,51] and among (in)vertebrates it is generally known that they can
490 recognise and choose their preferred food or habitat . Therefore, this pre-existing ability to
491 recognise the preferred niche may in fact "hijack" arising ecological divergence, protect it,
492 and even further reinforce the newly arising reproductive isolation by strengthening the
493 assortment. Furthermore, the degree of niche preference does not necessarily correlate
494 with the fitness advantage, which the niche provides to the organism. For example, bacteria
495 and yeast tend to prefer one food source over all others (a phenomenon known as a diauxic

shift), which in fact means a 100% preference even if the fitness advantage in growth rate is only minor [52,53].

Acknowledgement

We would like to thank Nick Barton, Joachim Hermisson, Stuart Baird and Taigo Paixo for their insightful comments on the earlier stages of this study. PP was supported by a bilateral Czech-Austrian program Aktion, and both JP and PP were supported by Austrian National Agency (FWF P-32896B).

References

1. Darwin C. 1859 *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* London: John Murray.

2. Mayr E. 1963 *Animal species and evolution*. Cambridge, MA: The Belknap Press of Harvard University Press.

3. Jiggins CD. 2006 Sympatric speciation: why the controversy? *Curr. Biol.* **16**, R333-334. (doi:10.1016/j.cub.2006.03.077)

4. Bolnick DI, Fitzpatrick BM. 2007 Sympatric Speciation: Models and Empirical Evidence. *Annu. Rev. Ecol. Evol. Syst.* **38**, 459–487. (doi:10.1146/annurev.ecolsys.38.091206.095804)

5. Via S. 2001 Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* **16**, 381–390.

6. Dobzhansky T. 1940 Speciation as a Stage in Evolutionary Divergence. *Am. Nat.* **74**, 312–321.

7. Felsenstein J. 1981 Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution* **35**, 124–138. (doi:10.2307/2407946)

8. Maynard-Smith JM. 1966 Sympatric Speciation. *The American Naturalist* **100**, 637–650.

9. Udovic D. 1980 Frequency-dependent selection, disruptive selection, and the evolution of reproductive isolation. *The American Naturalist* **116**, 621–641.

10. Gavrilets S. 2003 Perspective: models of speciation: what have we learned in 40 years? *Evolution* **57**, 2197–2215.

11. Barton NH, de Cara MAR. 2009 The evolution of strong reproductive isolation. *Evolution* **63**, 1171–1190. (doi:10.1111/j.1558-5646.2009.00622.x)

12. Berlocher SH, Feder JL. 2002 Sympatric Speciation in Phytophagous Insects: Moving Beyond Controversy? *Annu. Rev. Entomol.* **47**, 773–815. (doi:10.1146/annurev.ento.47.091201.145312)

13. Greenwood PJ. 1980 Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* **28**, 1140–1162. (doi:10.1016/S0003-3472(80)80103-5)

14. Lavagnino N, Serra F, Arbiza L, Dopazo H, Hasson E. 2012 Evolutionary Genomics of Genes Involved in Olfactory Behavior in the Drosophila melanogaster Species Group: *Evolutionary Bioinformatics* (doi:10.4137/EBO.S8484)

15. Rice WR, Salt GW. 1990 The evolution of reproductive isolation as a correlated character under sympatric conditions: experimental evidence. *Evolution* **44**, 1140–1152. (doi:10.1111/j.1558-5646.1990.tb05221.x)

16. Ravigné V, Dieckmann U, Olivieri I. 2009 Live Where You Thrive: Joint Evolution of Habitat Choice and Local Adaptation Facilitates Specialization and Promotes Diversity. *The American Naturalist* **174**, E141–E169. (doi:10.1086/605369)

17. Butlin R. 1987 A new approach to sympatric speciation. *Trends in Ecology & Evolution* **2**, 310–311. (doi:10.1016/0169-5347(87)90085-1)

18. Jaenike J, Holt RD. 1991 Genetic Variation for Habitat Preference: Evidence and Explanations. *The American Naturalist* **137**, S67–S90. (doi:10.1086/285140)

