

The evolutionary forces maintaining a wild polymorphism of *Littorina saxatilis*: model selection by computer simulations

A. PÉREZ-FIGUEROA, F. CRUZ, A. CARVAJAL-RODRÍGUEZ, E. ROLÁN-ALVAREZ
& A. CABALLERO

Departamento de Bioquímica, Genética e Inmunología, Facultad de Ciencias, Universidad de Vigo, Vigo, Spain

Keywords:

assortative mating;
genetic drift;
migration;
selection;
structured populations;
sympatric speciation.

Abstract

Two rocky shore ecotypes of *Littorina saxatilis* from north-west Spain live at different shore levels and habitats and have developed an incomplete reproductive isolation through size assortative mating. The system is regarded as an example of sympatric ecological speciation. Several experiments have indicated that different evolutionary forces (migration, assortative mating and habitat-dependent selection) play a role in maintaining the polymorphism. However, an assessment of the combined contributions of these forces supporting the observed pattern in the wild is absent. A model selection procedure using computer simulations was used to investigate the contribution of the different evolutionary forces towards the maintenance of the polymorphism. The agreement between alternative models and experimental estimates for a number of parameters was quantified by a least square method. The results of the analysis show that the fittest evolutionary model for the observed polymorphism is characterized by a high gene flow, intermediate-high reproductive isolation between ecotypes, and a moderate to strong selection against the nonresident ecotypes on each shore level. In addition, a substantial number of additive loci contributing to the selected trait and a narrow hybrid definition with respect to the phenotype are scenarios that better explain the polymorphism, whereas the ecotype fitnesses at the mid-shore, the level of phenotypic plasticity, and environmental effects are not key parameters.

Introduction

Speciation has been extensively studied from a theoretical point of view (reviewed by Turelli *et al.*, 2001). A key issue in the theoretical models concerns the presence or absence of gene flow in the evolution of reproductive isolation, but there is no theoretical problem with speciation occurring in the presence of gene flow (Turelli *et al.*, 2001). For example, theory suggests that speciation can occur as either a direct or a by-product effect of natural selection on particular traits, independently of the gene flow existing between the alternative genotypes (Lande, 1981; Dieckmann & Doebeli, 1999; Johannesson, 2001; Schluter, 2001). Recently, examples of by-product

ecological speciation in sympatry have been reported (Rundle *et al.*, 2000; Schluter, 2001; Nossil *et al.*, 2002), providing excellent case studies to investigate the relative contributions of different evolutionary factors involved in speciation processes. In sympatric scenarios, interaction between several complex factors is possible, and an approach that can be used to assess their relative contribution is model selection (Johnson & Omland, 2004), where competing models are compared with one another by evaluating how well they are supported by the observed data. This approach, applied to well-known case studies, can reveal the combination of evolutionary forces that quantitatively explain empirical observations, and is useful for investigating the possible contributions of new factors that have not been studied experimentally.

There have been several theoretical studies using computer simulation to describe the genetic background

Correspondence: Armando Caballero, Departamento de Bioquímica, Genética e Inmunología, Facultad de Ciencias, Universidad de Vigo, 36200 Vigo, Spain. Tel.: +34 986 812 568; fax: +34 986 812 556; e-mail: armando@uvigo.es

of sympatric speciation processes (Caisse & Antonovics, 1978; Boulding, 1990; Johannesson & Sundberg, 1992; Johnson *et al.*, 1996; Higashi *et al.*, 1999; Takimoto *et al.*, 2000), but few of these have dealt with a detailed quantitative comparison between theoretical simulations and empirical data. This kind of comparison seems particularly useful because many evolutionary factors may contribute to speciation but are experimentally untreatable. Here we focus on the first steps of speciation, investigating the evolutionary factors that fit a real, well-documented case of incomplete ecological speciation in sympatry (Johannesson *et al.*, 1993; Rolán-Alvarez *et al.*, 1997, 1999).

The ovoviviparous marine gastropod *Littorina saxatilis* (Olivi) is a very polymorphic species that inhabits different ecological niches of the North Atlantic shores, from estuaries to rocky shores or even salt marshes (Reid, 1996). Because of its polymorphic nature, this species has been the subject of several evolutionary and population genetic studies (reviewed by Reid, 1996; Johannesson & Tataronov, 1997; Small & Gosling, 2000; Wilding *et al.*, 2001). There is a striking and well-known polymorphism in *L. saxatilis* from the wave-exposed Galician rocky shores (north-west Spain), where two sympatric but still conspecific ecotypes overlap and hybridize (Johannesson *et al.*, 1993, 1995). These ecotypes can be associated with two different habitats. The RB ecotype inhabits the upper shore level, with high presence of barnacles, and it is characterized by a ridged and banded shell, whereas the SU ecotype inhabits the lower shore level, with high presence of mussels, and is characterized by a smooth and unbanded shell. In the mid shore, with a patched mixture of mussels and barnacles, both ecotypes live in true sympatry, although some microhabitat separation persists (Otero-Schmitt *et al.*, 1997; Erlandsson *et al.*, 1999). These two ecotypes differ for many morphological, anatomical and behavioural traits (reviewed in Table 1) but they are able to mate and produce fertile

hybrids (HY) in the wild, maintaining some gene flow between them (Rolán-Alvarez *et al.*, 1996). Hybrids show intermediate values for most characteristics presented in Table 1 (pers. obs.).

Transplant and laboratory experiments indicate that natural selection seems to be responsible for the adaptation of each ecotype to its own habitat and shore level (Rolán-Alvarez *et al.*, 1997). Likewise, the same studies suggest that hybrids of *L. saxatilis* in Galician populations survive as well as the pure forms in the mid shore. The analysis of other fitness components, such as fecundity or sexual selection, has not detected any clear hybrid disadvantage (Johannesson *et al.*, 1995, 2000; Rolán-Alvarez *et al.*, 1999). Even the fitness surfaces for multivariable morphological traits in the hybrids reveal that, whereas some traits are depressed, others are favoured by natural selection (Cruz *et al.*, 2001; Cruz & García, 2002, 2003). This implies the existence of extrinsic selection factors favouring every form (and trait) within their respective habitats. However, gene flow still occurs between ecotypes (Rolán-Alvarez *et al.*, 1996), the majority of migrants being adults because of the low survival rate of juveniles in unfamiliar habitats (Rolán-Alvarez *et al.*, 1997). The mean migration distance for adults in Galicia is approximately 1–2 m per month (Erlandsson *et al.*, 1998), which can generate, on a geographical scale, a scarce rate of gene flow. Nevertheless, previous studies (Johannesson *et al.*, 1993; Rolán-Alvarez *et al.*, 1996) have shown a relatively high gene flow at a local scale that was higher between ecotypes in the mid shore than the RB ecotype from upper shore and the SU ecotype from lower shore. Additionally, the ecotypes have evolved an incomplete assortative mating in the mid shore because of the contribution of ecotype microdistribution and mate choice (Rolán-Alvarez *et al.*, 1999). Therefore, this could be considered a case of incomplete sympatric speciation.

