PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Stankowski S, Westram AM, Zagrodzka ZB, Eyres I, Broquet T, Johannesson K, Butlin RK. 2020 The evolution of strong reproductive isolation between sympatric intertidal snails. *Phil. Trans. R. Soc. B* **375**: 20190545.
http://dx.doi.org/10.1098/rstb.2019.0545

Accepted: 25 March 2020

One contribution of 19 to a theme issue 'Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers'.

Subject Areas:
genetics, genomics, evolution

Keywords:
demographic inference, *Littorina arcana*, *Littorina saxatilis*, reproductive mode, speciation

Authors for correspondence:
Sean Stankowski
e-mail: seanstankowski@gmail.com
Roger K. Butlin
e-mail: r.k.butlin@sheffield.ac.uk

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5018468.

THE ROYAL SOCIETY PUBLISHING

The evolution of strong reproductive isolation between sympatric intertidal snails

Sean Stankowski[1], Anja M. Westram[2], Zuzanna B. Zagrodzka[1], Isobel Eyres[1], Thomas Broquet[3], Kerstin Johannesson[4] and Roger K. Butlin[1,4]

[1]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK
[2]Institute of Science and Technology Austria (IST Austria), Klosterneuburg, Austria
[3]CNRS and Sorbonne Université, Station Biologique de Roscoff, Roscoff, France
[4]Department of Marine Sciences, Tjärnö Marine Laboratory, University of Gothenburg, Strömstad, Sweden

SS, 0000-0003-0472-9299

The evolution of strong reproductive isolation (RI) is fundamental to the origins and maintenance of biological diversity, especially in situations where geographical distributions of taxa broadly overlap. But what is the history behind strong barriers currently acting in sympatry? Using whole-genome sequencing and single nucleotide polymorphism genotyping, we inferred (i) the evolutionary relationships, (ii) the strength of RI, and (iii) the demographic history of divergence between two broadly sympatric taxa of intertidal snail. Despite being cryptic, based on external morphology, *Littorina arcana* and *Littorina saxatilis* differ in their mode of female reproduction (egg-laying versus brooding), which may generate a strong post-zygotic barrier. We show that egg-laying and brooding snails are closely related, but genetically distinct. Genotyping of 3092 snails from three locations failed to recover any recent hybrid or backcrossed individuals, confirming that RI is strong. There was, however, evidence for a very low level of asymmetrical introgression, suggesting that isolation remains incomplete. The presence of strong, asymmetrical RI was further supported by demographic analysis of these populations. Although the taxa are currently broadly sympatric, demographic modelling suggests that they initially diverged during a short period of geographical separation involving very low gene flow. Our study suggests that some geographical separation may kick-start the evolution of strong RI, facilitating subsequent coexistence of taxa in sympatry. The strength of RI needed to achieve sympatry and the subsequent effect of sympatry on RI remain open questions.

This article is part of the theme issue 'Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers'.

1. Introduction

The evolution of strong reproductive isolation (RI) is fundamental to the origins and maintenance of biological diversity, especially in situations where geographical distributions of taxa broadly overlap [1,2]. But how do strong barriers acting in sympatry originate? In general terms, the evolution of strong RI is thought to arise most frequently through the coupling together of multiple components of isolation, leading to a stronger overall barrier to gene flow [3]. Because coupling often requires the buildup of linkage disequilibrium (LD) among barrier loci, this creates something of a paradox: the evolution of RI is most difficult in scenarios with high gene flow, which is where strong barriers are needed the most [4–6]. Understanding the circumstances and processes underlying the evolution of strong isolating barriers is, therefore, a major goal of speciation research [7,8].

One of the simplest and most common explanations for the presence of strong isolation in sympatry is that the evolution of barriers preceded the geographical overlap of the taxa in question [1,9]. This is because the coupling of barrier effects occurs automatically in models of speciation that include partial or complete

© 2020 The Author(s) Published by the Royal Society. All rights reserved.

geographical isolation [7]. For example, in allopatric divergence, the buildup of LD among barrier loci occurs as a simple by-product of geographical isolation, which may allow populations to coexist upon secondary contact (SC) [10,11]. In models of parapatric speciation, coupling occurs when multiple isolating traits are subject to divergent selection on opposite sides of a sharp ecological boundary [12,13]. This scenario may ultimately lead to the broad coexistence of populations if isolation is strong and contrasting habitats become patchily distributed. Although considered less common, it is possible for strong RI to evolve with little or no geographical separation [14]. For example, strong barriers may arise through the origin of novel traits that have a large effect on both local adaptation and assortment (i.e. multiple-effect or 'magic' traits) [15,16], and initially weak barriers may be strengthened via the coupling of additional barrier effects by reinforcement or related processes [3]. For pairs of sister taxa whose ranges overlap, with little or no current gene exchange, the empirical challenge is to disentangle the sequence of events that led to this strong RI. Was there a period without gene flow, or with low gene flow? Did this period result in strong RI or was isolation strengthened following SC and range overlap? What was the nature of the barriers to gene flow at different points in this history?

In this study, we focus on determining the current strength of RI and the history of divergence between two closely related, sympatric and syntopic intertidal snails in the genus *Littorina*. Recent studies of speciation in *Littorina* have focused mainly on the evolution of RI between parapatrically distributed ecotypes of *Littorina saxatilis* that inhabit different areas of the intertidal zone [17–20]. In many locations around the North Atlantic, including France, the UK, Sweden and Spain, it is possible to find phenotypically divergent populations of *L. saxatilis* in close proximity: a 'crab' ecotype, which has evolved thick, large shells in response to a high level of crab predation, and a 'wave' ecotype, which has evolved small, thin shells as an adaptation to withstand strong wave exposure [21]. Demographic reconstructions suggest that the ecotypes have diverged without periods of allopatry [17]. Despite the striking parallelism of their phenotypes, the strength of the barrier to gene flow appears to vary among locations, suggesting that there is a continuum of RI within this system. The current distributions of the ecotypes only ever overlap slightly, for example, in mosaic habitats covering a few metres in the mid-shore in Galicia [22,23]. It seems that RI between the ecotypes is never complete and, combined with spatial separation of habitats, this limits geographical coexistence to narrow regions with varying degrees of hybridization.

The relationship between *L. saxatilis* and its proposed sister species *Littorina arcana* provides a stark contrast to ecotype differentiation. The range of *L. arcana* (British Isles, northern France and a few localities in Norway) is completely included within the range of *L. saxatilis* (all North Atlantic coasts; [24]). On shores where they occur together, they are syntopic, often touching one another in the same rock crevices, with weak and variable spatial separation at most [25,26]. *Littorina arcana* shows parallel ecotypic variation to *L. saxatilis* at some sites [24]. The species can be distinguished based on differences in allozyme allele frequencies [27,28], but laboratory crosses generating viable and fertile offspring have been reported [27]. The only reliable phenotypic character distinguishing the species is their reproductive mode: while both species show direct development (i.e. they lack a planktonic larval stage), *L. arcana* lays egg masses on the substrate while *L. saxatilis* embryos remain in a brood pouch until they have completed development [24]. Brooding is a derived condition that appears to have evolved recently (1.7–0.06 Ma [29]) from an ancestral egg-laying population. In practice, the two species are indistinguishable in the field and only adult females can be distinguished reliably following dissection.

