Phenotypic divergence and genomic architecture at two different stages of speciation in a marine snail

Francesca Raffini [1,2,3, §*], Aurélien De Jode [4,5,6, §*], Kerstin Johannesson [4], Rui Faria [7,8], Zuzanna B. Zagrodzka [1], Anja M. Westram [1,9,10], Juan Galindo [11], Emilio Rolán-Alvarez [11], Roger K. Butlin [1,4]

* These authors contributed equally

[1] Ecology and Evolutionary Biology, School of Biosciences, The University of Sheffield, Sheffield S10 2TN, UK

[2] Department of Life Sciences and Biotechnology, University of Ferrara, 44121, Ferrara, Italy

[3] Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121, Naples, Italy

[4] Department of Marine Sciences-Tjärnö, University of Gothenburg, 45296 Strömstad, Sweden.

[5] Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, 35487, USA

[6] Dauphin Island Sea Lab, Dauphin Island, AL, 36525, USA

[7] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, *InBIO* Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal.

[8] BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal

[9] Faculty of Biosciences and Aquaculture, Nord University, 8049 Bodø, Norway

[10] Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria

[11] Centro de Investigación Mariña, Universidad de Vigo, Departamento de Bioquímica, Genética and Immunología, 36310 Vigo, Spain

[§] Corresponding authors: Francesca Raffini: francesca.raffini3@gmail.com; Aurelien De Jode: aurelien.dejode@laposte.net

Abstract

Speciation typically occurs in a time frame too long to be observed directly. This issue can be overcome by studying pairs of populations at different points in the speciation continuum, ideally within clades so that patterns are not confounded by differences among taxa. Such comparisons are possible in the marine snail *Littorina saxatilis* because it shows repeated occurrence of ecotypes adapted to either crab predation or wave action that differ in age and environmental context. Here, we explored transects spanning hybrid zones between the crab and wave ecotypes to contrast barriers to gene flow in Spain and Sweden, using low coverage whole genome sequencing, shell features, and behavioural traits. The two countries showed parallel divergence but distinct patterns of differentiation between the ecotypes: a continuous cline in Sweden but two highly genetically and phenotypically divergent, and partly spatially-overlapping clusters in Spain. Spanish early-generation hybrids were not observed but a low level of gene flow still seems to occur. In both countries, highly differentiated loci are clustered in genomic regions covered by chromosomal inversions but also occur in collinear regions. Despite being the same species and showing similar levels of phenotypic divergence, the Spanish ecotypes are closer to full reproductive isolation than the Swedish ecotypes. We discuss potential mechanisms contributing to the evolution of these different levels of reproductive isolation, particularly the age of the population, the strength of selection, the spatial context, and the role of assortative mating.

Introduction

The process of speciation involves the build-up of reproductive isolation. The precise meaning of 'reproductive isolation' has recently been debated (Westram *et al.*, 2022 and associated commentaries) but here we us the term in a general sense to include both reduction in interbreeding and reduction in gene exchange. The time-scale over which speciation occurs is highly variable (Coyne & Orr, 2004) but it is typically too long for direct observation. Therefore, inferences about the accumulation of reproductive isolation often depend on comparing contemporary pairs of populations that are at different points on a "speciation continuum" (Seehausen & Wagner, 2014; Stankowski & Ravinet, 2021). This approach can provide many insights, for example into the order in which components of reproductive isolation evolve and the patterns of gene exchange across the genome at different stages of divergence (Feder *et al.*, 2012), although a monotonic progress from weak to strong isolation cannot be assumed (Bolnick *et al.*, 2023; Stankowski & Ravinet, 2021). More studies of population pairs across the speciation continuum are needed, particularly within clades so that comparisons among different levels of reproductive isolation are not confounded with multiple differences among taxa that may not be relevant to the speciation process (Seehausen & Wagner, 2014). Significant progress has been made in this direction, especially in terms of genome-wide patterns of genetic differentiation thanks to the rapidly increasing accessibility of genomic data (e.g., Fang *et al.*, 2020, 2021; Jiggins, 2017; Reid *et al.*, 2021; Riesch *et al.*, 2017) but the number of these investigations in different study systems is

still limited and many questions about the mechanisms leading to speciation remain open, particularly in the absence of obvious physical barriers to gene flow.

The appearance of the first components of reproductive isolation is often relatively well understood, but explaining the later accumulation of additional components of isolation and the final cessation of all gene flow remains problematic (Butlin *et al.*, 2008; Butlin & Smadja, 2018; Kulmuni *et al.*, 2020). Many factors have been suggested that might lead to stronger reproductive isolation. Time is needed for the accumulation of incompatibility (e.g., Guerrero *et al.*, 2017), and possibly also for response to divergent selection if this is mutation limited (Barrett & Schluter, 2008), and so older population pairs might be more strongly isolated, as is often observed (e.g., Coyne & Orr, 1997). Stronger divergent selection, for example in more distinct habitats, should also lead to stronger isolation (Funk *et al.*, 2006). Some spatial arrangements of populations might be more conducive to the evolution of reproductive isolation than others, because gene flow opposes divergence (Coyne & Orr, 2004) but also because contact provides the opportunity for reinforcement (Servedio & Noor, 2003; Yukilevich, 2021). Similarly, the history of separation and contact between populations is likely to be important, because of the opportunity for gene flow and reinforcement but also because cycles of population expansion and contraction can act to bring components of reproductive isolation together (Butlin & Smadja, 2018; Hewitt, 1989). Opportunities for the evolution of assortative mating, either behaviour or due to habitat association, might vary with the reproductive biology of the taxa involved, and can lead to strong reproductive isolation particularly when it is aligned to divergent selection (Kopp *et al.*, 2018). One-allele barrier effects (Butlin *et al.*, 2021; Felsenstein, 1981), pleiotropy and multiple-effect traits (Servedio *et al.*, 2011; Smadja & Butlin, 2011) might promote the evolution of reproductive isolation. Finally, genomic architecture (numbers and effect sizes of barrier loci, their genomic distribution and patterns of recombination, including the effects of structural variants) may be important because the coupling of individual barrier loci and of barrier effects can be opposed by recombination (Butlin & Smadja, 2018; Felsenstein, 1981).

The intertidal gastropod, *Littorina saxatilis* and its close relatives, is a model system in which many of these issues can be addressed (Johannesson, 2016; Johannesson *et al.*, 2010, 2017; Rolán-Alvarez *et al.*, 2015). *L. saxatilis* is widespread and abundant on North Atlantic shores where its ovoviviparous reproduction and resulting short lifetime dispersal have allowed it to adapt to many different habitat types (Reid, 1996). Ecotypes adapted to sheltered habitats with a high risk of crab predation (the 'Crab' ecotype) or more exposed habitats with strong wave action but low predation risk (the 'Wave' ecotype) occur in close proximity on many shores and often meet in contact zones. They differ in a suite of adaptive traits including size, shell shape and thickness, and behaviour. Demographic models suggest that they evolved in parallel (Butlin *et al.*, 2014) although this is likely to have involved repeated use of at least some adaptive variants (Morales *et al.*, 2019), a proportion of which are found in putative chromosomal inversions (Faria *et al.*, 2019; Koch *et al.*, 2021, 2022), at least in Sweden. Background levels of genetic divergence between the Crab and Wave ecotypes and patterns of genomic differentiation determined using pooled sequencing data suggested that the strength of the barrier to gene flow between ecotypes varies widely among locations in Europe (Morales *et al.*, 2019).

In this study, we compared and contrasted the patterns of divergence between the Crab and Wave ecotypes in two regions in which the ecotypes have earlier been investigated: the Swedish west coast, colonised since the last glaciation, and the Galician coast in Spain, an older population that survived the Pleistocene glaciations *in situ* (Bosso *et al.*, 2022; Doellman *et al.*, 2011; Panova *et al.*, 2011). In Sweden, the tidal range is small, the Crab-Wave axis is parallel to the shore line (horizontal zonation), and the two ecotypes hybridize in narrow contact zones where sheltered boulder fields abut rocky headlands. There is a genome-wide barrier to gene exchange, but it is weak: background F_{ST} is about 0.04, clinal changes in SNP frequency are widespread in the genome but fixed differences are rare, loci putatively contributing to barrier effects occur both within and outside polymorphic inversions (Koch *et al.*, 2022; Westram *et al.*, 2018, 2021). The barrier appears to be due primarily to local adaptation, without intrinsic incompatibilities, but with some contribution due to size-assortative mating (Janson, 1983; Johannesson *et al.*, 2010, 2020; Le Pennec *et al.*, 2017; Perini *et al.*, 2020). In Spain, the tidal range is much greater and the two ecotypes are distributed on a perpendicular up-down shore axis (vertical zonation) with the Crab ecotype primarily in the barnacle belt in the high shore, where predation is most intense, and the Wave ecotype among blue mussels in the low shore, where wave action is strongest. The Spanish ecotypes overlap in a wide contact zone in the mid shore, characterized by a mosaic distribution of barnacle and mussel patches (Johannesson *et al.*, 1993, 1995; Morales *et al.*, 2019; Rolán-Alvarez *et al.*, 1997, 1999). There are some indications that the barrier to gene flow between ecotypes is stronger in Spain than in Sweden: background F_{ST} is higher, around 0.1 (Butlin *et al.*, 2014; Morales *et al.*, 2019; Westram *et al.*, 2021), and only a few intermediate genotypes were observed in putatively hybrid samples using reduced representation data (RAD-seq, Kess *et al.*, 2018). Yet, an investigation of the pattern of differentiation in Spain with a combination of genome-wide data and detailed spatial coverage of the contact zone has been lacking.