Trait	RB	SU	Reference
Shell height (mm)	5.43 (3.95)	3.48 (0.61)	Johannesson <i>et al.</i> , 1995
Growth rate (mm/month)	2.12	0.62	Johannesson <i>et al.</i> , 1997
Adult dispersal rates (mid shore) (m/month)	2.14 (1.734)	1.04 (0.720)	Erlandsson <i>et al.</i> , 1998
Female fecundity (embryos)	119.4 (50.55)	38.1 (16.72)	Cruz <i>et al.</i> , 1998
No. of teeth cusps in radulae	6.60 (3.95)	7.22 (3.95)	Rolán-Alvarez <i>et al.</i> , 1996
Shell diameter in embryos (mm)	0.55 (3.95)	0.65 (3.95)	Rolán-Alvarez <i>et al.</i> , 1996
Presence at upper (RB) and lower (SU) shores (range)	92–100%	66–97%	Johannesson <i>et al.</i> , 1993
Presence at mid shore	27% (10.15)	56.3% (15.22)	Johannesson <i>et al.</i> , 1993
F _{ST} (allozymes) between ecotypes		0.050 (0.024)	Rolán-Alvarez <i>et al.</i> , 1996
F _{ST} (allozymes) within ecotypes		0.011 (0.007)	Rolán-Alvarez <i>et al.</i> , 1996
Isolation (I _{PSI})*		0.77 (0.15)	Rolán-Alvarez <i>et al.</i> , 1999

Table 1 Ecotype characteristics in the hybrid zone.

*See Rolán-Alvarez & Caballero, 2000.

Numbers in parenthesis are standard deviations.

In this paper, we try to test different potential explanations of how various evolutionary forces (adaptive selection, gene flow and reproductive selection) have combined to produce the observed polymorphism of *L. saxatilis* in Galician shores. Our motivation has been to combine the information obtained from separate studies to investigate in quantitative terms the factors responsible for experimentally observed ecotype frequencies, neutral differentiation and assortative mating. A second objective of this study was to investigate the possible contribution of other factors to the *L. saxatilis* polymorphism that have not been experimentally investigated yet. We have tried to describe the genetic nature of this polymorphism through the number and effects of genes controlling the trait morph, the contribution of environmental effects, the hybrid ecotype definition and the existence of phenotypic plasticity. We have thus ran computer simulations of the behaviour of a *Littorina* subdivided population under different models and quantified, by a least square method, the agreement between the alternative models and empirical results from previous studies (Johannesson *et al.*, 1993; Rolán-Alvarez *et al.*, 1996, 1999).

Materials and methods

Simulation model

We simulated a two-dimensional metapopulation, with two neighbouring transects (equivalent to two vertical shore transects separated by a few meters) and three shore levels (upper shore, mid shore and lower shore) at each one (see Fig. 1a). The number of transects assumed is arbitrary, as the number of subpopulations along the coast is unknown and, in fact, there may be a continuum of populations. However, because the *L. saxatilis* polymorphism is believed to have evolved many times in parallel at a local scale (Cruz *et al.*, 2004), and because gene flow at a large geographical scale is negligible relative to that at microgeographical scale, we chose the minimum number of two transects as a local description. In each of the levels, there was a subpopulation with a constant size of $N = 1000$ individuals, equal numbers of males and females and discrete generations. The number of individuals of each subpopulation was chosen to be sufficiently large as to avoid important genetic drift effects on the timescale considered. Every generation each subpopulation exchanged Nm individuals (before selection) with the adjacent ones following a stepping-stone migration model (Kimura, 1953).

We tried to reflect a situation in which individuals are allocated to different ecotypes depending on the value of an underlying continuous quantitative trait. This trait was controlled by a variable number of unlinked additive loci. Each locus had two alleles, one of them with effect zero, and the other with a constant effect such that the range of genotypic values was always between 0 and 106

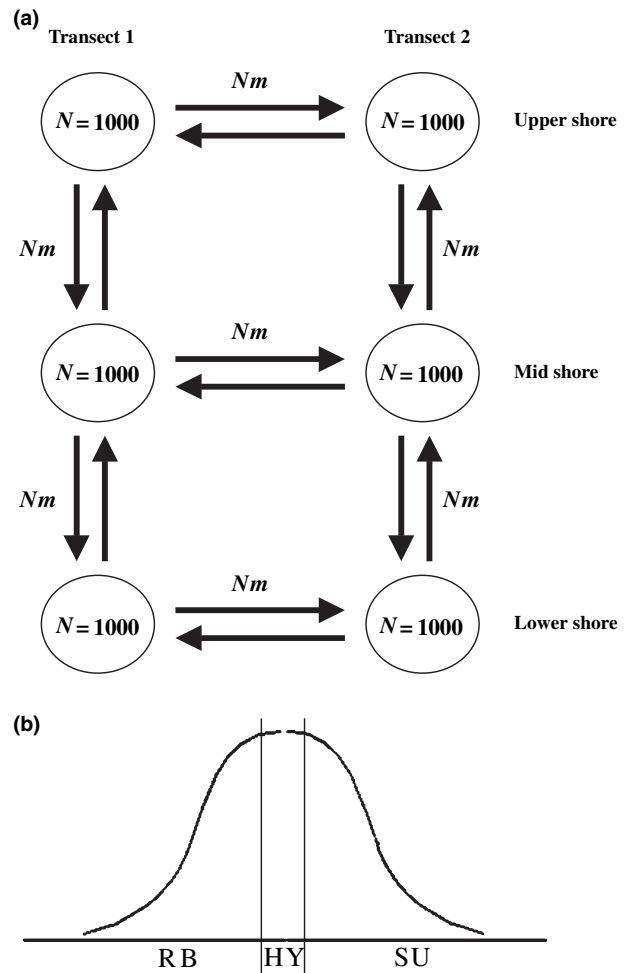


Fig. 1 (a) Structure of the simulated metapopulation. N = subpopulation size; Nm = number of migrants exchanged between subpopulations per generation. (b) Phenotypic distribution of the underlying quantitative trait determining the ecotype definition.

(see Table 2). The basic model considered four genes of large effect and 26 genes of small effect. This model assumes that the ecotype mainly depends on a few traits (such as ridges and bands) controlled by major genes, and a series of traits (such as size, growth rate, etc.) controlled by a number of genes of small effect (see Johannesson *et al.*, 1993; Carballo *et al.*, 2001). Other models assumed just one gene responsible for the ecotype trait, or 30 loci with equal effects. Finally, an infinitesimal model of genes (Fisher, 1918; Bulmer, 1980) was also simulated. In this case, the genotypic value of individuals was randomly taken from a normal distribution with mean the average genotypic value of the parents, and a variance equal to $V_{G_0}(1 - F)/2$, where V_{G_0} is the initial genetic variance and F is the average inbreeding coefficient of the parental generation. Phenotypic values were obtained by adding to the genotypic values an

Table 2 Parameters of simulations. See text for further details.

Simulation	Selected genes (homozygous effect)*	Environmental variance	HY definition†	Phenotypic plasticity variance
sim1	4 (20) + 26 (1)	50	45–55	0
sim2	30 (3.53)	50	45–55	0
sim3	1 (106)	50	45–55	0
sim4	Infinitesimal	50	45–55	0
sim5	1 (106)	0	45–55	0
sim6	4 (20) + 26 (1)	0	45–55	0
sim7	4 (20) + 26 (1)	100	45–55	0
sim8	4 (20) + 26 (1)	50	45–55	50
sim9	4 (20) + 26 (1)	50	5–95	0

*Number of genes affecting the quantitative trait (their homozygous effects in genotypic value units are in parenthesis).