19. Via S. 1991 The Genetic Structure of Host Plant Adaptation in a Spatial Patchwork: Demographic Variability Among Reciprocally Transplanted Pea Aphid Clones. *Evolution* **45**, 827–852. (doi:10.1111/j.1558-5646.1991.tb04353.x)

20. Mackenzie A. 1996 A trade-off for host plant utilization in the black bean aphid, Aphis fabae. *Evolution* **50**, 155–162. (doi:10.1111/j.1558-5646.1996.tb04482.x)

21. Sezer M, Butlin RK. 1998 The genetic basis of host plant adaptation in the brown planthopper (Nilaparvata lugens). *Heredity* **80**, 499–508. (doi:10.1046/j.1365-2540.1998.00316.x)

22. Rossi AM, Stiling P, Cattell MV, Bowdish TI. 1999 Evidence for host-associated races in a gall-forming midge: trade-offs in potential fecundity. *Ecological Entomology* **24**, 95–102. (doi:10.1046/j.1365-2311.1999.00164.x)

23. Craig TP, Horner JD, Itami JK. 2007 Genetics, experience, and host-plant preference in Eurosta solidaginis: implications for host shifts and speciation. *Evolution* **55**, 773–782. (doi:10.1111/j.0014-3820.2001.tb00813.x)

24. Forister ML, Dyer LA, Singer MS, Stireman JO, Lill JT. 2012 Revisiting the evolution of ecological specialization, with emphasis on insect–plant interactions. *Ecology* **93**, 981–991. (doi:10.1890/11-0650.1)

25. Gripenberg S, Mayhew PJ, Parnell M, Roslin T. 2010 A meta-analysis of preference–performance relationships in phytophagous insects. *Ecology Letters* **13**, 383–393. (doi:10.1111/j.1461-0248.2009.01433.x)

26. Christiansen FB. 1975 Hard and Soft Selection in a Subdivided Population. *The American Naturalist* **109**, 11–16.

28. Smith JM, Hoekstra R. 1980 Polymorphism in a varied environment: how robust are the models? *Genet. Res.* **35**, 45–57. (doi:10.1017/S0016672300013926)

29. Ravigné V, Olivieri I, Dieckmann U. 2004 Implications of habitat choice for protected polymorphisms. *Evolutionary Ecology Research* , 125–145.

30. Levene H. 1953 Genetic Equilibrium When More Than One Ecological Niche is Available. *The American Naturalist* **87**, 331.

31. Dempster ER. 1955 Maintenance of genetic heterogeneity. *Cold Spring Harb. Symp. Quant. Biol.* **20**, 25–31; discussion, 31–32. (doi:10.1101/sqb.1955.020.01.005)

32. Rieseberg LH. 2001 Chromosomal rearrangements and speciation. *Trends Ecol. Evol. (Amst.)* **16**, 351–358. (doi:10.1016/s0169-5347(01)02187-5)

33. Deakin MAB. 1966 Sufficent Conditions for Genetic Polymorphism. *The American Naturalist* **100**, 690–692. (doi:10.1086/282462)

34. Prout T. 1968 Sufficient conditions for multiple niche polymorphism. *The American Naturalist* **102**, 493–496.

35. Hoekstra RF, Bijlsma R, Dolman AJ. 1985 Polymorphism from environmental heterogeneity: models are only robust if the heterozygote is close in fitness to the favoured homozygote in each environment. *Genetics Research* **45**, 299–314. (doi:10.1017/S001667230002228X)

36. Gliddon C, Strobeck C. 1975 Necessary and Sufficient Conditions for Multiple-Niche Polymorphism in Haploids. *The American Naturalist* **109**, 233–235.