Here, we describe the phenotypic and genomic pattern of differentiation between Crab and Wave ecotypes in Spain using low-coverage whole-genome sequence data, shell features and behavioural traits for snails from a dense transect across the contact zone. We make a direct comparison with a transect in Sweden as well as with published analyses based on pooled or capture sequencing. We report a very different pattern in the Spanish contact zone compared to Sweden: in Spain, ecotypes partly overlap in space with evidence for only limited, almost unidirectional introgression, while in Sweden there is a gradual phenotypic as well as genetic transition from one ecotype to the other. We then consider the possible reasons for the very different positions of Spanish and Swedish snails on the speciation continuum for two populations of the same species.

Materials and Methods

Sampling

Snails were sampled in two areas of the core distribution of *L. saxatilis*: Spain and Sweden. In Sweden, snails were sampled in a single transect along the shore from a boulder field ('Crab' habitat) to a rocky headland ('Wave' habitat) on the island of Ramsö (58°49′27.8″N, 11°03′45.3″E; a re-sampling of

transect CZA_right from Westram *et al.*, 2021). Note that the tidal amplitude is only 35 cm in Sweden but high and low water level also vary with air pressure up to a maximum amplitude of 1.5 m: in all parts of the transect, snails were collected from positions scattered throughout their vertical distribution (~1m). Seven hundred snails were sampled in June 2015 by hand, haphazardly, without reference to phenotype but aiming to avoid juveniles. The position of each snail was recorded in three dimensions using a total station (Trimble M3). For spatial analysis, we placed each snail on a 'least cost path', as described in Westram *et al.* (2021), and distances were transformed to start from zero at the Crab ecotype end of the transect.

In Spain, snails were collected from Centinela on the west coast of Galicia (hereafter ER_EA; N 42° 4' 38.06", W 8° 53' 47.47") in spring (March) and autumn (September) of 2017. Each sample consisted of approximately 600 snails collected in the same way as in Sweden from two transects perpendicular to the shore, and about 5m wide and separated by 0-5 m, stretching from the upper limit of the *L. saxatilis* distribution in the splash zone to its lower limit close to low water of spring tides. Sampling was more dense in the lower part of the shore where hybridization between the two previously described ecotypes was expected (Galindo *et al.*, 2013). For spatial analysis, we used the position of each snail on a single shore-position axis running through the centre of each transect. Distances were transformed such that each transect started from zero and ended at roughly low water level as indicated by the lowest collected snails. To include habitat information, the presence or absence of barnacles (*Chthamalus* spp., typical of the mid to upper shore), mussels (*Mytilus galloprovincialis,* from mid to lower shore), and goose barnacles (*Pollicipes pollicipes*, lower shore) was recorded within 50mm of each snail position and for approximately uniformly distributed positions across the area containing the two sampling transects. Snails were collected over a vertical range of approximately 3m and shore height was recorded relative to the position of the lowest individual sampled.

Snails were stored in individual tubes, moistened with seawater and kept at 4°C until phenotyping was complete. Then, the head and foot of each snail was dissected and preserved in 99% ethanol.

Sequencing and read processing

Seventy-three adult snails from Spain in spring, 114 from Spain in autumn and 96 from Sweden were selected to cover the full sampling range within each country. DNA was extracted from foot tissue using a modified CTAB protocol (Panova *et al.*, 2016). In-house, high-throughput genomic DNA library preparation and whole genome sequencing (Illumina HiSeq4000, 150bp, eight lanes, paired-end) were performed by The Oxford Genomics Centre with a target coverage of 3x based on the estimated genome size of 1.35Gb (Westram *et al.*, 2018).

Raw reads were trimmed with Trimmomatic v. 0.38 (Bolger *et al.*, 2014), retaining reads with a minimum length of 70bp after filtering, and mapped to the *L. saxatilis* draft reference genome (We-

stram *et al.*, 2018) using bwa mem v. 0.7.17 (Li, 2013) and default settings. Positions with base or map quality lower than 20 were discarded using Samtools v. 1.7 (Li, 2011; Li *et al.*, 2009). PCR duplicates and overlap between paired-end reads were removed with Picard v. 2.0.1 (http://broadinstitute.github.io/picard/) and bamUtil (Jun *et al.*, 2015). As specimens were sequenced in paired-ends in eight lanes, resulting in 16 outputs for each snail, the files belonging to the same individual were sorted and merged with Samtools.

To explore patterns of diversity within Spain and Sweden, variants were called separately in each country using samtools mpileup and bcftools call v. 1.11 (Li, 2011) including only the longest contigs covering 90% of the reference genome. Allelic read counts for each SNP rather than genotype calls were retained due to the low-coverage. The resulting vcf files were filtered to retain only biallelic SNPs, positions where at least 50% individuals had a read depth between one and 18 irrespective of the reference or alternative allele, and a minor allele frequency higher than 0.05 using vcftools v. 0.1.14 (Danecek *et al.*, 2011) and vcffilterjs (Lindenbaum & Redon, 2018). Additionally, we retained only positions on the *L. saxatilis* linkage map (Westram *et al.*, 2018), i.e. within 1,000 bp of a SNP that could be positioned on this map.

Two types of dataset were generated for each country: unthinned and thinned. The first one included all the SNPs resulting from the processing described above without any further filtering. The second one was obtained by retaining one random SNP in each genomic window of 1,000 bp to reduce the impact of linkage disequilibrium and avoid overweighting regions with high SNP density. To achieve this aim, the reference genome was sliced into bins of 1,000 bp and one SNP was randomly picked in each window using the R v. 4.0.3 (R Core Team, 2020) package GenomicAlignments v. 1.26.0 (Lawrence *et al.*, 2013), vcftools and custom scripts. This random SNP subsampling was repeated three times in each country to create a total of six random SNP subsets from the unthinned datasets.

To investigate overall divergence between countries, variants were called jointly in Sweden and Spain and a thinned dataset was generated using the procedures described above.

In all datasets (within and between countries, thinned and unthinned), the reference and alternate allele read depth was extracted from the vcf files and one random read per position and individual was subsampled using vcftools and custom scripts. As for the random SNP subsets, the random read subsampling was repeated three times to create replicates for each SNP dataset and subset. These ‘single read’ data sets formed the basis of population genomic analysis, thereby avoiding the issues associated with calling genotypes from low-coverage data (Nielsen *et al.*, 2011).

Genomic analyses

Overall patterns of divergence within countries were explored in the thinned datasets through PCA and DAPC in the R packages adegenet (Jombart, 2008; Jombart & Ahmed, 2011) and factoextra v. 1.0.7 (https://github.com/kassambara/factoextra) treating individuals as haploid. If more than one ge-

netic cluster was identified, the within-cluster PC1 scores among subsets were tested for normality using the Shapiro test (Shapiro & Wilk, 1965). We compared cluster (if any) assignment of each individual between subsampled data sets and computed the correlation of the within-cluster PC1 scores using the Pearson or Spearman's coefficients, according to the data distribution, using the R package GGally v. 2.1.2 (https://ggobi.github.io/ggally). Loci with an allele frequency difference between clusters (if any) > 0.5 were used to compute a Hybrid Index for each individual (mean over genotypes expressed as 1 for the allele more common in the 'Crab' environment and 0 for the allele more common in the 'Wave' environment). Individual ancestries were estimated through a maximum likelihood approach and cross validation procedures using ADMIXTURE v. 1.3.0 (Alexander *et al.*, 2009), PLINK v. 1.90b6.5 (Purcell *et al.*, 2007) and custom scripts.

While the analyses described in the previous paragraph took advantage of the thinned datasets, the unthinned ones were used to compute global and per-locus F_{ST} (Weir & Cockerham, 1984) between genetic clusters (if any) or transect ends (defined as positions before 30 and after 80 m in Sweden), without imputation for missing values, using the R package hierfstat v. 0.5-7 (Goudet, 2005). At a larger geographic scale, divergence between countries was explored using the joint thinned datasets through PCA and DAPC as described above.

Chromosomal inversions are known to contribute to ecotype differentiation in the Swedish population (Faria *et al.*, 2019; Koch *et al.*, 2022; Westram *et al.*, 2021). To investigate patterns of differentiation along the genome and identify chromosomal inversions in Spain, the unthinned datasets and two approaches were used. First, per locus F_{ST} values between genetic clusters were plotted along each LG to produce Manhattan plots using custom scripts in R. Genomic inversions are expected to generate blocks of higher differentiation compared to non-inverted regions (Mérot, 2020). Second, we computed PCA by map position using the same procedure illustrated above and custom scripts to extract the SNPs falling on each linkage map position. PC1 scores per individual per map position were extracted and transformed by taking the inverse of the score when needed so that individuals belonging to the same genetic group always clustered on the same side of the PC1 axis. PC1 scores were then plotted along each LG and individuals were coloured according to their genetic cluster. Here, single chromosomal inversions tend to show three distinct groups corresponding to each homokaryotype (upper and lower groups) and the heterokaryotypes (middle group), while overlapping inversions display a characteristic triangular pattern, with apical cluster comprising homokaryotype individuals (Mérot, 2020). An inversion was considered to be present when both the PCA and Manhattan plots agreed on the inversion signal.