†Percentiles of the phenotypic distribution that define hybrids.

environmental deviation normally distributed with mean zero and a given environmental variance (Table 2).

The range of phenotypic values was divided into three portions (Fig. 1b). The lower one was appointed to the RB ecotype, the higher one to the SU ecotype, and the intermediate one to the HY ecotype. The cutting points between the ecotypes were given by the HY definition parameter, expressed as a proportion of the total phenotypic range. Phenotypic plasticity was included in some simulations, following Boulding (1990), by adding a deviate normally distributed with mean zero and a given phenotypic plasticity variance, to the phenotypic values of individuals. The deviation was negative for the upper shore and positive for the lower shore.

A set of 30 neutral two-allele loci unlinked to the quantitative trait loci were used to estimate the levels of genetic differentiation (F_{ST} calculated from heterozygosities; Wright, 1951; Nei, 1977). This was estimated among subpopulations of the same shore level (basically among individuals of the same ecotype), or among subpopulations of upper and lower levels (basically among individuals of different ecotype, RB and SU). The neutral markers also allowed estimation of the average inbreeding coefficient in each subpopulation and generation, which was necessary to implement the infinitesimal model of gene effects.

In the initial generation, gene frequencies for all loci were assigned using binomial sampling with probability 0.5 for any of the two alleles in each locus. For the quantitative trait, this implied a similar initial frequency of each of the three ecotypes for a hybrid definition between 45 and 55 percentiles of the distribution. Although the phenotypic population distribution in all cases began with a normal distribution (Fig. 1b), depending on the parameters of the simulation, this could end up as a bimodal distribution (with modes for RB and SU individuals). Because of the large subpopulation census sizes and the relatively low timescale considered, little

variation was lost by genetic drift in the course of the simulations. In addition, selection and migration, when present, were expected to be strong enough to neglect mutation in the quantitative trait loci. Thus, we did not generally consider mutation in the simulations. Some runs, however, were carried out including a reversible mutation rate of 10^{-5} per locus and generation both for neutral and quantitative trait loci, in order to confirm the unimportance of this factor.

In every generation, following migration of adults between subpopulations, matings were carried out between randomly chosen males and females of each subpopulation. The mating between the chosen individuals was always accepted, except in the case of a mating between RB and SU individuals, which was accepted with a given probability (see below). Gametes were generated by free recombination among all loci of parental homologous chromosomes. Male and female gametes were combined to produce offspring, whose genotypic and phenotypic values were calculated. The ecotypes' offspring were assigned depending on their phenotypic value (Fig. 1b). They survived with a probability equal to the fitness of their ecotype in the corresponding shore level (see below). This process was repeated until the subpopulation size (1000) was reached. Surviving individuals could then migrate between adjacent subpopulations. This process was repeated for 500 generations to reach a migration-selection-drift equilibrium, and 50 replicates were run for each of the simulation sets.

Evolutionary factors investigated

We ran nine sets of simulations with different combinations of genetic and biological parameters (Table 2). For every set of simulations, we ran the orthogonal combination of the four factors shown in Table 3 (256 combinations). The *gene flow* factor ranged from low gene flow ($Nm = 0.5$) to a very high one ($Nm = 40$). The *reproductive isolation* factor was simulated allowing, or not, mating between individuals depending on their ecotype.

Table 3 Evolutionary factors examined.

Gene flow	Reproductive isolation	Pure zone selection	Hybrid zone selection
0.5	1	1:1:1 (neutral)	0.5:1:0.5 (HY advantage)
8	0.67	1:1:0.5 (moderate)	1:1:1 (neutral)
20	0.33	1:0.67:0.33 (intense)	1:0.5:1 (HY disadvantage)
40	0	1:0:0 (extreme)	1:0:1 (HY inviability)

Gene Flow is expressed as number of migrants, Nm . Reproductive isolation as I , the joint isolation index applied on mating probabilities. Pure zone selection (fitness of RB in upper shore and SU in lower shore: fitness of HY: fitness of RB in lower shore and SU in upper shore). Hybrid zone selection (fitness of RB: fitness of HY: fitness of SU).

Mating was always accepted, except in the case of a mating between RB and SU individuals, which was accepted with a given probability (0, 0.2, 0.5 and 1). Thus, cases ranged from complete assortative mating between ecotypes ($I = 1$) to complete random mating ($I = 0$), where I is the joint isolation index (Merrel, 1950) applied on the mating probabilities (this is equivalent to the I_{PSI} index proposed by Rolán-Alvarez & Caballero, 2000). The *pure zone selection* factor represents the disruptive selection of the ecotypes in the upper and lower shore subpopulations, and it ranged from the neutral case (equal fitness of resident and nonresident ecotypes; 1:1:1, RB:HY:SU, respectively) to the case of zero viability for nonresident ecotypes (e.g. fitness 1:0:0: for RB:HY:SU, respectively in the upper shore). The *hybrid zone selection* factor ranged from hybrid inviability (1:0:1) to hybrid adaptive advantage (0.5:1:0.5) in the mid shore subpopulations.

We finally ran a new simulation (sim10) using the values of the above evolutionary factors (migration, reproductive isolation and selection) that gave better agreement with the empirical observations, in order to investigate the optimal values for the factors presented in Table 2. Thus, we considered four cases of *genetic structure* (1, 4 + 26, 30 genes and the infinitesimal model), three of *environmental variance* (0, 50 and 100), two of *phenotypic plasticity* (variance 0 and 50), and two of *hybrid definition* (hybrids defined within 45–55 and 5–95 percentiles of the phenotypic distribution).

The distance index

We used an index, that we call the *distance index*, to quantify the amount of fit between simulated competing models and empirical observations. Every simulation yielded results for the following six variables at generation 500: proportion of RB individuals in the upper shore level, proportion of SU individuals in the lower shore level, proportion of every ecotype in the mid shore level (averaged across ecotypes), F_{ST} between ecotypes and F_{ST} within ecotypes, and observed reproductive isolation between RB and SU ecotypes (measured as I_{PSI} , see Rolán-Alvarez & Caballero, 2000 for details about this index). These variables have been estimated in natural populations (Johannesson *et al.*, 1993; Rolán-Alvarez *et al.*, 1996, 1999), so we could compare simulated and empirical estimates (these latter are shown in the last five rows of Table 1). The fit of a particular simulation to the empirical means was calculated by the absolute difference between simulated and empirical observations relative to the empirical ones. For the frequency of morphs at upper and lower shores, a range instead of a mean value was available, and a distance zero was applied for simulation values within the range given. This procedure rendered a least squared distance coefficient for each of the variables compared, and the *distance index* was their sum, ranging from zero (maximum fit) to

infinity (maximum unfit). It is important to emphasize that the evolutionary factors studied above (levels of natural selection and migration, genetic structure, hybrid definition, environmental variance and phenotypic plasticity) were not used in the present index. An exception was reproductive isolation, for which we determined the *a priori* mate choices and measured the *a posteriori* estimates of sexual isolation from mating frequencies. Note that this does not pose a circularity problem, as other factors can interact with the mating frequencies so that the *a posteriori* estimates of sexual isolation can be biased (Rolán-Alvarez & Caballero, 2000).