In this paper, we address three questions surrounding the evolutionary and reproductive relationships between the sympatric taxa *L. arcana* and *L. saxatilis*. First, what are the evolutionary relationships between snails with different reproductive modes: are they distinct taxa, as suggested by the current taxonomy for the group? Second, is there evidence for RI between snails with different reproductive modes and, if so, what is its strength? Finally, what is the demographic history associated with divergence of these taxa? More specifically, does it appear that RI has evolved in sympatry, perhaps facilitated by the rapid divergence in reproductive mode, or were their sympatric distributions made possible by isolation that evolved during a period of physical separation?

2. Results and discussion

(a) Brooding and egg-laying snails are closely related, but genetically distinct

As a first step to understanding the evolutionary relationships among snails with different reproductive modes, we constructed a genome-wide phylogeny using female snails with brooding ($n = 11$) and egg-laying ($n = 13$) anatomy collected from wave-swept habitats at locations where *L. arcana* and *L. saxatilis* coexist in sympatry: (i) Ravenscar in England, (ii) Anglesey in Wales, and (iii) Roscoff in France (electronic supplementary material, figure S1). Coupled with the large distances between these sample locations and assuming that brooding is the derived strategy [24], the coexistence of the taxa at each location provides us with clear predictions for the patterns of clustering that we would expect to see under different divergence histories. For example, if the origin of brooding coincided with the evolution of strong genome-wide RI, we would expect brooding individuals to form a distinct monophyletic clade that is either sister to *L. arcana* or nested within *L. arcana*, depending on when and where brooding originated. In an alternative scenario, where gene flow between brooders and egg layers has been high across most of the genome and throughout history, we would expect individuals to cluster based on their sampling location rather than by reproductive anatomy.

The maximum-likelihood phylogeny, which was constructed from a concatenated alignment of approximately 13 million single nucleotide polymorphisms (SNPs) and rooted with the closely related species, *Littorina compressa* ($n = 2$), provides a well-supported set of relationships that is inconsistent with extensive genome-wide gene flow between snails with different modes of reproduction (figure 1*a*). Specifically, rather than clustering by their sampling location, samples formed monophyletic groups according to their reproductive anatomy. Within the brooding and egg-laying clades, samples then clustered by sample location, as is expected given the limited opportunity for gene flow between populations separated by large geographical distances. Despite forming distinct clades, the levels of allele frequency differentiation (F_{ST}) between *L. arcana* and *L. saxatilis* were relatively low compared with

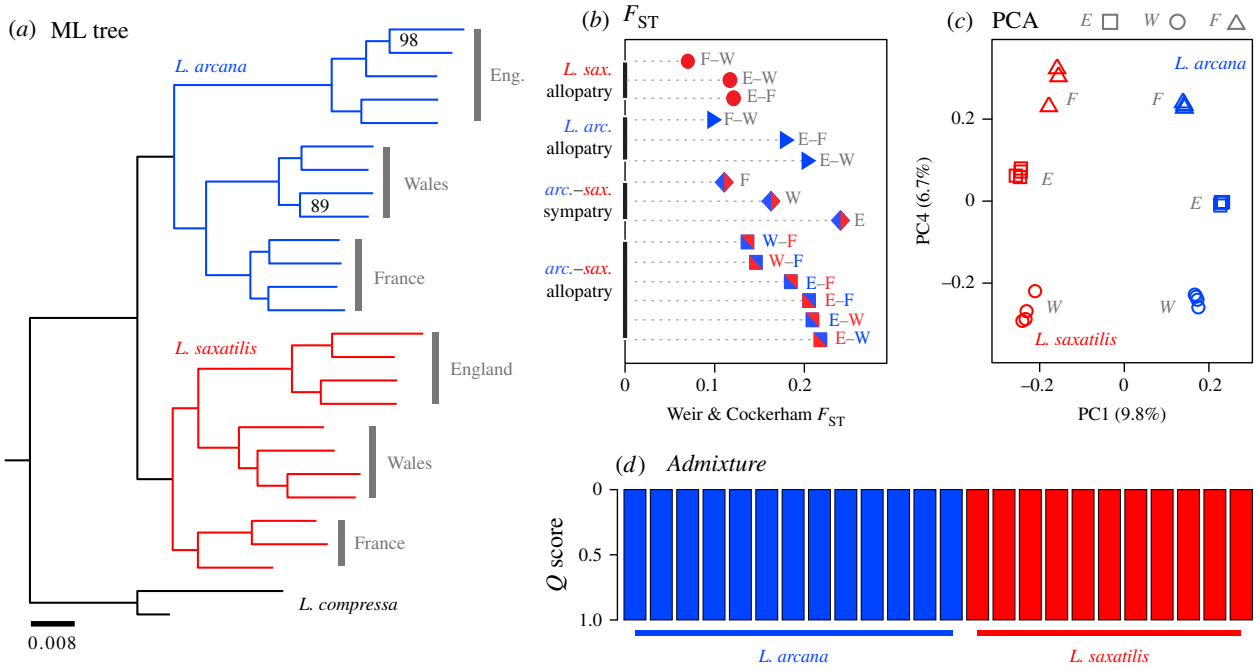

Figure 1. Evolutionary relationships between *Littorina arcana* (egg-laying) and *Littorina saxatilis* (brooding). (*a*) Maximum-likelihood (ML) phylogeny constructed for snails with egg-laying or brooding anatomy from three locations (England, France and Wales) based on a concatenated alignment of genome-wide SNPs. The phylogeny is rooted with the outgroup, *Littorina compressa*. All nodes have 100% bootstrap support, except for two nodes where the support is indicated by the numbers. (*b*) Estimates of F_{ST} between sample locations, both within and between *L. arcana* and *L. saxatilis*. E, England; F, France; W, Wales. (*c*) PC1 and PC4 from a principal component analysis conducted on the same dataset. (*d*) Ancestry assignment scores (*Q*) from a run of *Admixture* assuming two populations ($K = 2$) without prior information. Each horizontal bar shows the proportions of the genome assigned to each cluster. (Online version in colour.)

what one would generally expect when comparing intra- with interspecific levels of divergence (figure 1*b*). For the within-taxon comparisons, estimates of F_{ST} tended to be higher between populations of *L. arcana* (range = 0.10–0.20) than between populations of *L. saxatilis* (range = 0.07–0.12). This is consistent with a recent spread of *L. saxatilis* following its origin [24], but could also reflect higher levels of gene flow across its range compared with *L. arcana*. The only exception was the comparison between French and Welsh *L. arcana*, which were less differentiated than two of the three *L. saxatilis* comparisons. Of the nine possible comparisons that could be made between the taxa (three sympatric comparisons and six allopatric comparisons), six fell within the range of values for the within-taxon comparisons (figure 1*b*). The lowest between-taxon F_{ST} was observed between sympatric *L. arcana* and *L. saxatilis* in France ($F_{ST} = 0.11$), which were less differentiated than four of the six within-taxon comparisons, and the highest value was observed between the sympatric populations in England ($F_{ST} = 0.24$).