To investigate the effect of chromosomal inversions on the observed patterns of divergence within countries, the genetic analyses described above were repeated excluding SNPs falling within known inversions (Faria *et al.*, 2019; Koch *et al.*, 2022; Westram *et al.*, 2021) from the original Swedish and Spanish thinned datasets using vcftools.

Consistency among the randomly-sampled subsets was tested using Kendall's coefficient of concordance, W, computed using only the complete cases of the rankings and in the R package DescTools v. 0.99.48 (Kendall, 1948; Signorell, 2023).

Phenotypic analyses

Parallel morphological differentiation of ecotypes in Sweden and Spain has been described previously (Butlin *et al.*, 2014; Johannesson *et al.*, 2010). Continuous variation in multiple phenotypic traits across the Swedish transect has also previously been described (Koch *et al.*, 2021, 2022; Larsson *et al.*, 2020). Here, we focused on phenotypic divergence in the Spanish transect, where the following traits were recorded in each snail: maturity (adult, juvenile), sex, wet weight, shell thickness (using NeoteckDTI Digital Dial Indicator Probe, 0.001mm resolution at the widest point of aperture and average over three measures), shell ridging, shell banding, and boldness. To measure boldness, snails were disturbed to induce retraction and time until they emerged again (out time) and until they got back on the foot (crawl time) were recorded. This test was repeated three times for each snail and the average score for both out and crawl times between the three trials was used for subsequent analyses. Scores were attributed according to the out or crawl times in minutes, using the following categories for the two measures, respectively: [0,<1], score=1; [0-1,1-5]=2; [1-5,5-10]=3; [5-10,10-15]=4; [10-15, >15]=5. Each trial was stopped after 15 minutes independently of the snail's response. Details of this method deviate from previous studies (Koch *et al.*, 2021) but result in a similar distinction between 'bold' (low score) and 'wary' (high score) behaviours.

Shell length and growth parameters were obtained from standardized pictures of each snail. A scale was included for reference (Larsson *et al.*, 2020). These pictures were then analysed using *ShellShaper* (Larsson *et al.*, 2020) and the relative thickness (thickness/a0), aperture position parameters (r0_scaled=r0/shell_length, z0_scaled=z0/shell_length), aperture shape (extension factor c0/a0), aperture size (a0_scaled=a0/shell_length), height growth (log_gh), and width growth (log_gw) were computed as in previous studies (Koch *et al.*, 2021; Larsson *et al.*, 2020).

For each phenotypic trait, we tested differences between any genetic groups identified using the clustering analysis described above. Additionally, multivariate patterns were investigated through a PCA using the *prcomp* function from the R package *stats* v. 4.3.0 and the following phenotypes: a0_scaled, r0_scaled, z0_scaled, log_gh, log_gw, relative thickness, weight, thickness, and shell length.

To compare phenotypes between Sweden and Spain, we retrieved phenotypic data of the CZA transect from (Koch *et al.*, 2022), that corresponded to the same transect of this study but sampled at a different time. That dataset included the following variables: shell length, wet weight, aperture size, aperture position, aperture shape, height growth, width growth and bold score (crawl score). All these variables (except bold score) were used in a PCA including all individuals from both CZA

and Spain. In CZA, pure ecotypes were defined using the shore position, where pure Crab individuals are found in boulder habitat at least 15 meters away from the boulder-rock transition and pure Wave snails are found in the rocky habitats at least 40 meters away from the boulder-rock transition (Koch *et al.*, 2022). Then average relative differences between ecotypes were computed for each variable both in Sweden and Spain.

Results

Distinct patterns of genetic divergence in Sweden and Spain

We explored genetic divergence along transects spanning the Crab-hybrid-Wave axis in Sweden (boulder field to rocky headland) and Spain (high shore to low shore) using low coverage whole genome resequencing data in 283 snails. The Spanish transects did not show seasonal patterns (Fig. S1 in Supplementary Information), thus samples from spring and autumn were merged in the subsequent steps. A total of 311,549 (unthinned datasets) and 21,250 (thinned datasets) SNPs were obtained in Sweden, while the Spanish unthinned and thinned data included 339,614 and 21,212 SNPs, respectively. The joint thinned datasets, used to investigate overall divergence between countries, included 7,577 SNPs.

Sweden and Spain showed distinct patterns of genetic differentiation. In Sweden, Crab and Wave ends of the transect were distinguished by PC1 (3% of variation), but with a continuous range of intermediates distributed clinally along the transect so that cluster analysis identified only a single genetic group (Fig. 1, Fig. S2). This pattern was consistent among the random SNP and read subsets (Table S1) and in line with previous studies (Westram *et al.*, 2018, 2021).

In contrast to the Swedish population, the Spanish snails did not show a clinal pattern along the transect but instead formed two strongly genetically distinct groups: a more genetically variable one spanning almost all of the transect (corresponding to the Crab ecotype), while the other one was more homogeneous and mostly localized in the lower part of the shore (Wave ecotype). No intermediate genotype was observed between these two groups (Fig. 1, S2-S3). Some Spanish Crab samples showed a hybrid index value close to 0.6, indicating admixture and significant contributions of alleles typical of the Wave cluster in their genome, while individuals of the Wave cluster had little or no evidence for Crab ancestry. None of the analysed Spanish snails had a Crab ancestry proportion between 0.1 and 0.6, or a hybrid index between 0.2 and 0.55, whereas such individuals were common in the Swedish transect (Fig. 1, Table S2). These results were consistent among the random SNP and read replicates (Fig. S3, Table S1), including the assigned group membership, i.e., the same individuals were consistently classified as belonging to the Crab or Wave group. Within-cluster PC1 scores were highly correlated among the random SNP and read subsets (range: 0.7999 - 0.9965, Fig. S4). The two genetic clusters in Spain did not exhibit a clear association with habitat features such as shore height and presence of barnacles or mussels. Remarkably, some of the snails located in the lowest positions along the transect belonged to the Crab cluster (Fig.1, Fig. S5).

At a larger spatial scale, the joint analyses consistently identified three genetic groups: Sweden and the two Spanish clusters. Most of the differentiation was explained by geographic separation between the two regions, followed by the crab-wave axis in Spain. Interestingly, the divergence between the two Spanish groups was higher than variation in the whole Swedish transect (Fig. S6-S7).

Differentiation along the genome and chromosomal inversions

The genome-wide average differentiation between the two genetic clusters was higher in Spain compared to the ends of transect in Sweden (global F_{ST} values of 0.16 and 0.09 averaged among subsets, respectively, Table S3). The average F_{ST} by LG was higher in Sweden than in Spain in LG8 and LG15; similar between the two countries for LG7, LG11 and LG16; and higher in Spain than in Sweden in all the other LGs. (Table S3). Some regions displayed higher F_{ST} in Sweden compared to Spain, while other regions presented high differentiation in both countries, both in the form of islands or narrow peaks of high differentiation relative to adjacent regions (Fig. 2, S8-S10). A total of 53 contigs exhibited high differentiation in both countries (average F_{ST} per contig > 0.3 in both countries, Fig. S8) accounting for approximately 43% (53/122) and 11% (53/476) of contigs with F_{ST} values exceeding 0.3 in Sweden and Spain, respectively. Highly differentiated genomic regions in both countries were located on LG1, LG2, LG3, LG5, LG6, LG8, LG9, LG12, LG14, LG17, with LG6 and LG14 containing 26 and 10 of these contigs, respectively (Table S4). Forty of the 53 highly-differentiated contigs were located in known inversions that were polymorphic between ecotypes in Sweden (Faria *et al.*, 2019; Hearn *et al.*, 2022; Westram *et al.*, 2021) on LG6, LG9, LG12 and LG14.

To investigate the presence of inversions along the genome, we used Manhattan plots and PCA by map position plots along each of the 17 LGs. Multiple chromosomal inversions were identified in both countries in the same positions across the genome. Inversion patterns were congruent among methods (F_{ST} and PCA) and generally more pronounced in Spain than in Sweden with Manhattan plots showing blocks of high F_{ST} being more differentiated from the background and PCA plots showing more distinct clusters (Fig. 2, S9-S12). Overall, ten genomic regions in eight LGs exhibited a clear inversion pattern in the Spanish dataset. In LG1, LG2, LG4, LG7, LG9 and LG17 inversions were observed and corresponded to the Swedish inversions LGC1.2, LG2.1, LG4.1, LG7.2, LG9.1, and LGC17.1, respectively (Westram *et al.*, 2021). In LG6 and LG12, triangular PCA patterns (Fig. S13-S15) at several genomic positions along the LG suggested the presence of several potentially overlapping inversions, as previously described in Sweden (Faria *et al.*, 2019; Hearn *et al.*, 2022; Westram *et al.*, 2021). LG10, LG11 and LG14 showed ambiguous patterns of high differentiation in some regions but no obvious characteristics of inversions (Fig. 2, S9-S12). In LG11, a block of high F_{ST} was present in the region of the LGC11.1 inversion previously identified in Sweden, but it did not overlap entirely with the inversion's limits and the PCA plots did not show a clear characteristic pattern. Characteristic but less pronounced triangular PCA patterns of overlapping inversions were present in LG10 (Fig. S14) and LG14 (Fig. S16), similarly to Sweden where two consecutive inversions were previously observed in LG10 and three (two overlapping) inversions were identified in LG14.