To test the relative importance of the former evolutionary factors, we compared the mean distance values between treatments for every factor, but used the best-fit treatment combination for the rest of factors in each particular simulation (sim1–10). This allowed us to check the differences among treatments for the investigated factors by classical tests (Sokal & Rohlf, 1995). The total number of simulations performed was 15 200. Statistical analyses were carried out using SPSS/PC ver. 11.5.

Deterministic model

To both test the appropriateness of the factor levels chosen and check the Monte Carlo simulations, we also developed a quasi-deterministic simplified model. This was, basically, a one-locus model without environmental variation (equivalent to sim5 in Table 2). The initial frequencies for RB, HY and SU ecotypes (f_{RB} , f_{HY} , f_{SU}) were 0.25, 0.5 and 0.25, respectively, at each of the three shore levels (upper, mid and lower shore). Ecotype frequencies were computed deterministically, but there was also a nondeterministic mating procedure based on a Monte Carlo algorithm using the ecotype frequencies as parameters.

First, the ecotype frequency after selection and migration (f^*) was calculated. For example, the frequency of ecotype RB at the upper shore (f_{RBu}^*) was computed as

$$f_{RBu}^* = (1 - m)f_{RBu}W_{RBu} + mf_{RBm}W_{RBm}, \quad (1)$$

where $1 - m$ is the proportion of RB individuals which do not migrate from upper to mid shore, m is the proportion of RB individuals migrating from mid to upper shore, f_{RBu} and f_{RBm} are the frequencies before migration and selection of RB at upper and mid shore, respectively, and W_{RBu} and W_{RBm} are the fitnesses of RB at upper and mid shore, respectively. The frequency f_{RBu}^* was afterwards normalized with respect to the frequencies of all existing ecotypes at the upper shore. Frequencies of the other ecotypes were calculated in the same way.

Secondly, we resampled from a trinomial distribution (with probabilities f_{RBu}^* , f_{HYu}^* , f_{SUu}^*) the number of male and female individuals used for matings (500 per sex and shore level). Thus, the frequency of RB matings at the upper shore ($RB_uRB'_u$, the prime denotes females,

otherwise males) would be $f_{RBu}^* \times f_{RBu}^*$ on average, but the actual mating number would vary around this average due to sampling.

Finally, we computed the number of newborn individuals in the next generation. Thus, for example, the newborn RB from upper shore (N_{RBu}) were calculated as

$$N_{RBu} = RB_u RB'_u C_{RBRB} + \frac{1}{2} (RB_u HY'_u + HY_u RB'_u) C_{RBHY} + \frac{1}{4} (HY_u HY'_u C_{HYHY}), \quad (2)$$

where $RB_u RB'_u$ refers to the number of RB matings at the upper shore, and analogously for other ecotypes, C_{RBRB} , C_{RBHY} and C_{HYHY} are the mating probabilities for RB × RB, RB × HY and HY × HY pairs, respectively. N_{HYu} and N_{SUu} were similarly calculated. Then we obtained the morph frequency for the next generation (f) by dividing newborn ecotypes by the total number of individuals per shore level,

$$f_{RBu} = \frac{N_{RBu}}{N_{RBu} + N_{HYu} + N_{SUu}}, \quad (3)$$

and so forth for all other ecotype frequencies.

The deterministic model was much more restrictive than the simulated one, as it could not provide estimates of F_{ST} and referred only to a single locus, but it had the advantage of a very fast computing time, so it allowed for a large range of evolutionary parameters to be investigated. The evolutionary factors studied were the same as

before, but we could use 100 different values for each of them, covering their whole range. In the case of pure zone selection, we first maintained the fitness of resident ecotypes and hybrids equal to one, and reduced the fitness of the nonresident ecotypes (RB in the lower shore and SU in the upper shore) from 1 to 0.5 with a decrement of 0.01. Then we continued reducing the fitness of the nonresident ecotypes from 0.5 to 0 (decreasing by 0.01), but maintaining the fitness of the hybrids double the nonresident ones.

Results

Validation of the simulation procedure

Figure 2 shows results of the comparison between the deterministic (lines) and simulation (bars) models. The results are given for each factor assuming the null level of the others (reproductive isolation equal to zero, fitness equal to one for every ecotype in every shore level, and no gene flow between subpopulations). Overall, there was a good agreement and high and significant correlation between deterministic and simulation models for all the common combination of factors studied (averaging across replicates $r = 0.99$, $n = 13$, $P < 0.001$). The factor levels chosen in the simulations were good representatives of their whole range. The only exception was in the *hybrid zone selection* factor, where one of the representative points of the factor distribution (0:1:0, where only

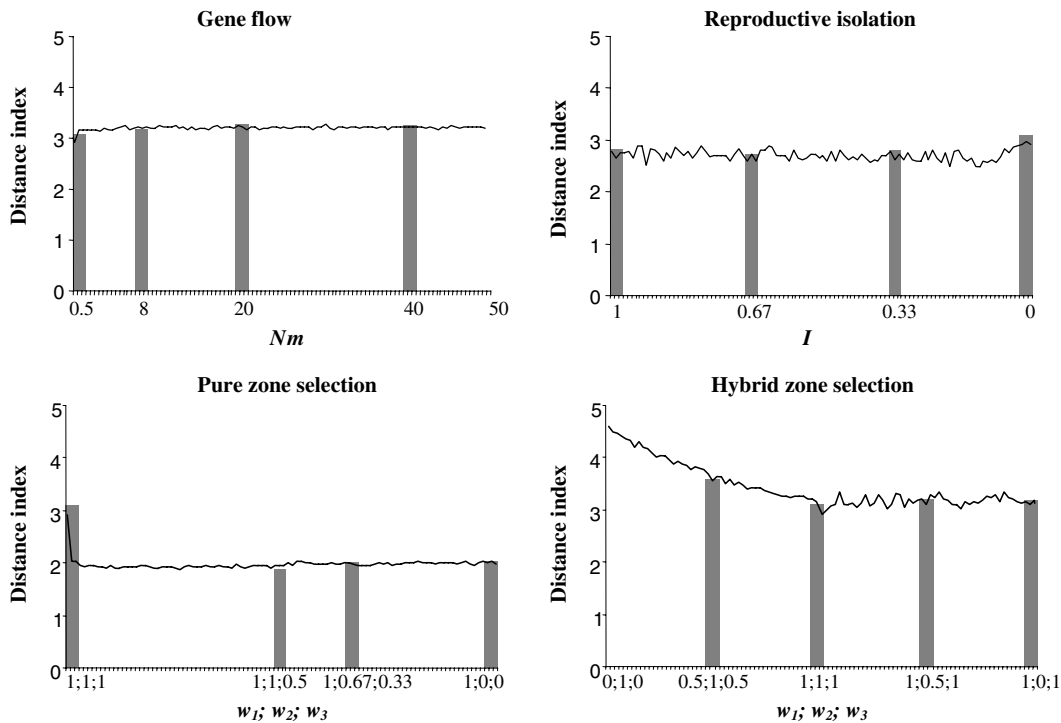


Fig. 2 Distance index for simulations (bars) and the deterministic model (lines) for the parameter set in sim5 (see Table 2).