Principal component analyses conducted on the genotype matrix, including different subsets of the data (see Material and methods), revealed the same pattern of clustering as the phylogenetic analysis, but also provided possible evidence for some gene flow between *L. arcana* and *L. saxatilis* (figure 1*c*). The first principal component (PC1), which explained 9.8% of the variation in the data, clearly separated egg-laying and brooding individuals. Consistent with levels of F_{ST}, separation on PC1 was largest in England and smallest in France. PC2 (8.2%) and PC3 (6.7%) were associated with differences that were specific to English *L. saxatilis* and English *L. arcana*, respectively. However, PC4 (6.7%) clearly separated samples according to the three locations but not by taxon (figure 1*c*). This pattern of clustering is not expected if two allopatric populations evolved complete RI prior to their broad coexistence, as patterns of variation would be expected to diverge randomly among locations as a result of drift. Instead, this pattern indicates that variation at a common group of markers has become shared between sympatric populations while also becoming differentiated among the localities. One explanation for this result is that the barrier to gene flow between *L. saxatilis* and *L. arcana* is porous, allowing a common pattern of genome-wide introgression to arise at all three locations owing to local gene flow. Another plausible but less parsimonious explanation is parallel evolution, where both taxa adapted to environmental conditions in England, Wales and France using the same set of loci.

Despite their close evolutionary relationships and possible evidence for gene flow, it is clear that this sequenced sample does not include any individuals that are the product of recent hybridization between *L. arcana* and *L. saxatilis*. If we had sampled early-generation hybrid or backcrossed individuals, we would expect to see evidence for samples with mixed ancestry from a model-based clustering analysis. Assuming a model with two populations ($K = 2$) and no prior information, all samples were strongly assigned to one of the two clusters that coincided with the alternative reproductive modes ($Q > 0.9999$ or <0.0001; figure 1*d*). As for the principal components analysis (PCA), nearly identical results were observed in analyses that used random or strategically selected (one SNP per assembly contig) subsets of the data. Given that we found no individuals of mixed ancestry, we can confidently conclude that none of the individuals is the product of recent hybridization between *L. arcana* and *L. saxatilis*.

(b) Strong reproductive isolation with evidence for limited ongoing introgression

Although the above analyses did not reveal any recent hybrids between *L. arcana* and *L. saxatilis*, they do not provide

a thorough test for the presence of hybrids in the field because we only sequenced a handful of adult females with clear brooding or egg-laying anatomy. Given that hybrids may have aberrant reproductive anatomies, and may be rare if RI is strong, we conducted a more thorough test for hybridization by genotyping 3092 individuals collected from the three locations at 80 SNPs that were highly differentiated between *L. arcana* and *L. saxatilis* based on the 24 whole-genome sequences ($F_{ST} > 0.95$). We then calculated a hybrid index (HI) score for each individual by polarizing the alleles at each SNP according to whether they were more common in *L. arcana* (0) or *L. saxatilis* (1) and expressing the sum of values across the 80 loci as a proportion of the total number of alleles genotyped for each individual, which varied if data at some loci were missing. Given the low level of allele sharing at these SNPs in the sequenced sample, pure *L. arcana* are expected to have HI scores near 0, while *L. saxatilis* individuals are expected to have scores near 1. Because the loci are spread broadly across the 17 linkage groups identified in *L. saxatilis* [18], the ancestry scores for recent hybrids and backcrosses are expected to follow patterns of segregation for multiple unlinked loci.

Similar to the analysis of whole-genome sequences, the genotyping assay revealed strongly bimodal distributions of HI score, indicative of strong RI between *L. arcana* and *L. saxatilis* (figure 2). The modes, which were positioned at opposite ends of the range, close to 0 and 1, coincided almost perfectly with the differences in reproductive strategy: over 95% of females (1117 out of 1173) had the reproductive anatomy predicted by their HI score. Conflicting assignments (56) were almost always explained by the anatomical misclassification of individuals in the early stages of sexual maturity, as the egg-laying and brooding structures look very similar early in their development [24]. The genetic assignments, which include male and reproductively immature snails, also show that the ratios of *L. arcana* and *L. saxatilis* were very different among the three locations ($\chi^2 = 174.57$, d.f. = 2, $p < 0.00001$; figure 2). The difference was driven mainly by France, where *L. arcana* was roughly three times more common than *L. saxatilis*. By contrast, the ratios of *L. arcana* to *L. saxatilis* were similar in England and Wales (without France: $\chi^2 = 2.81$, d.f. = 1, $p = 0.102$), where egg layers and brooders coexist at approximately equal frequency. Finally, the proportions inferred from genetic and morphological assignment were quite similar (figure 2), indicating that reproductive females can be used to provide accurate estimates of the relative abundance of these taxa in the field.

Hybrid populations simulated using the program *HYBRIDLAB* [30] clearly show that none of the individuals was an early-generation hybrid (F_1–F_3) or first-generation backcross (electronic supplementary material, figure S2), although hybrid index scores within the observed ranges can be generated after only a few backcross generations (electronic supplementary material, figure S2). Also, there was some variability in the hybrid index (HI) scores associated with each reproductive strategy (figure 2). We expected to see some variability, given that most of the loci that we chose were not perfectly diagnostic owing to low levels of within-species polymorphism (mean heterozygosity = 6.8%). However, shared polymorphism is also a predicted outcome of introgression, which could result from rare hybridization events that would be hard to detect even with very large samples. To determine

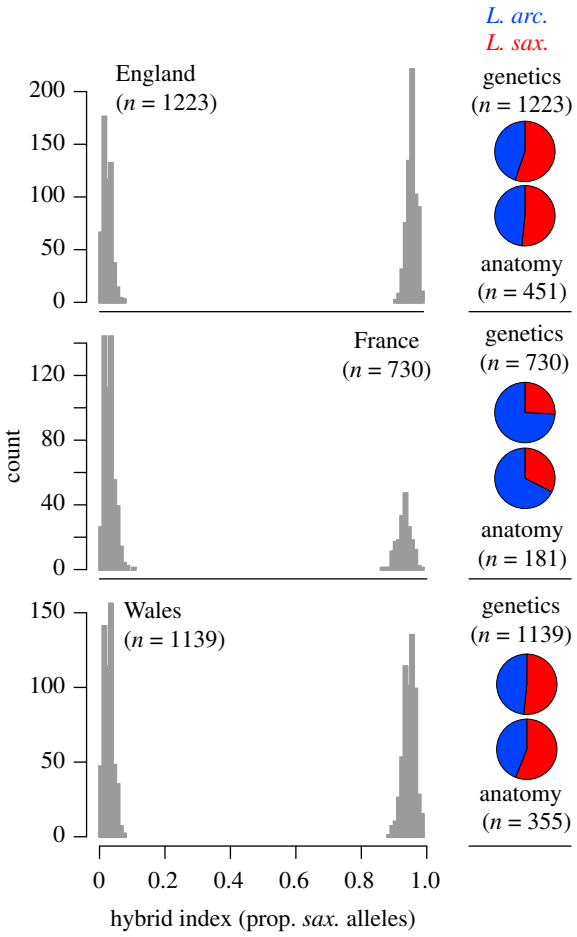

Figure 2. Hybrid index scores and species assignments determined by SNP genotyping. The histograms show hybrid index scores for 3092 male, female and reproductively immature snails sampled from England, Wales and France. The pie charts show the proportion of snails classified to each species using both the genetic data (all individuals) and female reproductive anatomy (reproductively mature females only). (Online version in colour.)