The inversions LG1.1 and LG7.1, previously identified in Sweden, were not observed in the Spanish populations. There was no clear evidence of any previously unknown inversion in Spain, though some regions displayed blocks of high differentiation that could indicate the presence of an inversion in LG5 from 15 cM to 46 cM and in LG8 from 30 cM to 35 cM (Fig. S10).

To investigate the contribution of chromosomal inversions to genetic divergence along transects, the genetic analyses described above were repeated using the thinned datasets and removing the SNPs in known inversions. In Sweden, clinal patterns remained consistent after removing the positions within known inversions (Fig. S17-S18, Table S2). A similar trend was observed in Spain, where no substantial differences were detected between the patterns with and without inversions: both showed two distinct genetic groups without intermediate genotypes (Fig.1, S2-S3, S17-S18). The Wave group was less homogeneous, and the percentage of variance attributed to the divergence between the two genetic clusters in PC1 was slightly lower (6.01%) compared to the analyses with inversions (6.93%, Fig. 1, S17). As in the analyses with inversions, results were consistent among the random subsets, which showed the same individual assignment to the Crab or Wave group and a high correlation of the within-group PC1 scores (Fig. S19, Table S1). After removing inversions, genome-wide average F_{ST} decreased slightly in Spain while it increased in Sweden (Table S3). The variability of F_{ST} among LGs (standard deviation of average F_{ST} among LG) was higher in Spain than in Sweden. Although it decreased in both countries with removal of the inversions (from 0.0421 to 0.0335 in Spain and from 0.0298 to 0.0161 in Sweden) it remained high in Spain, indicating heterogeneity in the barrier to gene flow over and above the effect of inversions.

Phenotypic divergence

Analyses of individual phenotypes in 185 Spanish snails suggested significant differences between the Crab and Wave groups. Some traits could not be measured in all snails (see sample sizes in Fig. S20). Overall, the Crab snails were bigger, heavier, and presented thicker shells with lower height and width growth, a higher aperture height, a smaller aperture size and higher size-independent relative thickness compared to the Wave snails (Fig. S20). Most of the Crab shells (93%) displayed ridges while they were present only in a single Wave snail (Fig. S20). In Sweden, ridging is found in the crab ecotype but is less pronounced than in Spain and is absent in the wave ecotype (personal observations). Banded shells (a phenotype never observed in Sweden) were observed in most snails except eleven individuals belonging to the Wave group. Bands were mostly black in the Crab group, while they were both black and brown in the Wave group (Fig. S20). Behavioural tests indicated that the Crab snails were more wary than the Wave group (Fig. S20). In fact, more than half of Crab snails took more than ten minutes to come out of their shell (median out boldness score of 4.33) whereas less than five minutes (median out boldness score of 1.33) were needed for most of the Wave snails (Fig. S20). Using data from Koch *et al.* (2022) for the CZA transect, relative differences between average individual phenotypic traits per ecotype varied between Spain and Sweden. While relative differences between average wet weight and shell length were similar between ecotypes in Sweden and

Spain, relative differences between average aperture size, aperture position, aperture shape, width growth and height growth were higher in Sweden compared to Spain (Fig. S21). Wherever differences between ecotypes were found in both regions, they were in the same direction.

The multivariate analysis indicated a more continuous variation in phenotypic than genomic data in the Spanish samples, and partial overlap between the Crab and Wave group (Fig. 3). The first axis of the PCA, accounting for around 51% of the observed variation, was highly correlated with shell length parameters while the second axis, describing around 19% of variance, was highly correlated with the aperture position (r0 and z0, Fig. 3a). Overall, the two genetic clusters were phenotypically distinct. However, a few individuals presented discordance between their genetic background and phenotype, i.e., phenotypic features that were typical of the other genetic cluster. The Crab snails showed a down-shore cline, as in the genomic analyses, with some individuals overlapping spatially and phenotypically with the Wave cluster: individuals located closer to the lower shore exhibited more Wave-like phenotypic characteristics (Fig. 3b). The joint phenotypic PCA, including the same subset of traits from both our Spanish snails and the Swedish CZA dataset from Koch *et al.* (2022), also showed continuous phenotypic variation between individuals (Fig. S22). The first axis accounted for around 74% of the total variation and was highly correlated with aperture size, aperture shape, aperture position, height growth and width growth. Despite this continuous variation, individuals clustered according to ecotypes and countries; distances between ecotype centroids were higher in Sweden than in Spain. The Swedish crab snails were the most distinct cluster and exhibited the highest levels of divergence from the Spanish wave samples.

Discussion

Contrasting populations at different stages of speciation can contribute to clarifying the processes underlying the emergence and accumulation of reproductive isolation, particularly in the absence of obvious barriers to gene flow (Hendry, 2009; Stankowski & Ravinet, 2021). The marine snail *Littorina saxatilis* is a renowned study system of repeated differentiation between Crab and Wave ecotypes, yet differences in the strength of reproductive isolation across its distribution has been largely overlooked. Few studies have directly compared ecotype divergence among different regions using the same methods (Butlin *et al.*, 2014; Morales *et al.*, 2019). Using low coverage whole genome sequencing and analyses of shell features and behavioural traits in two continuous transects in Spain and Sweden spanning the whole Crab-Wave axis, we confirmed parallel phenotypic and genetic divergence in these shores and the role of chromosomal inversions. However, we report distinctly different distributions of genetic diversity in space, with clearly separate genetic groups in Spain, lacking intermediates and overlapping in the wave-exposed habitat, in strong contrast to the continuous, unimodal cline observed in Sweden. We conclude that the populations of these two countries represent different stages along the 'speciation continuum' with the Spanish populations being closer to complete reproductive isolation.

The Spanish contact zone

Our analyses showed contrasting patterns of divergence along transects in Sweden and Spain. Genome-wide data from dense spatial sampling in the Spanish contact zone revealed that the two ecotypes are genetically distinct, with no early-generation hybrids and wide spatial overlap but indications of low levels of gene flow in one direction. A previous analysis of parental and 'intermediate' samples also suggested such as subdivision (Kess *et al.*, 2018). In Spain, phenotypic variation was strongly associated with the genetically-defined ecotypes but continuous overall, the lack of separation being largely due to variation along the shore height gradient within the Crab ecotype: low-shore snails of the genetically-defined Crab group were much more similar to Wave snails in size and shape than were high-shore Crab snails. However, one phenotypic trait (ridging) was almost perfectly associated with the Crab ecotype over the whole transect. This and other shell characteristics (colour and banding) corresponded to the previously described RB and SU ecotypes (Crab and Wave, respectively; Johannesson *et al.*, 1993, 1995; Rolán-Alvarez *et al.*, 1997, 1999).

The genetic pattern in Spain contrasts with continuous allele and inversion frequency clines between ecotypes in Sweden (Westram *et al.*, 2018, 2021), reflected in our analyses by a continuum of admixture proportions or hybrid index scores in our Swedish transect and a lower overall differentiation between ecotypes, when sampled away from the contact zone. Phenotypic differentiation was more contrasted between Spain and Sweden. While shell ornaments such as ridges and shell banding pattern presented higher differentiation in Spain, size and shell shape traits presented higher differentiation in Sweden, as previously observed (Butlin *et al.*, 2014). The different shore configurations in Spain and Sweden could explain the lower shell shape differentiation observed in Spain. While Swedish crab habitats (boulder fields) are always very protected from wave action, in Spain the crab habitat (upper part of the shore) can be battered by waves, particularly during storms. Therefore, maintaining shape traits allowing snails to cope with wave action in the upper part of Spanish shores would still be advantageous and could explain the pattern observed in our study.