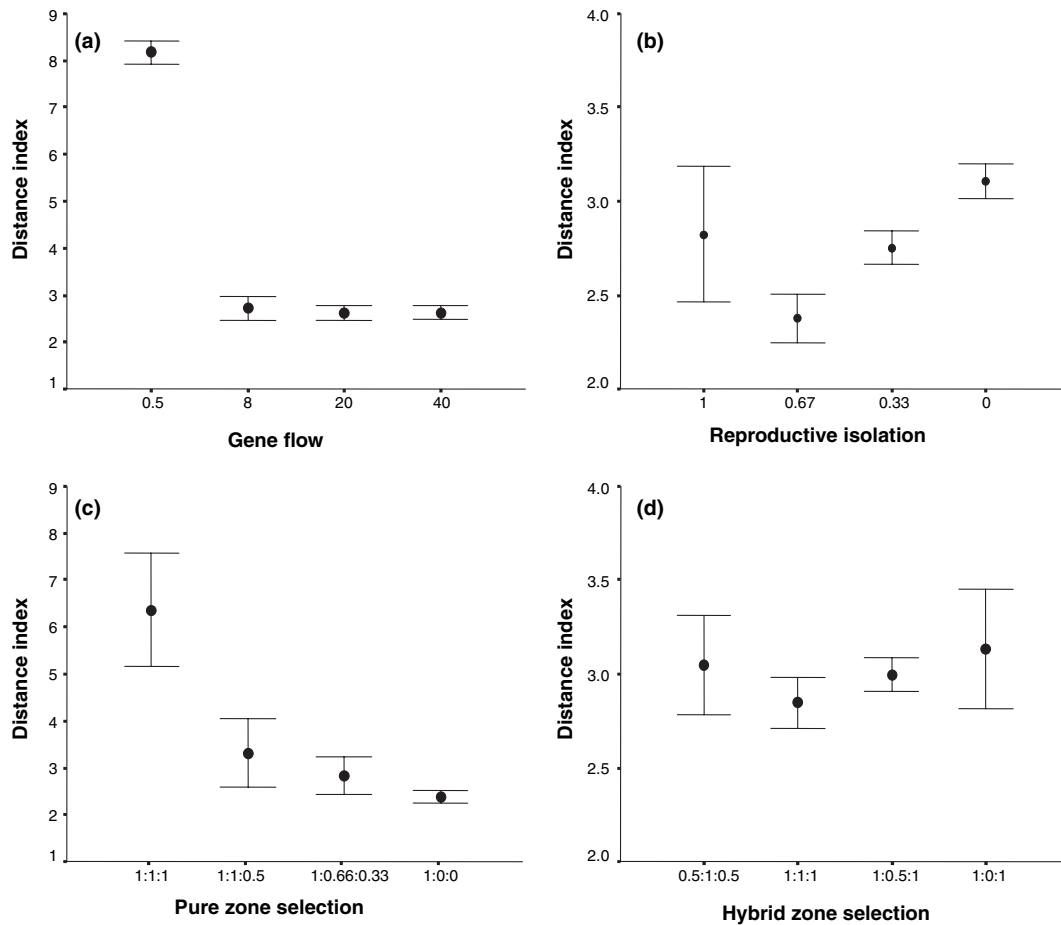


Fig. 3 Averaged distance index for the simulations presented in Table 2. Vertical bars represent the standardized errors for the nine simulations.

the hybrid can survive in the mid shore) was not chosen in the simulations. Nevertheless, this point is very unrealistic at the *L. saxatilis* hybrid zone (see Rolán-Alvarez *et al.*, 1997), as indicated by its high distance index. Below, we focus on the simulation model because the deterministic one was very limited.

Effects of gene flow, reproductive isolation and selection

Four evolutionary factors were separately studied by comparing mean distances across treatments, but using the best-fit treatment combination for the rest of the factors. Interestingly, all simulations rendered similar trends (see averages across sim1–9 in Fig. 3).

Averaged distance values for the four levels of *gene flow* are shown in Fig. 3a. The trend is very clear, the lowest level of gene flow ($Nm = 0.5$) showed the worst fit to the empirical observations across the different simulations, whereas the other levels ($Nm = 8, 20$ or

40) behaved better and rather similarly. In fact, the difference between the lowest Nm level (0.5) and the others could be shown to be significant in each of the nine simulations separately by a *posteriori* Student–Newman–Keuls (SNK) test ($P < 0.05$). The best-fit Nm value, using SNK tests ($P < 0.05$), was not always the same: being eight for sim1, sim3, sim5, sim6 and sim8; 20 for sim2; 20–40 for sim5 and 8–40 for sim7 and sim9. The reason for the low fit between simulations and observations for $Nm = 0.5$ was a high difference between simulated and observed F_{ST} between ecotypes in all simulations.

Averaged distance values for the levels of *reproductive isolation* are presented in Fig. 3b. The overall trend favours a reproductive isolation value of 0.67, although this was not shown in all simulations. Using SNK tests ($P < 0.05$) the best-fit reproductive isolation value was 1 for sim1 and sim6, and 0.67 for all the others. For most simulations the worst reproductive isolation value was 0 or 0.33, because of differences in F_{ST} between and within

ecotypes with respect to their corresponding empirical values.

The results on *pure zone selection* (Fig. 3c) implied disruptive selection, showing the best fit when the resident ecotype (RB on upper shore, SU on lower shore) had a higher fitness than the nonresident ones. All single simulations rendered significant differences by SNK tests between the null selection model (the worst fit model) and at least one of the disruptive selection models. However, different trends were evidenced. Some simulations (sim4, 5, 8 and 9) showed a significant negative relationship between the distance index and the degree of disruptive selection ($P < 0.05$), whereas the remainder (sim1, 2, 3, 6 and 7) showed a plateau for the disruptive selection values. In all simulations, the neutral case showed the worst fit because all simulated parameters (F_{ST} , morph frequency, isolation, etc.) were inconsistent with the empirical ones.

There was no clear trend across simulations for *hybrid zone selection* (Fig. 3d), although single simulations evidenced significant patterns. For sim6, the neutral case (equal fitness of ecotypes at the mid shore) showed the best model fitting by a SNK test ($P < 0.05$). Hybrid

unfitness showed the best model fitting in sim4 and sim9, whereas hybrid heterosis showed the best model fit in sim3 and sim8 ($P < 0.05$). The rest of the simulations (sim1, 2, 5 and 8) did not show any clear pattern. As it is shown in the figure, the mean differences between levels were typically small (even if occasionally significant) and this may therefore perhaps suggest that this factor plays a minor role in the maintenance of this particular polymorphism.