if the features of the observed distributions could be explained by ancestral polymorphism within two reproductively isolated populations, we simulated a large number of samples from well-mixed polymorphic populations without gene flow (1000 iterations for each of our six samples, with observed sample sizes and levels of missing data, using the allele frequencies observed within each distribution in each of the three locations). If the features of the observed distributions (i.e. the mean and variance) are similar to those obtained for the simulated ones, then the variability in the hybrid index scores can be explained without needing to invoke introgression. However, differences between the observed and simulated distributions may indicate that some gene flow is occurring between the populations. In the case of the mean HI, introgression from *L. saxatilis* into *L. arcana* and *L. arcana* into *L. saxatilis* would generate larger and smaller values in the recipient taxon, respectively, as a result of the increased skew generated by introgression. For the variance of the HI, an increase is expected in both directions, also mainly because of a tail of introgressed genotypes. For all six populations, the observed mean fell well within the distribution of simulated means (p-values ranging from 0.194 to 0.562) or, in the case of Welsh *L. arcana*, was slightly lower ($p = 0.013$), so consistent with the absence of gene flow between the populations. However, the variance was consistently larger than the simulated variances in all three populations of *L. arcana*

Table 1. Akaike information criterion (AIC) for nine demographic models fitted to JSFS for each location using *∂a∂i*. SI, ancestral population split without subsequent gene flow; SC, secondary contact; AM, ancient migration; IM, isolation with migration; IM2R8, isolation with different rates of migration in epochs 1 and 2. SC2*m*, AM2*m*, IM2*m* and IM2R82*m*, as above, but allowing migration rates to vary between two groups of loci in the genome (2*m*).

	England		France		Wales	
model	AIC	−ΔAIC	AIC	−ΔAIC	AIC	−ΔAIC
SI	1510.35	−833.92	979.21	−371.77	1595.98	−758.62
SC	854.69	−178.26	661.24	−53.80	953.69	−116.33
AM	817.58	−141.15	674.27	−66.83	956.76	−119.40
IM	852.69	−176.26	672.36	−64.92	954.79	−117.43
IM2R8	820.93	−144.40	664.15	−56.71	957.26	−119.90
SC2*m*	735.27	−58.84	611.78	−4.34	839.57	−2.12
AM2*m*	734.94	−58.51	615.62	−8.18	855.14	−17.78
IM2*m*	733.74	−57.31	613.62	−6.18	853.14	−15.78
IM2R82*m*	676.43	0	607.44	0	837.36	0

(Wales: $p = 0.010$, France: $p = 0.001$, England: $p = 0.041$), most notably in France, where all 1000 of the simulated datasets had smaller variances than the observed distribution. By contrast, only one population of *L. saxatilis*, England, had a variance that was significantly larger than expected ($p = 0.022$). Although it is important to acknowledge that violations of our simple model (e.g. population structure within locations and differences in patterns of recombination between species) could also cause differences between the simulated and observed distributions, this result suggests that there is a very low level of current gene flow between the taxa, with a stronger signal of introgression from *L. saxatilis* into *L. arcana*.

(c) Strong reproductive isolation may have evolved with a degree of geographical separation

The evidence for strong RI between these broadly overlapping, closely related taxa, raises the possibility that the barrier to gene flow has evolved in sympatry. To test this hypothesis, we used a modified version of *∂a∂i* [31,32] to model the demographic history of divergence from the joint-site-frequency spectrum (JSFS), separately in each location, using our whole-genome sequences.

The base model consisted of a single ancestral population that split into two populations and then diverged for T generations. We compared the fit of nine variants of this model (electronic supplementary material, figure S3), including (i) strict isolation (SI), where an ancestral population split without subsequent gene flow, (ii) SC, where a split coincided with a period of allopatry followed by gene flow upon recontact, (iii) ancient migration (AM), where gene flow occurred early in speciation but ceased at some point during the split, (iv) speciation with gene flow (IM), where gene flow occurred at a constant rate during divergence, and (v) a two-rate model of speciation with gene flow (IM2R8), where the migration rate changed at a point in time, T_m, during divergence. For the models that included migration (ii–v), we also fitted them allowing for heterogeneous gene flow across the genome (SC2*m*, AM2*m*, IM2*m* and IM2R82*m*) by allowing migration rates to vary between two groups of loci: a class that experienced migration between populations at a rate m, and a class that was impacted by RI, experiencing reduced migration at a rate m_e. Because *∂a∂i* can only model divergence of two populations and assumes no spatial structure, the model fitting was conducted separately for each of the three locations.

Demographic models with migration were always a much better fit for the observed data than the model of SI (table 1; electronic supplementary material, figure S4). For all three sample locations, the Akaike information criterion (AIC) score for the SI model was several hundred points higher than the worst-fitting model with gene flow. The difference was largest in England, where the SI model was 834 AIC points worse than the best-fitting model, and smallest in France, where it was 372 points worse than the best fit. Furthermore, of the models with gene flow, those with heterogeneous migration were a much better fit (more than 40 AIC points) than the models with a single migration rate. Together, these results rule out the completion of speciation prior to range overlap and suggest that the divergence history of *L. arcana* and *L. saxatilis* has involved at least some gene flow between them.

The specific details of the pattern of historical gene flow between *L. arcana* and *L. saxatilis* are, however, less clear, as most of the 2*m* models (i.e. models that include one migration rate for neutral loci and another for barrier loci experience that experience a reduction in m_e) tended to fit the data similarly well. In all three cases, the two-rate model of speciation with gene flow and heterogeneous migration (IM2R82*m*) was the best-fitting model. However, only in England was this model a clear front-runner over the IM2*m*, AM2*m* and SC2*m* models, which were between 57 and 59 AIC points worse than the IM2R82*m* model. In France, the other 2*m* models were within 10 AIC points of the IM2R82*m*, with the SC2*m* model, which was only 4 points worse, being the next best fit. The SC2*m* model was also the second-best-fitting model in Wales, being only 2 AIC points behind the IM2R82*m* model. In all three locations, the rates of migration were much lower in the first period of migration, which accounted for between 5% (France) and 11% (England) of the total split time. Thus, it appears that gene flow was lower than its current level early in divergence, possibly owing to a relatively short period of geographical isolation.