The Spanish population is characterized by two clearly distinct genetic clusters without intermediates, a pattern that is indicative of strong reproductive isolation between these two groups. Remarkably, the distributions of the two genetic clusters on the shore do not abut, as might be expected from adaptation to a simple habitat transition. In fact, the Spanish snails in the more heterogeneous "Crab" group were not limited to the upper shore but overlapped with the Wave ecotype, to the lowest levels sampled, and tended to show more "Wave"-like traits towards the lower end of the transect. These patterns are not fully consistent with previous findings at this and nearby localities, where sampling focused on groups representing phenotypic categories (e.g., Crab-like, Wave-like, and intermediate shells) or environmental areas (upper and lower shore) that might have mixed genetic groups and so not fully represented the intra-population diversity (Butlin *et al.*, 2014; Galindo *et al.*, 2013; Johannesson *et al.*, 1993, 1995; Kess *et al.*, 2018; Morales *et al.*, 2019; Rolán-Alvarez *et al.*, 1997, 1999). In Sweden (CZA), ecotype distributions do abut, separated by narrow hybrid zones, but a consistent displacement of cline centres from the habitat transition into the Wave habitat (Johannesson *et al.*, 2020; Westram *et al.*, 2018, 2021), partly mirrors the Spanish situation and suggests that factors

other than divergent selection along the Crab-Wave axis are involved in ecotype divergence in these shores. Our results highlight the power of analysing continuous spatial patterns across transects spanning fully replicated hybrid zones, which proved to be more informative than simple population comparisons in distinguishing processes underlying divergence (Westram *et al.*, 2021). However, greater resolution is still needed. A possible explanation for range overlap in Spain is a habitat mosaic on the mid- to low-shore, particularly if this is associated with local habitat choice (as suggested previously, e.g., Cruz *et al.*, 2004). Local associations with habitat variables were not apparent in our data but may be revealed by larger samples and other environmental descriptors. At the same time, wider sampling is clearly needed to establish whether the pattern at Centinela is repeated on other Iberian shores.

No individual with a genotype suggestive of F1 or early-generation hybrids was observed in Spain, suggesting a strong barrier to gene flow. However, low levels of gene flow, predominantly from Wave to Crab, could explain the increasing genetic and phenotypic similarity between ecotypes towards the lower shore, possibly due to between-ecotypes differences in longer-distance movement, habitat choice, or displacement by waves. The presence of Wave ancestry in the Crab snails could indicate local adaptation from standing genetic variation, and/or natural selection favouring introgressed alleles. These findings are in line with the observation of very few genetically admixed individuals in a group of snails selected for their intermediate morphology (Kess *et al.*, 2018). Interestingly, we observed discrepancies between genetic assignment and phenotypic features in a few snails without clear signatures of hybridization, backcrossing or introgression, that need further investigation.

Despite marked differences in the genomic architecture of ecotype divergence between Spain and Sweden, differentiation along the genome showed some degree of parallelism between the two countries, indicating that at least some parts of the genome were repeatedly implicated in the recurrent evolution of these ecotypes. Most of these common highly differentiated regions between Crab and Wave were located in LG6, LG12 and LG14, linkage groups that have been reported to show this pattern across the entire European distribution of *L. saxatilis* (Koch *et al.*, 2022; Morales *et al.*, 2019). Our results also suggested the presence of several chromosomal inversions in Spain in the same genomic regions as in Sweden, including LG6, LG12 and LG14. These patterns will be further corroborated by future studies as our ability to determine the precise extents of inversions was limited by the use of low coverage genomic data and the fragmented reference genome. Interestingly, inversions alone did not fully explain the two distinct genetic clusters observed in Spain, as is also true in Sweden (Koch *et al.*, 2022; Westram *et al.*, 2021); in fact, the two groups in Spain remained clearly discrete also after removing inversions, suggesting that gene flow between the Spanish ecotypes is highly restricted throughout the genome.

Distinct strength of reproductive isolation in Spain and Sweden

The Swedish and Spanish populations of *L. saxatilis* have experienced similar divergent selective pressures, as suggested by their phenotypic parallelism, and share similar genetic features such as

common inversions, yet reproductive isolation appears to be stronger in Spain, as shown by the higher levels of genomic divergence between ecotypes and the lack of early-generation hybrids. These results indicate that the Spanish ecotypes represent a more advanced stage of speciation than those in Sweden.

The evolution of these barriers to gene flow between ecotypes of different strengths might result from several and not mutually-exclusive mechanisms. First, the Spanish populations are older, having survived the Pleistocene glaciations *in situ*, whereas the Swedish populations are the result of post-glacial colonisation (Bosso *et al.*, 2022; Doellman *et al.*, 2011; Panova *et al.*, 2011). The spatial arrangement of habitats means that a much greater proportion of each ecotype experiences contact with the other one in Spain than in Sweden (because of the vertical rather than horizontal zonation), resulting in greater potential for gene flow but also more opportunity for reinforcement of prezygotic barriers (Servedio & Noor, 2003; Yukilevich, 2021). Periods of allopatry could have contributed to the build-up of ecotype differentiation but there is no evidence of past complete local isolation in either country (Butlin *et al.*, 2014; Galindo *et al.*, 2009; Kess *et al.*, 2018; Rolàn-Alvarez *et al.*, 2004). Divergent selection may be stronger in Spain because of the coincidence of the predation-exposure gradient with the up-down shore gradient (Morales *et al.*, 2019). Size-based assortative mating operates similarly in both regions, apparently being an ancestral feature (Ng *et al.* 2018). Size is a multiple-effect trait (Smadja & Butlin, 2011) and assortment in this species is by matching (Kopp *et al.*, 2018), both features conducive to the evolution of behavioural isolation (Ng *et al.*, 2019). The contribution of non-random mating (assortment and sexual selection) to reproductive isolation in Sweden nevertheless appears to be small because of the continuum of phenotypes in the hybrid zone (Perini *et al.*, 2020). With strongly divergent sizes, stronger assortment would be expected, and is observed in Spain (Johannesson *et al.*, 1995). This suggests that strong isolation in Spain may depend on an interaction between assortative mating and other factors that contribute to size differentiation such as stronger divergent selection or coupling with habitat choice and environmental patchiness. In Spain, habitat selection is strong in the contact zone (Cruz *et al.*, 2004; Erlandsson *et al.*, 1998, 1999; Johannesson *et al.*, 1995; Kostylev *et al.*, 1998) while in Sweden there are no signs of crab-wave habitat choice, which would be difficult to achieve as the contact zone is not a mosaic but a gradual environmental transition (Janson, 1983; Westram *et al.*, 2018), although there is evidence for some shore-level choice (Hearn, 2023). Finally, both single nucleotide and structural polymorphisms appeared to be shared between regions (Table S4, Fig. S8; Kess *et al.*, 2021; Morales *et al.*, 2019), and structural variants might enhance barriers to gene flow (Faria *et al.*, 2019) but their contributions to reproductive isolation might differ between geographic regions.

The speciation process can be seen as a continuum of levels of reproductive isolation (Stankowski & Ravinet, 2021). As reproductive isolation is not always easily measured or defined (Westram *et al.*, 2022), divergence or differentiation is often used as a proxy to place pairs of population along the speciation continuum. Population pairs with F_{ST} values between 0 and 0.19 have been considered to be in the 'divergent populations zone' of the continuum (where all the pairs still experience ongoing gene flow) while population pairs with F_{ST} values above 0.56 are considered 'true species' (where no gene flow is happening) and population pairs falling in between are considered to be

in the 'grey zone' of speciation (De Jode *et al.*, 2022; Roux *et al.*, 2016). Both the Swedish ecotypes and the Spanish ecotypes fall in the 'divergent populations' category with average genome-wide F_{ST} of 0.06 and 0.11, respectively. This is in accordance with our results for the Spanish ecotype: despite the rarity of hybridisation, some level of gene flow is probably still going on. Nevertheless, the Spanish ecotypes are clearly separated by stronger barriers to gene flow than the Swedish ecotypes. In Spain, differentiation in some parts of the genome (e.g., LG12 and LG14) reached the 'grey zone' (F_{ST} of 0.19). This offers the opportunity to understand the factors that result in increased reproductive isolation, through distinguishing the various possibilities outlined above, and so contributing to unravelling the mystery of speciation.

Acknowledgements

We thank Hernan Morales, Jenny Larsson, Pragya Chaube, Isobel Eyres, Samuel Perini, Richard Turney, Graciela Sotelo, and Alison Butlin for help with fieldwork and snail processing. We are grateful to Matteo Fumagalli, Andrea Benazzo, and Sean Stankowski for suggestions on read processing. We also thank Jenny Larsson for guidance on using the ShellShaper program. The analysis of genomic data was conducted on the University of Sheffield high performance computing cluster, ShARC. Funding was provided by European Research Council grant 693030- BARRIERS to RKB; the Swedish Science Research Council (grant number 2021-04191) to KJ; Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2019-2022), the European Union (European Regional Development Fund - ERDF), Xunta de Galicia (ED431C 2020-05) and Ministerio de Ciencia, Innovación y Universidades (PID2021-124930NB-I00) to JG and ERA.