Effects of genetic structure, environmental variance, phenotypic plasticity and hybrid definition

We analysed the orthogonal combination of four additional factors using the best-fit values of the previous ones (migration with $Nm = 20$, reproductive isolation with $I = 0.67$, extreme pure selection and neutral hybrid zone performance). There were significant differences in *genetic structure* by a *t* test ($P < 0.05$; Fig. 4a), with the best-fit model being that with 30 genes of equal additive effects. The *hybrid definition* was also an important factor, showing very significant differences ($P < 0.001$) between the two levels used (Fig. 4b). This suggests that hybrid

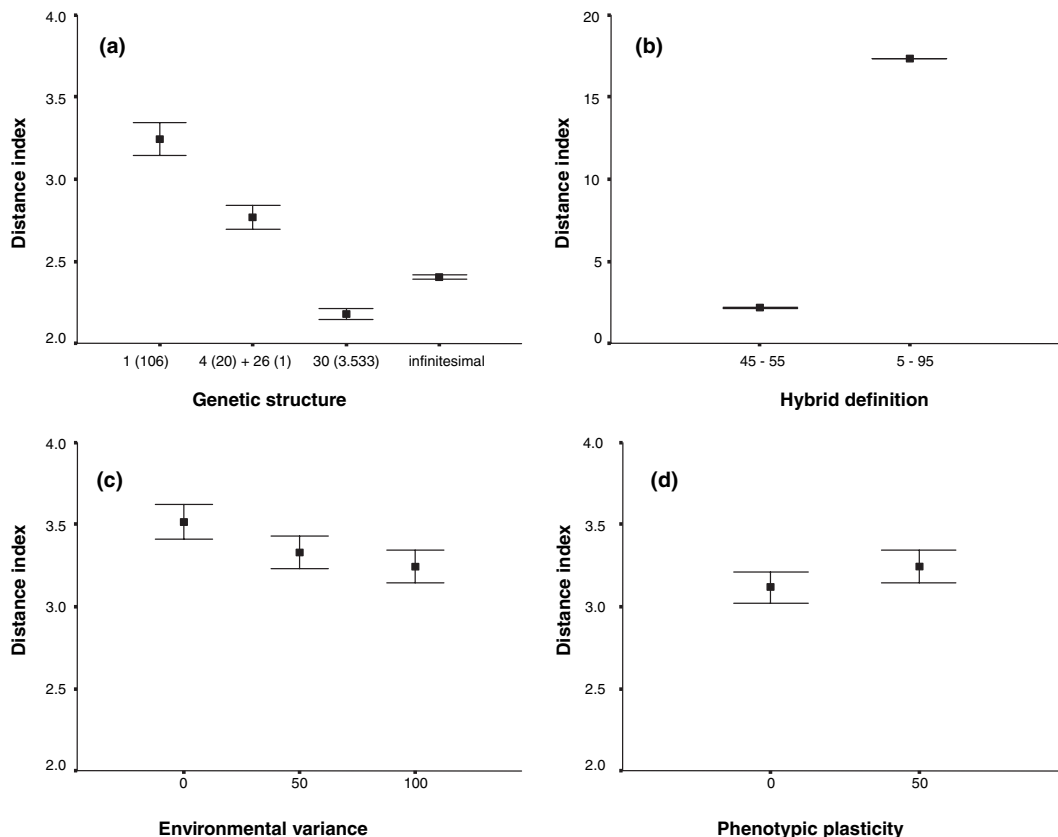


Fig. 4 Averaged distance index for the simulation used to investigate the importance of the genetic and environmental parameters of Table 2. Vertical bars represent the standardized errors for the 50 replicates simulated.

ecotypes should be relatively homogeneous genotypes. *Environmental variance* and *phenotypic plasticity* did not show clear differences (Fig. 4c,d), suggesting that they are not determinant factors for explaining the present polymorphism.

Discussion

Understanding the factors that maintain the Galician polymorphism of *L. saxatilis* is of key importance from an evolutionary point of view, as this polymorphism has been suggested to occur as a result of an incomplete sympatric speciation process (Johannesson *et al.*, 1993, 1995; Rolán-Alvarez *et al.*, 1997, 1999). In this context, we carried out a model selection procedure (Johnson & Omland, 2004) to investigate the relative support of different combinations of evolutionary factors (selection intensities, migration rates and reproductive isolation levels) on the maintenance of the polymorphism.

Under the models investigated, the migration rate must be high (larger than about eight migrants per generation) to explain the polymorphism (Fig. 3), supporting a sympatric scenario. Experimental estimates of Nm showed a wide range (5–75; Rolán-Alvarez *et al.*, 1996), compatible with this result. However, the high gene flow deduced from the present study is somewhat surprising given the low dispersal ability of the species (reviewed by Reid, 1996). Although the distribution of the ecotypes is apparently micro-parapatric, i.e. the two ecotypes live preferentially at distinct shore levels (upper and lower shores; Johannesson *et al.*, 1993), following Futuyma & Mayer (1980), the *Littorina* model must be considered sympatric. This is so because every individual from a particular locality has (at least theoretically) the capability to meet individuals from the other ecotype (adults migrate a few meters per month and the two habitats are separated between one and two dozen meters; Johannesson *et al.*, 1993, 1995). We have assumed two parallel close transects in the simulations because gene flow is only relevant at microgeographical distances. However, to incorporate further transects would mainly affect values of F_{ST} within ecotypes, and it would require larger gene flow than that presented above to explain the observed differentiation. This would not affect the main conclusion that the system needs a high level of gene flow between subpopulations to explain the observed parameters.

The best results of reproductive isolation were reached for $I = 0.67$ (Fig. 3). This is in agreement with the experimental estimates that yielded a reproductive isolation value of approximately 0.7 (Rolán-Alvarez *et al.*, 1999). At present, only prezygotic isolation has been detected experimentally and our simulation results also suggest that hybrid inviability (a type of post-zygotic isolation), if present, is not relevant to the polymorphism. Indeed, the simulation results suggest that the polymorphism is intraspecific (Johannesson *et al.*, 1993;

Rolán-Alvarez *et al.*, 1997) as the lowest distance index values were given by simulations with high gene flow, lack of post-zygotic effects and incomplete reproductive isolation.

The simulations are also in agreement with experimental evidence for the existence of a selective gradient acting over the ecotypes (Rolán-Alvarez *et al.*, 1997), as the results obtained are optimal when opposite selection (viability) gradients of variable intensity are applied to the ecotypes in the upper and lower shores (Fig. 3). However, the analysis suggests that even moderate divergent selection is sufficient to get a good model fitting with respect to the empirical observations. This result indicates that less restrictive conditions are required to maintain a polymorphism that may lead to sympatric speciation. This is true although we have not included habitat choice in the simulations, a mechanism that favours polymorphism maintenance (García-Dorado, 1986).

The role of hybrid fitness in maintaining this polymorphism was much less clear (Fig. 3). The most frequent trend across simulations was a hybrid fitness equal to that of pure ecotypes, as observed for viability in transplant experiments made in the wild (Rolán-Alvarez *et al.*, 1997). However, as has been shown for hybrid fecundity estimates obtained in the wild (Cruz & García, 2003), a few simulations supported both hybrid unfitness and hybrid advantage. In general, the simulations showed rather small differences between the levels of hybrid fitness (Fig. 3), implying that this is not a critical factor for maintaining this particular polymorphism. Nevertheless, the high level of gene flow observed and the absence of post-zygotic reproductive isolation suggest that hybrid fitness must be determined by habitat (exogenous) selection (Arnold & Hodges, 1995) rather than by endogenous selection (Barton & Hewitt, 1985).