37. de Meeûs T, Michalakis Y, Renaud F, Olivieri I. 1993 Polymorphism in heterogeneous environments, evolution of habitat selection and sympatric speciation: Soft and hard selection models. *Evol Ecol* **7**, 175–198. (doi:10.1007/BF01239387)

38. Nosil P, Feder JL, Flaxman SM, Gompert Z. 2017 Tipping points in the dynamics of speciation. *Nat Ecol Evol* **1**, 0001. (doi:10.1038/s41559-016-0001)

39. Wilson DS, Turelli M. 1986 Stable Underdominance and the Evolutionary Invasion of Empty Niches. *The American Naturalist* **127**, 835–850. (doi:10.1086/284528)

40. Nagylaki T. 2009 Evolution under the multilocus Levene model without epistasis. *Theoretical Population Biology* **76**, 197–213. (doi:10.1016/j.tpb.2009.07.003)

41. Bürger R. 2010 Evolution and polymorphism in the multilocus Levene model with no or weak epistasis. *Theor Popul Biol* **78**, 123–138. (doi:10.1016/j.tpb.2010.06.002)

42. Barton NH. 2010 What role does natural selection play in speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 1825–1840. (doi:10.1098/rstb.2010.0001)

601 43. Shpak M, Kondrashov AS. 1999 Applicability of the Hypergeometric Phenotypic Model
602 to Haploid and Diploid Populations. *Evolution* **53**, 600–604. (doi:10.1111/j.1558-
603 5646.1999.tb03794.x)

604 44. Egas M, Dieckmann U, Sabelis MW. 2004 Evolution Restricts the Coexistence of
605 Specialists and Generalists: The Role of Trade-off Structure. *The American Naturalist*
606 **163**, 518–531. (doi:10.1086/382599)

607 45. Priklopil T, Kisdi E, Gyllenberg M. 2015 Evolutionarily stable mating decisions for
608 sequentially searching females and the stability of reproductive isolation by assortative
609 mating: EVOLUTIONARILY STABLE MATING DECISIONS. *Evolution* **69**, 1015–
610 1026. (doi:10.1111/evo.12618)

611 46. Priklopil T. 2012 On invasion boundaries and the unprotected coexistence of two
612 strategies. *J. Math. Biol.* **64**, 1137–1156. (doi:10.1007/s00285-011-0448-y)

613 47. Beltman JB, Haccou P, Cate C ten. 2004 LEARNING AND COLONIZATION OF NEW
614 NICHES: A FIRST STEP TOWARD SPECIATION. *Evolution* **58**, 35–46.
615 (doi:10.1111/j.0014-3820.2004.tb01571.x)

616 48. Beltman JB, Metz JAJ. 2005 Speciation: more likely through a genetic or through a
617 learned habitat preference? *Proc. R. Soc. B* **272**, 1455–1463.
618 (doi:10.1098/rspb.2005.3104)

619 49. Wells DM, French AP, Naeem A, Ishaq O, Traini R, Hijazi H, Bennett MJ, Pridmore TP.
620 2012 Recovering the dynamics of root growth and development using novel image
621 acquisition and analysis methods. *Philos Trans R Soc Lond B Biol Sci* **367**, 1517–1524.
622 (doi:10.1098/rstb.2011.0291)

623 50. Arkowitz RA. 1999 Responding to attraction: chemotaxis and chemotropism in
624 Dictyostelium and yeast. *Trends Cell Biol.* **9**, 20–27. (doi:10.1016/s0962-8924(98)01412-
625 3)

626 51. Xie L, Altindal T, Chattopadhyay S, Wu X-L. 2011 Bacterial flagellum as a propeller and
627 as a rudder for efficient chemotaxis. *Proceedings of the National Academy of Sciences*
628 **108**, 2246–2251. (doi:10.1073/pnas.1011953108)

629 52. Berthels NJ, Cordero Otero RR, Bauer FF, Thevelein JM, Pretorius IS. 2004 Discrepancy
630 in glucose and fructose utilisation during fermentation by Saccharomyces cerevisiae wine
631 yeast strains. *FEMS Yeast Res* **4**, 683–689. (doi:10.1016/j.femsyr.2004.02.005)

632 53. Bren A, Park JO, Towbin BD, Dekel E, Rabinowitz JD, Alon U. 2016 Glucose becomes
633 one of the worst carbon sources for E.coli on poor nitrogen sources due to suboptimal
634 levels of cAMP. *Sci Rep* **6**, 1–10. (doi:10.1038/srep24834)

635