Article:

Stankowski, S., Zagrodzka, Z.B., Galindo, J. et al. (2023) Whole-genome phylogeography of the intertidal snail Littorina saxatilis. Evolutionary Journal of the Linnean Society, 2 (1). kzad002. ISSN: 2752-938X

https://doi.org/10.1093/evolinnean/kzad002

Evolutionary Journal of the Linnean Society, 2023, **2**, 1–13
https://doi.org/10.1093/evolinnean/kzad002
Advance access publication 17 August 2023
Original Article

OXFORD

Original Article

Whole-genome phylogeography of the intertidal snail *Littorina saxatilis*

Sean Stankowski[1,2,*], Zuzanna B. Zagrodzka[1], Juan Galindo[3,4], Mauricio Montaño-Rendón[5], Rui Faria[6,7], Natalia Mikhailova[8], April M.H. Blakeslee[9], Einar Arnason[10], Thomas Broquet[11], Hernán E. Morales[12], John W. Grahame[13,14], Anja M. Westram[2,15], Kerstin Johannesson[16], Roger K. Butlin[1,16,*]

[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, S10 2TN Sheffield, UK
[2]Institute of Science and Technology Austria (IST Austria), 3400 Klosterneuburg, Austria
[3]Centro de Investigación Mariña, Departamento de Bioquímica, Genética e Inmunología, Universidade de Vigo, 36310 Vigo, Spain
[4]Centro de Investigación Mariña (CIM-UVIGO), Illa de Toralla, 36331 Vigo, Spain
[5]Sheffield Children's NHS Foundation Trust, Clarkson St, Broomhall, Sheffield S10 2TH, UK
[6]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal
[7]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
[8]Institute of Cytology Russian Academy of Sciences, St. Petersburg, 194064, Russia
[9]Biology Department, East Carolina University, Greenville, NC 27858, USA
[10]Faculty of Life and Environmental Sciences, University of Iceland, Reykjavík 107, Iceland
[11]CNRS & Sorbonne Université, Station Biologique de Roscoff, Roscoff 29680, France
[12]Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, DK-1353 Copenhagen, Denmark
[13]School of Biology, University of Leeds, Leeds LS2 9JT, UK
[14]School of Ocean Sciences, Bangor University, Menai Bridge, Anglesey LL59 5AB, UK
[15]Faculty of Biosciences and Aquaculture, Nord University, N-8049 Bodø, Norway
[16]Department of Marine Sciences, Tjärnö, University of Gothenburg, 405 30 Strömstad, Sweden

[*]Corresponding author. Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK. E-mail: sean.stankowski@ist.ac.at

ABSTRACT

Understanding the factors that have shaped the current distributions and diversity of species is a central and longstanding aim of evolutionary biology. The recent inclusion of genomic data into phylogeographic studies has dramatically improved our understanding in organisms where evolutionary relationships have been challenging to infer. We used whole-genome sequences to study the phylogeography of the intertidal snail *Littorina saxatilis*, which has successfully colonized and diversified across a broad range of coastal environments in the Northern Hemisphere amid repeated cycles of glaciation. Building on past studies based on short DNA sequences, we used genome-wide data to provide a clearer picture of the relationships among samples spanning most of the species natural range. Our results confirm the trans-Atlantic colonization of North America from Europe, and have allowed us to identify rough locations of glacial refugia and to infer likely routes of colonization within Europe. We also investigated the signals in different datasets to account for the effects of genomic architecture and non-neutral evolution, which provides new insights about diversification of four ecotypes of *L. saxatilis* (the crab, wave, barnacle, and brackish ecotypes) at different spatial scales. Overall, we provide a much clearer picture of the biogeography of *L. saxatilis*, providing a foundation for more detailed phylogenomic and demographic studies.

Keywords: ecotype formation; parallel evolution; postglacial colonization; refugia; species history

INTRODUCTION

Understanding the factors that have shaped the current distributions and diversity of species is a central and longstanding aim of biogeographic investigations (Avise 2000, Hominick 2002, Barry Cox *et al.* 2016, Kumar and Kumar 2018). The relatively recent inclusion of genomic data into phylogeographic studies has provided detailed insight into the histories experienced by many organisms from a range of terrestrial and marine

---

Received April 14 2023; revised June 26 2023; accepted for publication August 16 2023

environments (Chase *et al.* 2017, Rosser *et al.* 2017, Berv *et al.* 2021). For example, it is possible to gain detailed insight into how species distributions have been impacted by major historical perturbations like geological events and drastic changes in the earth's climate. By combining molecular and morphological information, we can also gain insight into the relative importance of local distributional changes and natural selection in shaping patterns of phenotypic variation across species geographic ranges (Stankowski and Johnson 2014). These insights not only inform our general understanding of the evolutionary process, but are also relevant to more pragmatic goals like taxonomy and biodiversity conservation (Whittaker *et al.* 2005, Ladle and Whittaker 2011).