Summarizing, we obtained a reasonable, although not perfect, fit between simulations and empirical observations (mean distance values about two for the best combinations, with a possible range between zero and infinity). The results, however, further corroborate previous interpretations of this polymorphism (Johannesson *et al.*, 1993; Rolán-Alvarez *et al.*, 1997, 1999; Cruz *et al.*, 2004): the Galician polymorphism of *L. saxatilis* can be explained by a high gene flow ($Nm \geq 8$), an incomplete reproductive isolation ($I = 0.67$) and a moderate to strong disruptive selection favouring ecotypes in their own habitat (RB in the upper shore and SU in the lower shore). This is an important outcome of our study, as the joint effect of the different evolutionary forces could interact in unexpected ways when studied in combination. In addition, the high general fit between simulations and experimental observations confirms that no other important contributing factors are missing in the model.

This study also casts light on other factors that have not yet been investigated experimentally, such as the genetic

architecture of the trait defining the ecotype morphology, the norm of reaction and environmental and plastic sources of variation. The hybrid definition factor could have relevance in explaining the polymorphism, as the observations seem incompatible with a large proportion (95%) of the phenotypic distribution of the trait defining the hybrids. This can be interpreted as the RB and SU ecotypes from mid shore being more genetically heterogeneous than the same ecotypes from the upper and lower shores. In the simulations used to specifically assess the genetic structure of the ecotype trait (Fig. 4), the results indicate that a moderate number of loci responsible for the ecotype morphology produce the closest agreement with the observations, strengthening the hypothesis of a polygenic quantitative trait contributing to morphology. A model with a substantial number of loci appears to be more likely than models with a few loci or with an infinitesimal model. Phenotypic plasticity and the amount of environmental variance are not key factors in maintenance of the polymorphism (although some environmental variation is needed; see Fig. 4). Nature is certainly far more complex than what we have simulated here, and the inclusion of new factors, such as habitat choice at the mid shore or asymmetrical gene flow within and between ecotypes, would perhaps improve the similarity between simulation outputs and what is observed in the wild. Mutation was not included in the simulations because a low impact of mutation would be expected considering the large subpopulation sizes used, the relative small number of generations simulated, and the negligible effects of mutation relative to migration and selection. Nevertheless, we carried out sim1 (for all 256 model combinations in Table 3) with reversible mutation rates of 10^{-5} per locus and generation, and got identical results to those without mutation, confirming our expectation.

In a previous simulation study, Boulding (1990) developed a model to investigate the causes of genetic differentiation between *Littorina sitkana* and *Littorina sp.* In this model, individuals were defined by a quantitative trait controlled by eight loci of equal effect, and their fitnesses depended on their phenotypes and the class of subpopulations (wave-exposed or wave-protected) in which they lived. Assortative mating was not considered. Her basic conclusion was that genetic differentiation could be generated by the action of selective gradients in spite of the existence of migration. In our work the reduction of genetic differentiation by migration is compensated for by the increase of differentiation caused by assortative mating. Thus, the results of Boulding (1990) could have been similar to ours if assortative mating had been considered.

In another simulation study on *Littorina*, Johannesson & Sundberg (1992) developed a speciation model to study the Swedish populations of *L. saxatilis*. They considered a model of two loci, one defining the ecotype and fitness and the other defining the assortative mating.

In that work, subpopulations were distributed linearly on a one-dimensional stepping-stone model. Their results are in disagreement with ours, as the maintenance of their polymorphism was compatible with a low gene flow and a hybrid advantage. However, the model was one-dimensional, because there is no vertical gradient on Swedish coasts (spring tides can reach about 30 cm), whereas this is not the case on Galician shores (spring tides up to about 4 m). The migration model of Johannesson & Sundberg (1992) implied selection against hybrids (as hybrids always migrated to a habitat where they had very low fitness) and, thus reproductive isolation genes became fixed. In contrast, in our model mid shore hybrids can migrate to three different demes (upper shore, lower shore or mid shore in the other transect), where they may exhibit low to high fitness (depending on the *pure zone selection* factor). Therefore, our model does not systematically select against hybrids. Furthermore, in our simulations with hybrid advantage and high gene flow, the frequency of hybrids in the mid shore is far higher than in empirical observations. Additionally, the high hybrid frequency tends to reduce the frequencies of RB and SU in their respective pure zones (by HY migration), and the distance index values are increased.

Sympatric speciation is a controversial topic in evolutionary biology and there is not conclusive evidence for it, but there is a strong and growing belief that sympatric speciation may be a real evolutionary process (Turelli *et al.*, 2001; Via, 2001). Our results indicate that the Galician polymorphism of *L. saxatilis* is compatible with some of the conditions thought to facilitate sympatric speciation (reviewed in Via, 2001), namely strong disruptive natural selection on habitat use, genotype-environment interaction, positive genetic correlation between divergently selected characters and assortative mating. In fact, the Galician littorinid system can be described as a case of by-product ecological speciation (*sensu* Schluter, 2001; Cruz *et al.*, 2004). The next effort on theoretical grounds should be directed towards modelling how reproductive isolation may evolve under realistic conditions, as it has been claimed in this case that reproductive isolation could emerge as a side-effect of natural selection driving the ecotype size differences along the environmental gradient.

Acknowledgments

We are grateful to Nick Barton and to two anonymous referees for useful comments on the manuscript, and to Darren Martin for English corrections. This work was supported by EUMAR project from European Union (EVK3-CT-2001-00048) and grants from Xunta de Galicia (PGDIT02PXIC30101PM) and Universidade de Vigo (64102C124). A. Pérez-Figueroa was supported by a FPI fellowship (Ministerio de Ciencia y Tecnología).