References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolnick, D. I., Hund, A. K., Nosil, P., Peng, F., Ravinet, M., Stankowski, S., Subramanian, S., Wolf, J. B. W., & Yukilevich, R. (2023). A multivariate view of the speciation continuum. *Evolution*, *77*(1), 318–328. https://doi.org/10.1093/evolut/qpac004

Bosso, L., Smeraldo, S., Russo, D., Chiusano, M. L., Bertorelle, G., Johannesson, K., Butlin, R. K., Danovaro, R., & Raffini, F. (2022). The rise and fall of an alien: Why the successful colonizer

Littorina saxatilis failed to invade the Mediterranean Sea. *Biological Invasions*, *24*(10), 3169–3187. https://doi.org/10.1007/s10530-022-02838-y

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric: The most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 2997–3007. https://doi.org/10.1098/rstb.2008.0076

Butlin, R. K., Servedio, M. R., Smadja, C. M., Bank, C., Barton, N. H., Flaxman, S. M., Giraud, T., Hopkins, R., Larson, E. L., Maan, M. E., Meier, J., Merrill, R., Noor, M. A. F., Ortiz-Barrientos, D., & Qvarnström, A. (2021). Homage to Felsenstein 1981, or why are there so few/many species? *Evolution*, *75*(5), 978–988. https://doi.org/10.1111/evo.14235

Butlin, R. K., & Smadja, C. M. (2018). Coupling, Reinforcement, and Speciation. *The American Naturalist*, *191*(2), 155–172. https://doi.org/10.1086/695136

Butlin, R. K., Saura, M., Charrier, G., Jackson, B., Andrè, C., Caballero, A., Coyne, J. A., Galindo, J., Grahame, J. W., Hollander, J., Kemppainen, P., Martinez-Fernandez, M., Panova, M., Quesada, H., Johannesson, K., & Rolán-Alvarez, E. (2014). Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution*, *68*(4), 935–949. https://doi.org/10.1111/evo.12329

Carballo, M., Garcia, C., & Rolan-Alvarez, E. (2001). Heritability of shell traits in wild Littorina saxatilis populations: Results across a hybrid zone. *Journal of Shellfish Research*, *20*(1), 415–422.

Conde-Padín, P., Caballero, A., & Rolán-Alvarez, E. (2009). Relative role of genetic determination and plastic response during ontogeny for shell-shape traits subjected to diversifying selection. *Evolution*, *63*(5), 1356–1363. https://doi.org/10.1111/j.1558-5646.2009.00636.x

Coyne, J. A., & Orr, H. A. (1997). "Patterns of Speciation in Drosophila" Revisited. *Evolution*, *51*(1), 295–303. JSTOR. https://doi.org/10.2307/2410984

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.

Cruz, R., Vilas, C., Mosquera, J., & García, C. (2004). Relative contribution of dispersal and natural selection to the maintenenace of a hybrid zone in Littorina. *Evolution*, *58*(12), 2734–2746. https://doi.org/10.1554/03-644

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Jode, A., Le Moan, A., Johannesson, K., Faria, R., Stankowski, S., Westram, A. M., Butlin, R. K., Rafajlović, M., & Fraïsse, C. (2022). Ten years of demographic modelling of divergence and speciation in the sea. *Evolutionary Applications*, *16*(2), 542–559. https://doi.org/10.1111/eva.13428

Doellman, M. M., Trussell, G. C., Grahame, J. W., & Vollmer, S. V. (2011). Phylogeographic analysis reveals a deep lineage split within North Atlantic Littorina saxatilis. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1722), 3175–3183. https://doi.org/10.1098/rspb.2011.0346

Erlandsson, J., Kostylev, V., & Rolán-Alvarez, E. (1999). Mate search and aggregation behaviour in the Galician hybrid zone of Littorina saxatilis. *Journal of Evolutionary Biology*, *12*(5), 891–896. https://doi.org/10.1046/j.1420-9101.1999.00087.x

Erlandsson, J., Rolán-Alvarez, E., & Johannesson, K. (1998). Migratory differences between ecotypes of the snail Littorina saxatilis on Galician rocky shores. *Evolutionary Ecology*, *12*(8), 913–924. https://doi.org/10.1023/A:1006559904596

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., & Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution*, *4*(8), 1105–1115. https://doi.org/10.1038/s41559-020-1222-6

Fang, B., Kemppainen, P., Momigliano, P., & Merilä, J. (2021). Population Structure Limits Parallel Evolution in Sticklebacks. *Molecular Biology and Evolution*, *38*(10), 4205–4221. https://doi.org/10.1093/molbev/msab144

Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., Rafajlović, M., Panova, M., Ravinet, M., & Johannesson, K. (2019). Multiple chromosomal rearrangements in a hybrid zone between Littorina saxatilis ecotypes. *Molecular Ecology*, *28*(6), 1375–1393. https://doi.org/10.1111/mec.14972

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Felsenstein, J. (1981). Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution*, *35*(1), 124–138. JSTOR. https://doi.org/10.2307/2407946

Funk, D. J., Nosil, P., & Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences*, *103*(9), 3209–3213. https://doi.org/10.1073/pnas.0508653103

Galindo, J., Cacheda, D., Caballero, A., & Rolán-Alvarez, E. (2019). Untangling the contribution of genetic and environmental effects to shell differentiation across an environmental cline in a marine snail. *Journal of Experimental Marine Biology and Ecology*, *513*, 27–34. https://doi.org/10.1016/j.jembe.2019.02.004

Galindo, J., Martínez-Fernández, M., Rodríguez-Ramilo, S. T., & Rolán-Alvarez, E. (2013). The role of local ecology during hybridization at the initial stages of ecological speciation in a marine snail. *Journal of Evolutionary Biology*, *26*(7), 1472–1487. https://doi.org/10.1111/jeb.12152

Galindo, J., P. Moràn, & E. Rolàn-Alvarez. (2009). Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail Littorina saxatilis. *Molecular Ecology*, *18*(5), 919–930. https://doi.org/10.1111/j.1365-294X.2008.04076.x

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Guerrero, R. F., Muir, C. D., Josway, S., & Moyle, L. C. (2017). Pervasive antagonistic interactions among hybrid incompatibility loci. *PLOS Genetics*, *13*(6), e1006817. https://doi.org/10.1371/journal.pgen.1006817

Hearn, K. E. (2023). *Sex determination, adaptive divergence, and the role of inversions in ecotypes of the intertidal snail Littorina saxatilis.* [PhD thesis]. University of Sheffield.

Hearn, K. E., Koch, E. L., Stankowski, S., Butlin, R. K., Faria, R., Johannesson, K., & Westram, A. M. (2022). Differing associations between sex determination and sex-linked inversions in two ecotypes of Littorina saxatilis. *Evolution Letters*, *6*(5), 358–374. https://doi.org/10.1002/evl3.295

Hendry, A. P. (2009). Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(8), 1383–1398. https://doi.org/10.1139/F09-074

Hewitt, G. M. (1989). The subdivision of species by hybrid zones. In *Speciation and its Consequences* (pp. 85–110). Sinauer Associates.

Hollander, J., & Butlin, R. K. (2010). The adaptive value of phenotypic plasticity in two ecotypes of a marine gastropod. *BMC Evolutionary Biology*, *10*(1), 333. https://doi.org/10.1186/1471-2148-10-333

Janson, K. (1983). Selection and migration in two distinct phenotypes of Littorina saxatilis in Sweden. *Oecologia*, *59*(1), 58–61. https://doi.org/10.1007/BF00388072

Jiggins, C. D. (2017). *The ecology and evolution of Heliconius butterflies*. Oxford University Press.

Johannesson, K. (2016). What can be learnt from a snail? *Evolutionary Applications*, *9*(1), 153–165. https://doi.org/10.1111/eva.12277

Johannesson, K., Butlin, R. K., Panova, M., & Westram, A. M. (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*.

Johannesson, K., Johannesson, B., & Rolan-Alvarez, E. (1993). Morphological differentiation and genetic cohesiveness over a microenvironmental gradient in the marine snail Littorina saxatilis. *Evolution*, *47*(6), 1770–1787. JSTOR. https://doi.org/10.2307/2410220

Johannesson, K., Panova, M., Kemppainen, P., André, C., Rolán-Alvarez, E., & Butlin, R. K. (2010). Repeated evolution of reproductive isolation in a marine snail: Unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1547), 1735–1747. https://doi.org/10.1098/rstb.2009.0256

Johannesson, K., Rolán□Alvarez, E., & Ekendahl, A. (1995). Incipient reproductive isolation between two sympatric morphs of the intertidal snail Littorina saxatilis. *Evolution*, *49*(6), 1180–1190. https://doi.org/10.1111/j.1558-5646.1995.tb04445.x

Johannesson, K., Zagrodzka, Z., Faria, R., Marie Westram, A., & Butlin, R. K. (2020). Is embryo abortion a post-zygotic barrier to gene flow between Littorina ecotypes? *Journal of Evolutionary Biology*, *33*(3), 342–351. https://doi.org/10.1111/jeb.13570

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Research*, *25*(6), 918–925. https://doi.org/10.1101/gr.176552.114

Kendall, M. (1948). *Rank correlation methods*. Griffin.