In general, the ML parameters from the best-fitting model provide further evidence for strong RI of *L. arcana* and *L. saxatilis* (table 2). For the best-fitting model (IM2R82*m* in

Table 2. Parameter estimates from the best-fitting demographic model (IM2R82*m*) at each location, where *L. saxatilis* is population 1 and *L. arcana* is population 2. $\theta_1 = 4\text{Nref}\mu$, where μ is the mutation rate. *n*1, size of population 1 after split proportional to the effective size of the ancestral population, Nref; *n*2, size of population 2 after split proportional to Nref. M_{12}, migration from pop. 2 to pop. 1, which is equal to $2\text{Nref}m_{12}$, where m_{12} is the proportion of chromosomes in each generation that are new migrants from population 2 to population 1. M_{12}, migration from pop. 1 to pop. 2; Me_{21}, effective migration from pop. 1 to pop. 2. T_s, the scaled time between the split and present (in units of 2Nref generations); T_m, the time of change of migration rates; *P*, the proportion of the genome evolving without a barrier to gene exchange.

				migration period 1			migration period 2						
	θ	*n*1	*n*2	M_{12}	M_{21}	Me_{12}	Me_{21}	M_{12}	M_{21}	Me_{12}	Me_{21}	T_s	T_m	*P*
Wales	4410	2.2	3.81	2.38×10^{-7}	2.38×10^{-7}	1.30×10^{-18}	3.79×10^{-5}	2.58	29.6	0.669	0.35	0.947	0.0635	0.235
France	4260	3.22	4.71	8.07×10^{-5}	8.07×10^{-5}	1.18×10^{-6}	2.71×10^{-13}	3.67	29	0.581	0.0504	0.947	0.0472	0.329
England	3730	1.02	2.44	6.18×10^{-4}	6.18×10^{-4}	1.04	1.84×10^{-9}	0.67	11.4	0.135	0.0462	1.53	0.169	0.237

all cases), the estimated proportion of the genome experiencing reduced migration as a result of RI was high (mean 73%), ranging from 67% (France) to 77% (Wales). Moreover, the migration rates inferred between the taxa were highly asymmetrical (table 2). For the IM2R82*m* model, this was most notable in the second period of gene flow, where the migration rate from *L. saxatilis* to *L. arcana* was between 8 (France) and 17 (England) times higher than from *L. arcana* to *L. saxatilis*. Given that the taxa are roughly equally common at two of the three sites (figure 2), the asymmetrical gene flow cannot be explained by differences in their relative abundance. Rather, the result is more likely to reflect asymmetry in the strength of RI, with a stronger barrier from *L. arcana* to *L. saxatilis*.

(d) Conclusions, broader implications and next steps

Using whole-genome sequences and SNP genotyping of large field-collected samples, our study has clarified the evolutionary and reproductive relationships between *L. arcana* and *L. saxatilis* using samples spanning a large part of their range [24]. Despite some controversy in the past about the taxonomic status of snails with divergent reproductive modes [24], patterns of genome-wide variation clearly show that the snails with different reproductive anatomies share a recent common ancestor and represent distinct taxa. SNP genotyping of more than 3000 snails failed to recover any individuals that were consistent with being recent hybrid or backcross individuals, confirming that RI between *L. arcana* and *L. saxatilis* is very strong. There is, however, evidence for a very low level of ongoing introgression between them, primarily in one direction (from *L. saxatilis* to *L. arcana*), despite using a set of SNPs selected on the basis of high differentiation. The presence of strong, yet asymmetrical RI is supported by demographic analysis of these populations, providing further evidence that isolation between these taxa is not complete.

After inferring the evolutionary relationships and strength of RI in this system, we wanted to gain insight into the demographic history underlying the divergence of these taxa. Although some caution should be exercised because several different models fit the data quite well, our results suggest that *L. arcana* and *L. saxatilis* have coexisted for a long period with ongoing gene flow, but may have diverged initially during a period of strong but incomplete geographical separation. In the light of the current distributions of these taxa, we propose the following biogeographic history as a working hypothesis for their divergence: the brooding strategy—the trait that defines *L. saxatilis*—probably arose in a location where the egg-laying reproductive strategy is now absent. The success of the brooding strategy allowed the lineage to expand, where it may have replaced some existing egg-laying populations, possibly facilitated by broad-scale environmental change. However, brooding snails subsequently arrived in areas where sufficient ecological opportunity and strong enough RI allowed reproductive strategies to co-occur. The evolution of RI was not completed in allopatry. Sympatry may have led to further strengthening of RI but it is not yet complete, despite a long period of coexistence. Although RI is very strong, its effects are still uneven around the genome. Further analyses, including more sequenced genomes, more locations and more complex demographic models, may enhance our understanding of the demographic history of these populations in the future.

In addition to learning more about speciation between these taxa, our new understanding highlights *Littorina* as a model for understanding how different forces influence the evolution of RI. To date, most studies of speciation in *L. saxatilis* have focused on the parallel evolution of RI between locally adapted 'crab' and 'wave' ecotypes owing to divergent selection in different areas of the intertidal zone [17,18,20,33,34]. Patterns of molecular and phenotypic variation suggest that the strength of the barrier between the ecotypes varies among locations, suggesting that there may be a continuum of RI in this system. For example, in the British Isles and France, the level of genetic differentiation is very low (F_{ST} range = 0.01–0.03 [20]) and there is extensive hybridization in areas where the ecotypes meet [35]. In Sweden, the level of genetic differentiation is higher (F_{ST} range = 0.05–0.07 [20]) and the phenotypic divergence of the ecotypes is more substantial [17], potentially owing to the presence of a stronger overall barrier to gene flow between them [18]. RI appears to be even stronger in Spain (F_{ST} range = 0.09–0.12 [20]), as the ecotypes are able to coexist in a narrow zone with much lower admixture than in Sweden [23]. Our comparison of *L. arcana* and *L. saxatilis* (F_{ST} range = 0.11–0.24) extends the continuum to a point where RI is almost complete. However, the route towards strong RI between these taxa is very different from the formation of ecotypes in *L. saxatilis*, as it is unlikely to have evolved as a result of divergent selection acting across a sharp ecological gradient. Rather, it appears more substantial geographical separation may have facilitated the evolution of RI, possibly facilitating the divergence in reproductive mode and evolution of strong intrinsic isolation.

The findings of this study also set the stage for more detailed investigations into the genomic basis of RI between these more strongly isolated taxa. In *L. saxatilis*, the genomic landscape of ecotype divergence is characterized by a small number of highly differentiated regions, most of which overlap with putative chromosomal inversions that appear to be involved in local adaption in multiple locations [18–20]. In places where hybridization is common, the effects of gene flow are widespread across the genome [18], as is predicted when isolating barriers are highly porous. Given that the barrier to gene flow between *L. arcana* and *L. saxatilis* is very strong and involves most of the genome (73%), future studies are expected to reveal a very different genomic landscape of speciation in this system. Specifically, we expect highly differentiated loci to be widespread across the genome rather than being restricted to a handful of genome regions. However, chromosomal inversions may also play a role between these taxa, as has been observed in *L. saxatilis*. The inclusion of populations of *L. saxatilis* from areas where *L. arcana* is absent will also enable more powerful tests of our hypothesis that the same genomic landscape of introgression has evolved in each location owing to ongoing gene flow in sympatry.