References

- Arnold, M.L. & Hodges, S.A. 1995. Are the natural hybrids unfit relative to their parents? *TREE* **10**: 67–71.
- Barton, N.H. & Hewitt, G.M. 1985. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **16**: 113–148.
- Boulding, E.G. 1990. Are the opposing pressures on exposed and protected shores sufficient to maintain genetic differentiation between gastropod populations with high intermigration rates? *Hydrobiologia* **193**: 41–52.
- Bulmer, M.G. 1980. *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.
- Caisse, M. & Antonovics, J. 1978. Evolution in closely adjacent plant populations. IX. Evolution of reproductive isolation in clinal populations. *Heredity* **40**: 371–384.
- Carballo, M., García, C. & Rolán-Alvarez, E. 2001. Heritability of shell traits in wild *Littorina saxatilis* populations: results across a hybrid zone. *J. Shell. Res.* **20**: 415–422.
- Cruz, R. & García, C. 2002. Disruptive selection on female reproductive characters in a hybrid zone of *Littorina saxatilis*. *Evol. Ecol.* **15**: 167–182.
- Cruz, R. & García, C. 2003. Using environmental effects on fecundity to compare the adaptive characteristics of the morphs in a hybrid zone of *Littorina saxatilis*. *Evol. Ecol.* **17**: 157–173.
- Cruz, R., Rolán-Alvarez, E. & García, C. 1998. Natural selection on a vertical environmental gradient in *Littorina saxatilis*: analysis of fecundity. *Hydrobiologia* **378**: 89–94.
- Cruz, R., Rolán-Alvarez, E. & García, C. 2001. Sexual selection on phenotypic traits in a hybrid zone of *Littorina saxatilis* (Oliv). *J. Evol. Biol.* **14**: 773–785.
- Cruz, R., Carballo, M., Conde-Padín, P. & Rolán-Alvarez, E. 2004. Testing alternative models for sexual isolation in natural populations of *Littorina saxatilis*: an indirect support for by-product ecological speciation? *J. Evol. Biol.* **17**: 288–293.
- Dieckmann, U. & Doebeli, M. 1999. On the origin of species by sympatric speciation. *Nature* **400**: 354–357.
- Erlandsson, J., Rolán-Alvarez, E. & Johannesson, K. 1998. Migratory differences between ecotypes of the snail *Littorina saxatilis* on Galician rocky shores. *Evol. Ecol.* **12**: 913–924.
- Erlandsson, J., Kostylev, V. & Rolán-Alvarez, E. 1999. Mate search and aggregation behaviour in the Galician hybrid zone of *Littorina saxatilis*. *J. Evol. Biol.* **12**: 891–896.
- Fisher, R.A. 1918. The correlation between relatives under the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* **52**: 399–433.
- Futuyma, D.J. & Mayer, G.C. 1980. Non-allopatric speciation in animals. *Syst. Zool.* **29**: 254–271.
- García-Dorado, A. 1986. The effect of niche preference on polymorphism protection in a heterogeneous environment. *Evolution* **40**: 936–945.
- Higashi, M., Takimoto, G. & Yamamura, N. 1999. Sympatric speciation by sexual selection. *Nature* **402**: 523–526.
- Johannesson, K. 2001. Parallel speciation: a key to sympatric divergence. *TREE* **16**: 148–153.
- Johannesson, K. & Sundberg, P. 1992. Speciation in *Littorina saxatilis* (Oliv)? A one-dimensional selection-migration model. In: *Proceedings of the Third International Symposium on Littorinid Biology*. (J. Grahame, P.J. Miu & D.G. Reid, eds), pp. 1–8. The Malacological Society of London.
- Johannesson, K. & Tatarenkov, A.N. 1997. Allozyme variation in a snail (*Littorina saxatilis*)- deconfounding the effects of microhabitat and gene flow. *Evolution* **51**: 402–409.
- Johannesson, K., Johannesson, B. & Rolán-Alvarez, E. 1993. Morphological differentiation and genetic cohesiveness over a microenvironmental gradient in the marine snail *Littorina saxatilis*. *Evolution* **47**: 1770–1787.
- Johannesson, K., Rolán-Alvarez, E. & Ekendahl, A. 1995. Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution* **49**: 1180–1190.
- Johannesson, K., Rolán-Alvarez, E. & Erlandsson, J. 1997. Growth rate differences between upper and lower shore ecotypes of the marine snail *Littorina saxatilis* (Oliv) (Gastropoda). *Biol. J. Linn. Soc.* **61**: 267–279.
- Johannesson, K., Larsson, A., Cruz, R., García, C. & Rolán-Alvarez, E. (2000). Hybrid fitness seems not to be an explanation for the partial reproductive isolation between ecotypes of Galician *Littorina saxatilis*. *J. Moll. Stud.* **66**: 149–156.
- Johnson, J.B. & Omland, K.S. 2004. Model selection in ecology and evolution. *TREE* **19**: 101–108.
- Johnson, P.A., Hoppensteadt, F.C., Smith, J.J. & Bush, G.L. 1996. Conditions for sympatric speciation: a diploid model incorporating habitat fidelity and non-habitat assortative mating. *Evol. Ecol.* **10**: 187–205.
- Kimura, M. 1953. 'Stepping-stone' model of population. *Ann. Rep. Natl. Inst. Genet. Japan* **3**: 62–63.
- Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. *Proc. Natl. Acad. Sci. USA* **78**: 3721–3725.
- Merrel, D.J. 1950. Measurement of sexual isolation and selective mating. *Evolution* **4**: 326–331.
- Nei, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.* **41**: 225–233.
- Nossil, P., Crespi, B.J. & Sandoval, C.P. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* **417**: 440–443.
- Otero-Schmitt, J., Cruz, R., García, C. & Rolán-Alvarez, E. 1997. Feeding strategy and habitat choice in *Littorina saxatilis* (Gastropoda: Prosobranchia) and their role in the origin and maintenance of a sympatric polymorphism. *Ophelia* **46**: 205–216.
- Reid, D.G. 1996. *Sytematics and Evolution of Littorina*. Dorset Press, Dorchester.
- Rolán-Alvarez, E. & Caballero, A. 2000. Estimating sexual selection and sexual isolation effects from mating frequencies. *Evolution* **54**: 30–36.
- Rolán-Alvarez, E., Rolán, E. & Johannesson, K. 1996. Differentiation in radular and embryonic characters, and further comments on gene flow, between two sympatric morphs of *Littorina saxatilis* (Oliv). *Ophelia* **45**: 1–15.
- Rolán-Alvarez, E., Johannesson, K. & Erlandsson, J. 1997. The maintenance of a cline in the marine snail *Littorina saxatilis*: the role of home site advantage and hybrid fitness. *Evolution* **51**: 1838–1847.
- Rolán-Alvarez, E., Erlandsson, J., Johannesson, K. & Cruz, R. 1999. Mechanisms of incomplete prezygotic reproductive isolation in an intertidal snail: testing behavioural models in wild populations. *J. Evol. Biol.* **12**: 879–890.
- Rundle, H.D., Nagel, L., Boughman, J.W. & Schluter, D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**: 306–308.

- Schluter, D. 2001. Ecology and the origin of species. *TREE* **16**: 372–380.
- Small, M.P. & Gosling, E.M. 2000. Genetic structure and relationships in the snail species complex *Littorina arcana* Hannaford Ellis, *L. compressa* Jeffreys and *L. saxatilis* (Olivier) in the British Isles using SSCPs of cytochrome-b fragments. *Heredity* **84**: 692–701.
- Sokal, R.R. & Rohlf, F.J. 1995. *Biometry*, 3th edn. Freeman, New York.
- Takimoto, G., Higashi, M. & Yamamura, N. 2000. A deterministic genetic model for sympatric speciation by sexual selection. *Evolution* **54**: 1870–1881.
- Turelli, M., Barton, N.H. & Coyne, J.A. 2001. Theory and speciation. *TREE* **16**: 330–343.
- Via, S. 2001. Sympatric speciation in animals: the ugly duck grows up. *TREE* **16**: 381–390.
- Wilding, C.S., Butlin, R.K. & Grahame, J. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected by AFLP markers. *J. Evol. Biol.* **14**: 611–619.
- Wright, S. 1951. The genetical structure of populations. *Ann. Eugen.* **15**: 323–354.

Received 5 March 2004; revised 30 April 2004; accepted 5 May 2004