Kess, T., Brachmann, M., & Boulding, E. G. (2021). Putative chromosomal rearrangements are associated primarily with ecotype divergence rather than geographic separation in an intertidal, poorly dispersing snail. *Journal of Evolutionary Biology*, *34*(1), 193–207. https://doi.org/10.1111/jeb.13724

Kess, T., Galindo, J., & Boulding, E. G. (2018). Genomic divergence between Spanish Littorina saxatilis ecotypes unravels limited admixture and extensive parallelism associated with population history. *Ecology and Evolution*, *8*(16), 8311–8327. https://doi.org/10.1002/ece3.4304

Koch, E. L., Morales, H. E., Larsson, J., Westram, A. M., Faria, R., Lemmon, A. R., Lemmon, E. M., Johannesson, K., & Butlin, R. K. (2021). Genetic variation for adaptive traits is associated with polymorphic inversions in Littorina saxatilis. *Evolution Letters*, *5*(3), 196–213. https://doi.org/10.1002/evl3.227

Koch, E. L., Ravinet, M., Westram, A. M., Johannesson, K., & Butlin, R. K. (2022). Genetic architecture of repeated phenotypic divergence in Littorina saxatilis ecotype evolution. *Evolution*, *76*(10), 2332–2346. https://doi.org/10.1111/evo.14602

Kopp, M., Servedio, M. R., Mendelson, T. C., Safran, R. J., Rodríguez, R. L., Hauber, M. E., Scordato, E. C., Symes, L. B., Balakrishnan, C. N., Zonana, D. M., & van Doorn, G. S. (2018). Mechanisms of Assortative Mating in Speciation with Gene Flow: Connecting Theory and Empirical Research. *The American Naturalist*, *191*(1), 1–20. https://doi.org/10.1086/694889

Kostylev, V., Erlandsson, J., & Johannesson, K. (1998). Microdistribution of the polymorphic snail Littorina saxatilis (olivi) in a patchy rocky shore habitat. *Oceanographic Literature Review*, *45*(3), 513.

Kulmuni, J., Butlin, R. K., Lucek, K., Savolainen, V., & Westram, A. M. (2020). Towards the completion of speciation: The evolution of reproductive isolation beyond the first barriers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1806), 20190528. https://doi.org/10.1098/rstb.2019.0528

Larsson, J., Westram, A. M., Bengmark, S., Lundh, T., & Butlin, R. K. (2020). A developmentally descriptive method for quantifying shape in gastropod shells. *Journal of The Royal Society Interface*, *17*(163), 20190721. https://doi.org/10.1098/rsif.2019.0721

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Le Pennec, G., Butlin, R. K., Jonsson, P. R., Larsson, A. I., Lindborg, J., Bergström, E., Westram, A. M., & Johannesson, K. (2017). Adaptation to dislodgement risk on wave-swept rocky shores in the snail Littorina saxatilis. *PLOS ONE*, *12*(10), e0186901. https://doi.org/10.1371/journal.pone.0186901

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics*. https://doi.org/arXiv:1303.3997v2

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lindenbaum, P., & Redon, R. (2018). bioalcidae, samjs and vcffilterjs: Object-oriented formatters and filters for bioinformatics files. *Bioinformatics*, *34*(7), 1224–1225. https://doi.org/10.1093/bioinformatics/btx734

Mérot, C. (2020). Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Molecular Ecology*, *29*(14), 2513–2516. https://doi.org/10.1111/mec.15500

Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., & Butlin, R. K. (2019). Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. *Science Advances*, *5*(12), eaav9963. https://doi.org/10.1126/sciadv.aav9963

Ng, T. P. T., Rolán-Alvarez, E., Dahlén, S. S., Davies, M. S., Estévez, D., Stafford, R., & Williams, G. A. (2019). The causal relationship between sexual selection and sexual size dimorphism in marine gastropods. *Animal Behaviour*, *148*, 53–62. https://doi.org/10.1016/j.anbehav.2018.12.005

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, *12*(6), 443–451. https://doi.org/10.1038/nrg2986

Nonaka, E., Svanbäck, R., Thibert-Plante, X., Englund, G., & Brännström, Å. (2015). Mechanisms by Which Phenotypic Plasticity Affects Adaptive Divergence and Ecological Speciation. *The American Naturalist*, *186*(5), E126–E143. https://doi.org/10.1086/683231

Panova, M., Aronsson, H., Cameron, A. R., Dahl, P., Godhe, A., Lind, U., Ortega-Martinez, O., Pereyra, R., Tesson, S. V. M., Wrange, A.-L., Blomberg, A., & Johannesson, K. (2016). DNA extraction protocols for whole-genome sequencing in marine organisms. In *Marine Genomics, methods in molecular biology* (Vol. 1452, pp. 13–44). Springer.

Panova, M., Blakeslee, A. M. H., Miller, A. W., Mäkinen, T., Ruiz, G. M., Johannesson, K., & André, C. (2011). Glacial History of the North Atlantic Marine Snail, Littorina saxatilis, Inferred from Distribution of Mitochondrial DNA Lineages. *PLOS ONE*, *6*(3), e17511. https://doi.org/10.1371/journal.pone.0017511

Perini, S., Rafajlović, M., Westram, A. M., Johannesson, K., & Butlin, R. K. (2020). Assortative mating, sexual selection, and their consequences for gene flow in Littorina. *Evolution*, *74*(7), 1482–1497. https://doi.org/10.1111/evo.14027

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/

Reid, D. G. (1996). *Systematics and evolution of Littorina* (2009/05/11 ed., Vol. 76). The Ray Society; Cambridge Core. https://www.cambridge.org/core/article/dg-reid-systematics-and-evolution-of-littorina-x-463p-london-the-ray-society-1996-volume-164-of-the-series/C42A72C8BA749D9BEDA00F2D76BAE7CB

Reid, K., Bell, M. A., & Veeramah, K. R. (2021). Threespine Stickleback: A Model System For Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, *22*(1), 357–383. https://doi.org/10.1146/annurev-genom-111720-081402

Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S. R., de Carvalho, C. F., Safran, R. J., Sandoval, C. P., Feder, J., Gries, R., Crespi, B. J., Gries, G., Gompert, Z., & Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, *1*(4), 0082. https://doi.org/10.1038/s41559-017-0082

Rolán-Alvarez, E., Austin, C. J., & Boulding, E. G. (2015). The Contribution of the Genus Littorina to the Field of Evolutionary Ecology. In *Oceanography and Marine Biology* (pp. 166–223). CRC Press.

Rolàn-Alvarez, E., Carballo, M., Juan Galindo, Moràn, Fernàndez, B., Caballero, A., Cruz, R., Boulding, E. G., & Johannesson, K. (2004). Nonallopatric and parallel origin of local reproductive barriers between two snail ecotypes. *Molecular Ecology*, *13*(11), 3415–3424. https://doi.org/10.1111/j.1365-294X.2004.02330.x

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. *Journal of Evolutionary Biology*, *12*(5), 879–890. https://doi.org/10.1046/j.1420-9101.1999.00086.x

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*(6), 1838–1847. JSTOR. https://doi.org/10.2307/2411006

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLOS Biology*, *14*(12), e2000234. https://doi.org/10.1371/journal.pbio.2000234

Seehausen, O., & Wagner, C. E. (2014). Speciation in Freshwater Fishes. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 621–651. https://doi.org/10.1146/annurev-ecolsys-120213-091818

Servedio, M. R., Doorn, G. S. V., Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: 'Magic' but not rare? *Trends in Ecology & Evolution*, *26*(8), 389–397. https://doi.org/10.1016/j.tree.2011.04.005

Servedio, M. R., & Noor, M. A. F. (2003). The Role of Reinforcement in Speciation: Theory and Data. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 339–364. https://doi.org/10.1146/annurev.ecolsys.34.011802.132412

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples)†. *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Signorell, A. (2023). *DescTools: Tools for Descriptive Statistics* (0.99.48) [Computer software]. https://cran.r-project.org/package=DescTools

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*(24), 5123–5140. https://doi.org/10.1111/j.1365-294X.2011.05350.x

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution*, *75*(6), 1256–1273. https://doi.org/10.1111/evo.14215

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. JSTOR. https://doi.org/10.2307/2408641

Westram, A. M., Faria, R., Johannesson, K., & Butlin, R. (2021). Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular Ecology*, *30*(15), 3797–3814. https://doi.org/10.1111/mec.15861

Westram, A. M., Rafajlović, M., Chaube, P., Faria, R., Larsson, T., Panova, M., Ravinet, M., Blomberg, A., Mehlig, B., Johannesson, K., & Butlin, R. (2018). Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. *Evolution Letters*, *2*(4), 297–309. https://doi.org/10.1002/evl3.74

Westram, A. M., Stankowski, S., Surendranadh, P., & Barton, N. (2022). What is reproductive isolation? *Journal of Evolutionary Biology*, *35*(9), 1143–1164. https://doi.org/10.1111/jeb.14005

Yukilevich, R. (2021). Reproductive Character Displacement Drives Diversification of Male Courtship Songs in Drosophila. *The American Naturalist*, *197*(6), 690–707. https://doi.org/10.1086/714046

Data Accessibility

Raw sequence reads are deposited in NCBI SRA (BioProject accession ID PRJNA781449 that will be available upon publication). Phenotypic data (maturity, sex, wet weight, shell thickness, shell ridging, shell banding, shell length, shell growth parameters, and boldness scores) will be archived in Zenodo and scripts on Github upon publication.

Author Contributions

RKB, KJ, AMW, FR, and ADJ designed the study. RKB, KJ, AMW, ZZ, RF, JG, ERA collected data. RKB, FR, ADJ and RF analysed data. FR, ADJ and RKB drafted the manuscript. All authors revised and agreed to the manuscript.