The intertidal snail *Littorina saxatilis* (Olivi, 1792) is an example of a species that has successfully colonized and diversified across a broad range of coastal environments in the Northern Hemisphere amid repeated cycles of glaciation (Reid 1996). Although *L. saxatilis* has become an important model organism for eco-evolutionary studies, obtaining a clear picture of the range-wide history of the species has been challenging. Our current understanding of the recent phylogeographic history of *L. saxatilis* comes by way of two phylogeographic analyses, published a decade ago (Panova *et al.* 2011, Doellman *et al.* 2011; recently reanalysed by Blakeslee *et al.* 2021). Based on short mitochondrial DNA sequences (58 and 125 variable sites, respectively) from range-wide samples, these studies revealed highly complex patterns of haplotype sharing across very broad areas of the postglacial North Atlantic coastline. The authors interpreted this pattern as arising from long-term persistence of *L. saxatilis* during multiple glacial cycles coupled with the rapid recolonization of heavily glaciated northern areas of Europe from multiple independent refugia. However, most previously glaciated areas exhibited high haplotype diversity and low haplotype divergence, making the identification of possible refugia and colonization routes challenging. Another possible explanation for the complexity is that the patterns of mtDNA variation do not reflect the demographic history of *L. saxatilis*, either because the mitochondrial genome has a discordant evolutionary history (due to selection or the stochastic effects of demography) or because the information contained in short sequences is low (Avise 1994, Ballard and Whitlock 2004). In these respects, inferences made from genome-wide variation may help to shed new light on the recent history of this species.

A clear understanding of the biogeographic history of *L. saxatilis* is paramount to understanding the repeated diversification of the species into locally adapted ecotypes at many locations across its broad range (Fig. 1). The best-studied ecotypes are the divergent 'crab' and 'wave' morphs that can be found in close proximity to one another on the rocky shores of Spain, Sweden, and the UK. At each location, these ecotypes exhibit divergent shell and behavioural traits that are thought to be adaptations to high levels of crab predation, and wave exposure, respectively (Johannesson *et al.* 2010). Detailed population genetic analyses indicate that these ecotypes have evolved repeatedly at numerous locations rather than having a single origin and then colonizing other locations by dispersal (Butlin *et al.* 2014). In addition to the crab- and wave-adapted populations, two other ecotypes are known from multiple locations across Europe, but have been less well studied. The first is the brackish ecotype, which has been recorded from salt marshes and estuaries in the UK, Sweden, Spain, South Africa, and Nova Scotia with some populations thought to reflect recent human introductions (Knight *et al.* 1987, Reid 1996). This ecotype was originally described as a separate species based on its shell shape and colour pattern (*Littorina tenebrosa*), but after molecular investigation by Janson and Ward (1984) it was synonymized by Reid (1996). The fourth ecotype, the barnacle ecotype, has been recorded at many locations in the UK and France and is the most phenotypically divergent morph of *L. saxatilis*, having an adult shell diameter of just a few millimetres (Fig. 1). Its small size enables it to shelter in and around empty barnacles allowing it to inhabit the lower intertidal zone despite strong wave action (Reid 1996). The barnacle ecotype was also originally described as a separate species (*Littorina neglecta*) but was later made a synonym of *L. saxatilis* (Reid 1996) following molecular investigation (Johannesson and Johannesson 1990). However, for the brackish and barnacle ecotypes, it is unclear if their broad distributions reflect repeated adaptation to a common selective pressure or a single origin and subsequent dispersal to multiple locations (Johannesson *et al.* 2010).

In this study, we used whole-genome sequences to advance our understanding of the phylogeography and diversification of the intertidal snail *L. saxatilis*. We had the following aims. First, we wanted to clarify relationships across the species geographic range, thereby revealing potential glacial refugia and major colonization paths that can be evaluated in light of historical changes in oceanic connectivity. Second, we wanted to infer the relationships among the ecotypes to evaluate evidence for their parallel origins. This was conducted at the regional scale for all four ecotypes (i.e. between countries) and also at a local scale (at pairs of nearby locations within countries) for the crab and wave ecotypes.