Another obvious question arising from this study concerns the basis of RI between *L. arcana* and *L. saxatilis*: specifically, what are the isolating barriers that are responsible for keeping these taxa distinct? The coexistence of these taxa across such a large area implies that they are occupying distinct ecological niches on the shore, which may also contribute to isolation between them. However, the habitat differences are not obvious, as these taxa occupy the same areas of the intertidal zone. The most obvious candidate for a barrier trait is the difference in female reproductive strategy between the taxa, which includes a difference in both the mode (egg-laying in *L. arcana* versus brooding in *L. saxatilis*) and timing (seasonal in *L. arcana* versus year-round in *L. saxatilis*) of reproduction. However, other barriers, including assortative mating and microhabitat choice, may also play important roles, or may even be the primary barrier to gene flow. Further studies, including mating trials, laboratory crosses and fine-scale ecological measurements, are currently being conducted to address these questions.

A major challenge, in this system and others, is to determine the nature and extent of the barrier to gene flow at the time when their distributions became broadly overlapping: how much isolation was required to permit their coexistence and how much has evolved subsequently owing to reinforcement or related processes? Three-spine sticklebacks show a similar continuum of levels of isolation, from clinal introgression at habitat boundaries in lake–stream [36] and marine–freshwater [37] ecotype pairs to very low levels of gene flow in benthic–limnetic ecotype pairs [38] and between Pacific and Japan Sea species [39]. In this case, there is an interesting contrast between sympatry of the benthic–limnetic pairs, where divergent adaptation clearly played a part in generating RI [40], and parapatry of the Pacific–Japan Sea species, where a period of spatial separation may have been more important than ecological differentiation in generating isolation [39]. A combination of ecological differentiation, range overlap and low levels of introgression can be found in a diverse range of other systems, such as flatfish [41,42], poplars [43], lampreys [32] and flies [44]. However, this is also true for cases where the history is more likely to be dominated by divergence in allopatry (e.g. newts [45]; gulls [46]). This suggests two alternative routes to strong isolation, dominated by accumulation of incompatibilities in allopatry or by ecological divergence (cf. [47]), but understanding the similarities and differences between these pathways requires much more work on the accumulation of RI in these species, including *Littorina*.

3. Material and methods

(a) Sampling and anatomical assessment

More than 3300 snails were collected across three locations in the North Atlantic intertidal zone: near Ravenscar in England (54°24′28″ N ; 0°29′33″ W), on Anglesey in Wales (53°17′58″ N; 4°40′48″ W) and near Roscoff in France (48°41′42″ N; 4°06′42″ W). Samples were returned to the University of Sheffield and dissected to identify sex, reproductive status and reproductive mode according to Reid [24]. Snails were classified as adult if male or female reproductive anatomy could be identified. Samples without visible reproductive anatomy were classified as juvenile. Females were assigned to *L. saxatilis* if a brood pouch containing embryos was present. *Littorina arcana* were identified if a jelly gland was present. Foot tissue was stored in 100% ethanol prior to DNA extraction.

(b) Whole-genome sequencing, filtering, mapping and variant calling

From the samples described above, we selected reproductively mature females of each taxon (*L. arcana*: $n = 13$, *L. saxatilis*: $n = 11$ and *L. compressa*: $n = 2$) and prepared samples for whole-genome sequencing. DNA was extracted from a piece of foot tissue using a CTAB protocol [48]. Sequencing libraries were prepared using a TrueSeq DNA Nano gel-free library prep with a 350 bp insert and then sequenced on a HiSeq X (150 PE) to a

theoretical average depth of 15× coverage. Library preparation and sequencing were conducted by Edinburgh Genomics at the University of Edinburgh, UK.

Sequencing adaptors were removed and sequences were trimmed for low quality using Trimmomatic, and reads shorter than 70 bp were discarded. Raw reads were mapped to the *L. saxatilis* version 1 reference genome [18] using the BWAmem algorithm [49]. PCR duplicates were removed with biobambam2 [50]. Variant calling was performed using GATK 4.0.7 [51] following best practice recommendations and by executing steps documented in the short variant discovery pipeline [52]. Briefly, HaplotypeCaller was used to simultaneously call SNPs and indels and produce a gVCF file for each sample. To make this feasible with a large, highly fragmented reference genome, we performed this step on subsets of 1000 assembly contigs to produce 389 gVCFs per individual. GenotypeGVCFs was then used to perform joint genotyping across the samples, to produce a multi-sample VCF for each subset of contigs. The multi-sample VCFs were then concatenated together using bcftools to produce a complete VCF. Indel variants were removed from the VCF file. We retained bi-allelic sites with a quality score (Q) of 20 or greater and removed sites with a mean depth of fewer than 5 reads and more than 35 reads. Scripts used to perform these steps are available at https://github.com/seanstankowski/Littorina_A.

(c) Phylogenetic analysis, principal components analysis, admixture analysis and calculation of F_{ST}

We used *RAxML* v8 [53] to reconstruct the evolutionary relationships among the 24 samples from a concatenated alignment of all variant sites from across the genome using the HPC-PTHREADS-SSE3 implementation of the program with the GTRGAMMA model. Support for each node was determined via bootstrap analysis (100 replicates). The topology was rooted with *L. compressa* and rendered using Figtree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

PCA were conducted in *PLINK* v. 1.90 [54]. Initially, we ran the PCA including all of the variant sites, but we also performed additional analysis using random subsets of 10 000 SNPs and with one SNP per assembly contig.

We also used the program *Admixture* [55] to test for evidence of admixture between *L. arcana* and *L. saxatilis*. *Admixture* used the same statistical model as the program *Structure* [56], but calculates ancestry much more rapidly and accurately by using a numerical optimization algorithm [55]. We performed 10 replicate runs at $K = 2$ using different random seeds and they produced near-identical ancestry scores for each individual. We initially performed the analysis using all of the variant sites, but we also performed additional analysis using random subsets of 10 000 SNPs and with one SNP per assembly contig.

F_{ST} was estimated between all possible pairs of populations, both within and between taxa, according to Weir & Cockerham [57] using VCFtools [58].

(d) SNP genotyping assay and simulations

Because male and reproductively immature specimens cannot be reliably assigned to taxon using morphological or anatomical structures, we developed a genotyping assay with LGC Biosearch Technologies that we used to assign individuals to taxon and detect individuals of mixed ancestry. Using the 24 whole-genome sequences described above, we first identified markers that were highly differentiated between brooding and egg-laying females by calculating F_{ST} for each locus in VCFtools [58]. Because the sample sizes for the comparison are relatively small, and we wanted to ensure that SNPs were highly differentiated, we excluded sites with missing data as candidates. We excluded sites with a quality score less than 20, with more than two alleles and with an F_{ST} less than 0.95. Aided by a recently published genetic map for *L. saxatilis* [18], we selected SNPs on each linkage group, attempting to space them as evenly along each linkage group (LG) as possible.

In addition, we identified seven SNPs, each on a different LG, that we could use to identify and exclude any individuals of *L. compressa* from our analyses using the same criteria outlined above. Although *L. compressa* is relatively easy to identify based on shell morphology, these markers allowed us to find any misidentified individuals and exclude them from our analysis. All marker development, DNA extraction and genotyping was conducted by LGC Bioresearch Technologies. Details of the final set of markers can be found on the Dryad Digital Repository. Information needed to conduct the assay with LGC is provided on GitHub.