Figures

Figure 1. Patterns of genomic divergence in Spain (**a, c, e, g, i**) and Sweden (**b, d, f, h, l**) as shown by PCA (**a, b, c, d**), admixture (**e, f**) and the Hybrid Index (**g-l**). Transect positions closer to zero correspond to the high shore (Spain) or boulder field (Sweden) while higher values represent the low shore (Spain) or rocky headland (Sweden). Snails in the admixture plots (e, f) were ordered according to their position along shore. In Sweden, a sampling gap occurred at the transect positions 90-120 as seen in the plot of PCA1 and Hybrid Index along shore (d, h) and a single genetic cluster was identified but two groups are shown in the admixture plot (e) to facilitate comparisons with Spain. The Crab and Wave groups are coloured in blue and red, respectively.

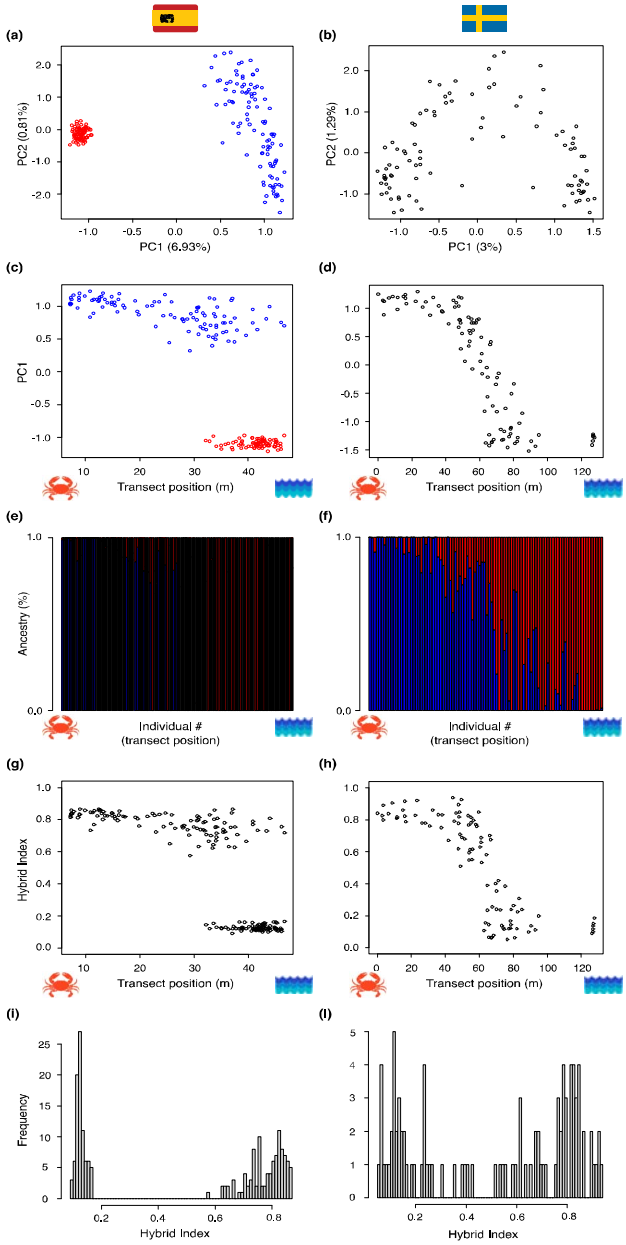

Figure 2. Manhattan plots of per locus F_{ST} values along the genome in Sweden (**a**) and Spain (**b**). The linkage groups and positions in centimorgans are indicated in blue and red, respectively. Chromosomal inversions identified in previous studies are highlighted in red.

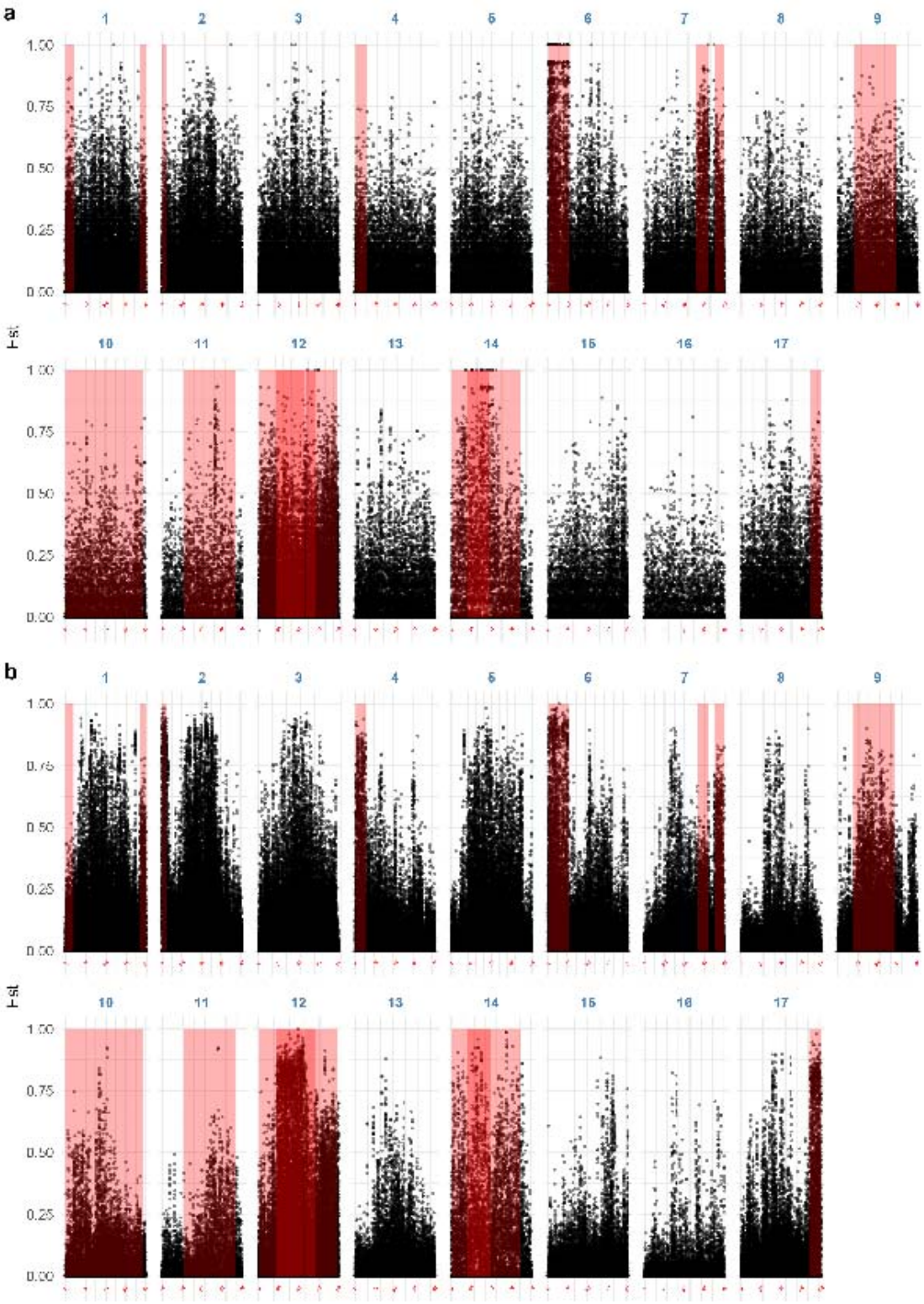
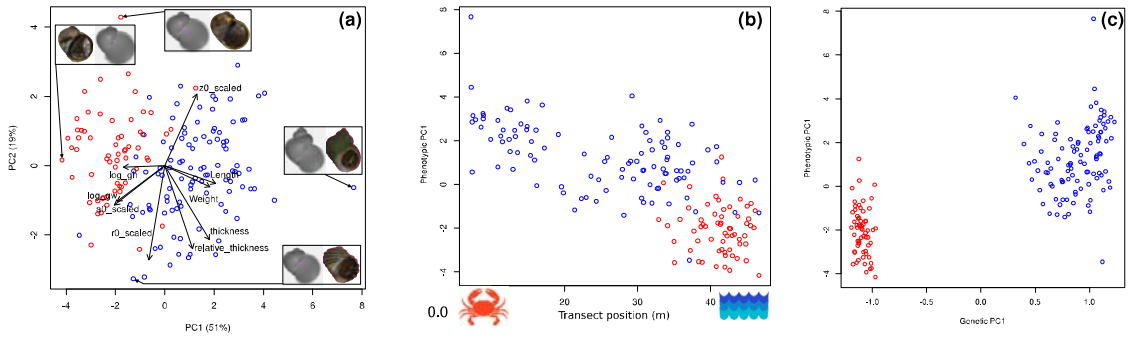
Figure 3. Multivariate phenotypic divergence in Spain as shown by the first two PCA axis (**a**), along transect (**b**), and its relation to genomic differentiation (**c**). Transect positions closer to zero correspond to the high shore while higher values represent the low shore. The Crab and Wave groups are coloured in blue and red, respectively.