MATERIALS AND METHODS

Sampling and anatomical assessment

To address these questions, we collected *L. saxatilis* from 18 locations across the North Atlantic, attempting to capture as much of the geographic distribution and phenotypic variation as possible, but with a focus mainly on Europe where the species is most common and the diversity of ecotypes is highest (Fig. 1; Supporting Information, Table S1). The current distribution of *L. saxatilis* also includes populations in the Mediterranean (Venice), South Atlantic (South Africa), and Pacific (San Francisco Bay) but these are thought to be recent introductions (Reid 1996).

When possible, samples of crab, wave, barnacle, and brackish ecotypes of *L. saxatilis* were collected. However, the sampling of ecotypes in many locations was not exhaustive. Also, in many of the sampled locations, the habitat and shell phenotypes were not typical of the four recognized ecotypes and habitats that have been primarily described from mainland European populations (Reid 1996) so were classified as 'other'. At two of the locations, Silleiro in Spain and the Koster area in Sweden, individuals of the crab and wave ecotypes were collected only metres apart, but in two different locations that were separated by several kilometres.

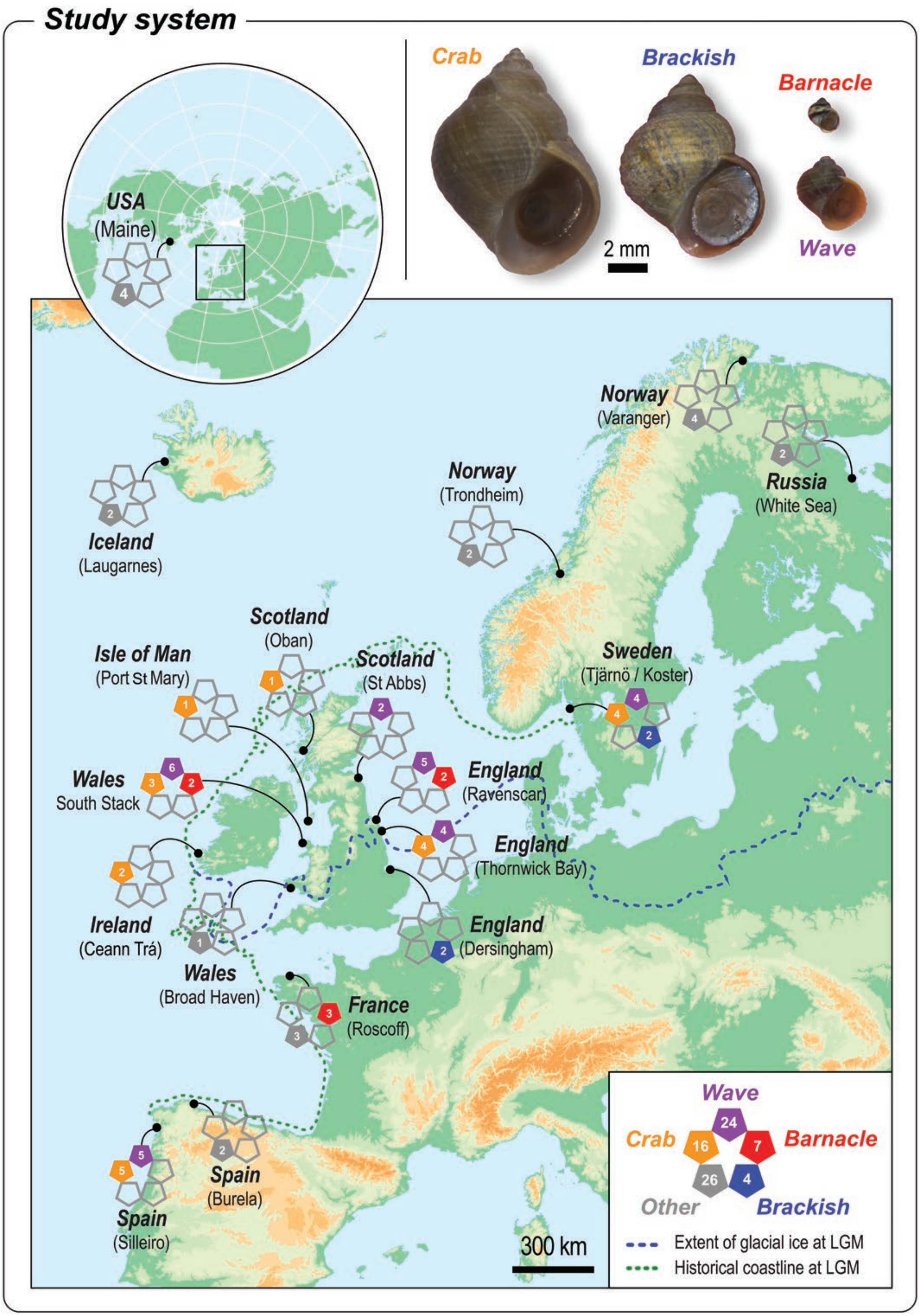

Figure 1. Sample locations and representative images of adult snails of the four ecotypes of *L. saxatilis*. Examples of the shells of reproductively mature crab (from Sweden), wave (Sweden), brackish (Sweden), and barnacle (UK) ecotypes are shown. The black points show the locations of the 17 sites where samples were collected. The number of each ecotype collected is indicated within the five pentagons; 'other' refers to shell phenotypes that could not be classified into one of the four ecotypes (see Material and methods for more details). The area of the main map is represented by the box on the circular map. The dashed blue line in the main map shows the maximum southern extent of glacial ice on land (Jaunsproge 2013), and the dashed green line shows the maximum expansion of the historical coastline (Wang *et al.* 2019) during the LGM. Although not shown on the circular map, glacial ice extended below the USA sample site during the LGM.

Sample processing, DNA extraction, and whole-genome sequencing

We only selected reproductively mature individuals for sequencing. Maturity was determined by examining the reproductive anatomy (Reid 1996). In the UK, France, and Norway, *L. saxatilis* is sympatric with closely-related species (*Littorina arcana* and *Littorina compressa*) and cannot be reliably distinguished based on shell morphology. In these locations, only female *L. saxatilis* were sequenced. These were identified by the presence of a brood pouch, which is a species-diagnostic trait (Reid 1996). Males from these locations were not sequenced because they cannot be reliably assigned to *L. saxatilis* or *L. arcana*. Foot tissue was stored in 99% ethanol prior to DNA extraction.

DNA was extracted from a small piece of foot tissue using a CTAB protocol (Panova *et al.* 2016) at the University of Sheffield, UK. 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 was conducted by Edinburgh Genomics at the University of Edinburgh, UK.