After filtering out individuals with more than 50% missing data (fewer than 40 genotypes), we were left with 3145 individuals with an average of 5.9% missing data. We polarized the alleles at each marker based on whether they were initially more common in *L. arcana* (0) or *L. saxatilis* (1) and then estimated a hybrid index score as their sum divided by the number of successfully genotyped alleles. The same procedure was used to estimate an '*L. compressa* score', using the seven additional SNP markers. Using this assay, we identified 54 individuals with *compressa* scores of 1. This included 20 deliberately sampled control individuals, and many individuals that were identified as *L. compressa* during dissection. No other individuals had an *L. compressa* allele at these loci. We, therefore, removed the individuals from further analyses, leaving 3092 samples.

SNP data were used to calculate an ancestry score or hybrid index as the proportion of alleles in an individual that were typical of *L. saxatilis*. To compare the distributions of these scores with expectations in the presence of gene flow or the absence of gene flow, we simulated genotypes under scenarios with and without hybridization between the two observed HI modes. For the first simulation, we used the program *HYBRIDLAB* [30] to simulate genotypes expected from early-generation hybrids (F_1–F_3) and backcrosses (BC_1–BC_4). For simplicity, we restricted this analysis to France, where the difference in the hybrid index scores between the left and right modes was the smallest. One thousand simulated F_1s were generated by randomly intercrossing individuals from each mode using the observed genotypes. Subsequent classes were generated by crossing simulated individuals either with the real populations (for backcrosses) or with themselves (for the F_2 and F_3). To simulate reproduction within each mode without gene flow, we generated individuals for each mode using the observed sample sizes in each location (six groups) by drawing an allele frequency for each locus from a binomial distribution with the observed allele frequency and sample size (of alleles) and then used this frequency to draw genotypes for each locus, by binomial sampling with sample size 2. We introduced missing values at random, according to the observed numbers of missing genotypes, and then calculated the hybrid index score for each individual. We generated 1000 simulated distributions for each group and compared the observed means and variances with the distributions of simulated means and variances by counting how many times the observed value exceeded the simulated values (or was lower than the simulated values for means of the *L. saxatilis* groups). For both simulations, hybrid index scores were calculated as described for the real data.

(e) Demographic analyses

We inferred the historical pattern of gene flow between *L. arcana* and *L. saxatilis* by fitting demographic models to the JSFS using a version of the program *∂a∂i* [31] that includes an improved optimization method and the ability to model semipermeable migration across the genome [32]. To minimize LD between

SNPs, we thinned the dataset so that no two SNPs were within 5 kb of each other. We then calculated the unfolded JSFS between *L. arcana* and *L. saxatilis*, separately for each location, using *L. compressa* as an outgroup to polarize alleles as ancestral or derived. After constriction, each JSFS was projected at a scale that maximized the number of segregating sites according to the *∂a∂i* manual (England: projection = 8,8, sites = 23 591; Wales: projection = 8,8, sites = 28 932; France: projection = 6,6, sites = 26 385). For each model, we performed 100 independent runs using randomly generated starting parameters, and we report the results for the run with the lowest AIC score.

Data accessibility. Raw sequence data are available on the short-read archive (SRA) as part of bioproject ID PRJNA626520. The SNP genotype dataset, SNP design information and scripts used to analyse the data are available at https://github.com/seanstankowski/Littorina_Philospohical_Transactions_B_2020.

Authors' contributions. S.S., A.M.W., R.K.B. and K.J. designed the study. S.S., A.M.W., R.K.B., Z.B.Z., T.B. and I.E. were involved in the organization and execution of fieldwork and/or the processing of samples. S.S. and R.K.B. analysed the data. S.S. wrote the manuscript with input from the other authors.

Competing interests. We declare we have no competing interests.

Funding. Funding was provided by the Natural Environment Research Council (NERC) and the European Research Council.

Acknowledgements. We thank Rui Faria, Nicola Nadeau, Martin Garlovsky and Hernan Morales for advice and/or useful discussion during the project. Richard Turney, Graciela Sotelo, Jenny Larson, Stéphane Loisel and Meghan Wharton participated in the collection and processing of samples. Mark Dunning helped with the development of bioinformatic pipelines. The analysis of genomic data was conducted on the University of Sheffield High-performance computer, ShARC. Jeffrey Feder and an anonymous reviewer provided comments that improved the manuscript.