Whole-genome sequencing, mapping, and variant calling

Sequencing adaptors were removed and sequences were trimmed for low quality using Trimmomatic (Bolger *et al.* 2014) and reads shorter than 70 bp were discarded. Raw reads were mapped to the *L. saxatilis* V2 reference genome (Westram *et al.* 2018) using the BWAmem algorithm (Li 2013). PCR duplicates were removed with biobambam2 (Tischler and Leonard 2014). Variant calling was performed using GATK 4.0.7 by executing steps documented in the short variant discovery pipeline (McKenna *et al.* 2010). Briefly, HaplotypeCaller was used to call Single Nucleotide Polymorphisms (SNPs) and indels simultaneously and produce a gzipped variant call format (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 variant call format (VCF) for each subset of contigs. The multi-sample VCFs were then concatenated 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 30 or greater and removed sites with a mean depth of < 5 reads and > 35 reads, Finally, we only retained sites within contigs assigned to one of the 17 *L. saxatilis* linkage groups (for reasons outlined below), and removed all sites with missing data.

Given prior knowledge about the genomic basis of ecotype formation in *L. saxatilis*, we generated three datasets to both understand and minimize the impact of selection on our inference of evolutionary relationships among the samples. The main dataset that we examined, hereafter referred to as the '*inversion-free*' dataset, includes all of the variant sites in the genome except for those that fell within the putative chromosomal rearrangements identified by Faria *et al.* (2019), at least some of which are consistently associated with the divergence between crab and wave ecotypes (Faria *et al.* 2019, Morales *et al.* 2019). Because some of these regions are large (Faria *et al.* 2019), subject to divergent selection (Westram *et al.* 2018, 2021), and associated with parallel ecotype formation across Europe, they have the potential to confound signatures of the demographic history. We therefore removed them according to the outermost inversion coordinates from Westram *et al.* (2018, 2021), including a small (~1 cM) buffer region on each side because the precise breakpoints are not clear (Supplement 1). The second dataset, referred to as '*inversion-and-outlier-free*' dataset, is a subset of the sites in the inversion-free data set, but filtered to reduce the impact of parallel selection on inferences of population history still further. Because the *L. saxatilis* genome consists of hundreds of thousands of very short contigs, we simply removed all assembly contigs that contained outlier loci according to analyses performed in Morales *et al.* (2019) who compared crab and wave samples from many locations across Europe. The outlier loci included SNPs with an F_{ST} in the top 1% of the crab-wave empirical distribution or a BayPass (Gautier 2015) Bayes factor greater than 20. These criteria are strict because many assembly contigs contained outlier loci, meaning that tightly linked loci were also discarded. The third dataset, referred to as the '*full dataset*', included all variable sites that passed our initial depth and quality filters including putative inversions and all outlier loci.