References

1. Coyne JA, Orr HA. 2004 *Speciation*, pp. 276–281. Sunderland, MA: Sinauer Associates.
2. Price T. 2008 *Speciation in birds*. Greenwood Village, CO: Roberts & Company Publishers.
3. Butlin RK, Smadja CM. 2018 Coupling, reinforcement, and speciation. *Am. Nat.* **191**, 155–172. (doi:10.1086/695136)
4. Nosil P. 2013 Degree of sympatry affects reinforcement in *Drosophila*. *Evolution* **67**, 868–872. (doi:10.1111/j.1558-5646.2012.01817.x)
5. Nosil P, Crespi BJ, Sandoval CP. 2003 Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proc. R. Soc. Lond. B* **270**, 1911–1918. (doi:10.1098/rspb.2003.2457)
6. Yukilevich R, True JR. 2006 Divergent outcomes of reinforcement speciation: the relative importance of assortative mating and migration modification. *Am. Nat.* **167**, 638–654. (doi:10.1086/503120)
7. 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)
8. Butlin R *et al.* 2012 What do we need to know about speciation? *Trends Ecol. Evol.* **27**, 27–39. (doi:10.1016/j.tree.2011.09.002)
9. Mayr E. 1942 *Systematics and the origin of species*. New York, NY: Columbia University Press.
10. Orr HA. 1996 Dobzhansky, Bateson, and the genetics of speciation. *Genetics* **144**, 1331–1335.
11. Gavrilets S. 2004 *Fitness landscapes and the origin of species (MPB-41)*. Princeton, NJ: Princeton University Press.
12. Endler JA. 1977 Geographic variation, speciation, and clines. *Monogr. Popul. Biol.* **10**, 1–246.
13. Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011 The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol. Ecol.* **20**, 2044–2072. (doi:10.1111/j.1365-294X.2011.05080.x)
14. 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)
15. Servedio MR, Kopp M. 2012 Sexual selection and magic traits in speciation with gene flow. *Curr. Zool.* **58**, 510–516. (doi:10.1093/czoolo/58.3.510)
16. Smadja CM, Butlin RK. 2011 A framework for comparing processes of speciation in the presence of gene flow. *Mol. Ecol.* **20**, 5123–5140. (doi:10.1111/j.1365-294X.2011.05350.x)
17. Butlin RK *et al.* 2014 Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* **68**, 935–949. (doi:10.1111/evo.12329)
18. Westram AM *et al.* 2018 Clines on the seashore: the genomic architecture underlying rapid divergence in the face of gene flow. *Evol. Lett.* **2**, 297–309. (doi:10.1002/evl3.74)
19. Faria R, Johannesson K, Butlin RK, Westram AM. 2019 Evolving inversions. *Trends Ecol. Evol.* **34**, 239–248. (doi:10.1016/j.tree.2018.12.005)
20. Morales HE, Faria R, Johannesson K, Larsson T, Panova M, Westram AM, Butlin RK. 2019 Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. *Sci. Adv.* **5**, eaav9963. (doi:10.1126/sciadv.aav9963)
21. Johannesson K, Panova M, Kemppainen P, André C, Rolán-Alvarez E, Butlin RK. 2010 Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Phil. Trans. R. Soc. B* **365**, 1735–1747. (doi:10.1098/rstb.2009.0256)
22. Galindo J, Martínez-Fernández M, Rodríguez-Ramilo ST, Rolán-Alvarez E. 2013 The role of local ecology during hybridization at the initial stages of ecological speciation in a marine snail. *J. Evol. Biol.* **26**, 1472–1487. (doi:10.1111/jeb.12152)
23. Kess T, Galindo J, Boulding EG. 2018 Genomic divergence between Spanish *Littorina saxatilis* ecotypes unravels limited admixture and extensive parallelism associated with population history. *Ecol. Evol.* **8**, 8311–8327. (doi:10.1002/ece3.4304)
24. Reid DG. 1996 *Systematics and evolution of Littorina*. London, UK: The Ray Society.
25. Hannaford ECJ. 1985 The breeding migration of *Littorina arcana* Hannaford Ellis, 1978 (Prosobranchia: Littorinidae). *Zool. J. Linn. Soc.* **84**, 91–96. (doi:10.1111/j.1096-3642.1985.tb01718.x)
26. Hannaford ECJ. 1979 Morphology of the oviparous rough winkle, *Littorina arcana* Hannaford Ellis, 1978, with notes on the taxonomy of the *L. saxatilis* species-complex (Prosobranchia: Littorinidae). *J. Conchol.* **30**, 43–56.
27. Warwick T, Knight AJ, Ward RD. 1990 Hybridisation in the *Littorina saxatilis* species complex (Prosobranchia: Mollusca). *Hydrobiologia* **193**, 109–116. (doi:10.1007/BF00028070)
28. Ward RD, Janson K. 1985 A genetic analysis of sympatric subpopulations of the sibling species *Littorina saxatilis* (*olivi*) and *Littorina arcana* Hannaford Ellis. *J. Mollusc. Stud.* **51**, 86–94. (doi:10.1093/oxfordjournals.mollus.a065888)
29. Reid DG, Dyal P, Williams ST. 2012 A global molecular phylogeny of 147 periwinkle species (Gastropoda, Littorininae): phylogeny of Littorininae. *Zool. Scr.* **41**, 125–136. (doi:10.1111/j.1463-6409.2011.00505.x)
30. Nielsen EEG, Bach LA, Kotlicki P. 2006 HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples: program note. *Mol. Ecol. Notes* **6**, 971–973. (doi:10.1111/j.1471-8286.2006.01433.x)
31. Gutenkunst R, Hernandez R, Williamson S, Bustamante C. 2010 Diffusion approximations for demographic inference: DaDi. *Nat. Prec.* (doi:10.1038/npre.2010.4594.1)
32. Rougemont Q, Gagnaire P-A, Perrier C, Genthon C, Besnard A-L, Launey S, Evanno G. 2017 Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol. Ecol.* **26**, 142–162. (doi:10.1111/mec.13664)
33. Johannesson K, Butlin RK, Panova M, Westram AM. 2017 Mechanisms of adaptive divergence and speciation in *Littorina saxatilis*: integrating knowledge from ecology and genetics with new data emerging from genomic studies. In *Population genomics: marine organisms* (eds MF Oleksiak, OP Rajora), pp. 277–301. Cham, Switzerland: Springer.

34. Faria R *et al.* 2019 Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Mol. Ecol.* **28**, 1375–1393. (doi:10.1111/mec.14972)
35. Grahame JW, Wilding CS, Butlin RK. 2006 Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. *Evolution* **60**, 268–278. (doi:10.1111/j.0014-3820.2006.tb01105.x)
36. Roesti M, Hendry AP, Salzburger W, Berner D. 2012 Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. *Mol. Ecol.* **21**, 2852–2862. (doi:10.1111/j.1365-294X.2012.05509.x)
37. Jones FC *et al*. 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61. (doi:10.1038/nature10944)
38. Rundle HD, Nagel L, Boughman JW, Schluter D. 2000 Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306–308. (doi:10.1126/science.287.5451.306)
39. Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. 2018 The genomic landscape at a late stage of stickleback speciation: high genomic divergence interspersed by small localized regions of introgression. *PLoS Genet.* **14**, e1007358. (doi:10.1371/journal.pgen.1007358)
40. Arnegard ME *et al.* 2014 Genetics of ecological divergence during speciation. *Nature* **511**, 307–311. (doi:10.1038/nature13301)
41. Momigliano P, Jokinen H, Fraimout A, Florin A-B, Norkko A, Merilä J. 2017 Extraordinarily rapid speciation in a marine fish. *Proc. Natl Acad. Sci. USA* **114**, 6074–6079. (doi:10.1073/pnas.1615109114)
42. Le Moan A, Gaggiotti O, Henriques R, Martinez P. 2019 Beyond parallel evolution: when several species colonize the same environmental gradient. *bioRxiv*, 662569. (doi:10.1101/662569)
43. Christe C, Stölting KN, Bresadola L, Fussi B, Heinze B, Wegmann D, Lexer C. 2016 Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Mol. Ecol.* **25**, 2482–2498. (doi:10.1111/mec.13587)
44. Egan SP, Ragland GJ, Assour L, Powell THQ, Hood GR, Emrich S, Nosil P, Feder JL. 2015 Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecol. Lett.* **18**, 817–825. (doi:10.1111/ele.12460)
45. Zieliński P, Dudek K, Arntzen JW, Palomar G, Niedzicka M, Fijarczyk A, Liana M, Cogălniceanu D, Babik W. 2019 Differential introgression across newt hybrid zones: evidence from replicated transects. *Mol. Ecol.* **28**, 4811–4824. (doi:10.1111/mec.15251)
46. Gay L, Crochet P-A, Bell DA, Lenormand T. 2008 Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution* **62**, 2789–2806. (doi:10.1111/j.1558-5646.2008.00491.x)
47. Seehausen O *et al.* 2014 Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192. (doi:10.1038/nrg3644)
48. Panova M *et al.* 2016 DNA extraction protocols for whole-genome sequencing in marine organisms. *Methods Mol. Biol.* **1452**, 13–44. (doi:10.1007/978-1-4939-3774-5_2)
49. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
50. Tischler G, Leonard S. 2014 biobambam: Tools for read pair collation based algorithms on BAM files. *Source Code Biol. Med.* **9**, 13. (doi:10.1186/1751-0473-9-13)
51. McKenna A *et al.* 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. (doi:10.1101/gr.107524.110)
52. Van der Auwera GA *et al.* 2013 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.1–11.10.33.
53. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
54. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7. (doi:10.1186/s13742-015-0047-8)
55. Alexander DH, Novembre J, Lange K. 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664. (doi:10.1101/gr.094052.109)
56. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
57. Weir BS, Cockerham CC. 1984 Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.
58. Danecek P *et al.* 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. (doi:10.1093/bioinformatics/btr330)