Inference of evolutionary relationships

We used maximum likelihood phylogenetic analysis to reconstruct the evolutionary relationships among the sequenced samples. For all three datasets, we converted SNP calls in VCF format to a concatenated FASTA alignment using a custom perl script. For tree construction, we used the HPC-PTHREADS-SSE3 implementation of the program *RAxML* v.8 (Stamatakis 2014) with the GTRGAMMA model. Support for each node in the ML phylogeny was determined via bootstrap analysis (100 replicates). Resulting topologies were rooted with *L. compressa* (four individuals sequenced for this study) and rendered using Figtree 1.4.4 (Rambaut 2009).

As an alternative to a bifurcating tree, we used the program Admixture 1.3.0 (Alexander *et al.* 2009) to infer relationships between samples based on the inversion-free dataset. Admixture is a maximum-likelihood implementation of the Structure algorithm (Pritchard *et al.* 2000) and estimates the fraction of each individual's genome deriving from *K* clusters. Before the analysis, the dataset was thinned to reduce linkage disequilibrium between linked SNPs as recommended by the authors. This was done in the program plink 2.0 (Chang *et al.* 2015) by removing SNPs that had an r^2 value greater than 0.1 with any other SNP within a 50-SNP sliding window that was advanced by 10 SNPs each time. We tested values of *K* ranging from 2 to 5, performing five independent runs at each value to ensure convergence. The cross-validation error was estimated and assessed to identify suitable values of *K*.

Population genetic analyses

Patterns of within- and between-population genetic diversity are expected to be impacted by recent demographic changes, including population bottlenecks and patterns of connectivity that coincide with recent colonization events (Nei *et al.* 1975). To quantify variation in the level of genetic diversity among the samples, we calculated the proportion of heterozygous sites for each sequenced individual considering only loci that were variable within the set of sequenced samples (i.e. only sites that were variant). We refer to this hereafter as the observed heterozygosity (H_o). To quantify population structure across the range of *L. saxatilis*, we calculated the fixation index F_{ST} (θ) between sample populations (grouping all ecotypes from the same location) with more than two individuals according to Weir and Cockerham (1984). Both of these analyses were conducted using VCFtools (Danecek *et al.* 2011).

RESULTS

Phylogenetic relationships

After variant calling and filtering, we retained roughly 8.9 million variant sites across the sequenced samples. After removing variants that fell within known chromosomal inversions, we retained 5.7 million variable sites upon which we based our main analyses. The further removal of assembly contigs containing loci previously associated with crab and wave divergence reduced the number of variable sites to 1.5 million.

The ML phylogeny for the inversion-free dataset yielded a clear set of relationships among the 77 individuals of *L. saxatilis*, with most nodes showing high support (Fig. 2). However, a few nodes had very low bootstrap support (Fig. 2). Examination of the bootstrap trees reveals that this was due to variation in the placement of samples from the Atlantic coast of the UK, Iceland, and the USA. The placement of these samples also varied among the trees constructed from the three datasets (Supporting Information, Fig. S1).

Samples collected from the same site almost always clustered together in the inversion-free phylogeny with high bootstrap support. This was also the case for the inversion-and-outlier-free dataset, whereas for the full dataset, individuals from the French (Roscoff) and one Welsh site (South Stack) formed multiple clades that were interdigitated with samples from other locations (Supporting Information, Fig. S1). However, the main results of these analyses are consistent among datasets. We now focus primarily on the inversion-free phylogeny, but mention differences between the datasets where relevant.

Of the populations sampled, individuals from the Atlantic coast of western and northern Spain formed the most basal clade of *L. saxatilis* (Fig. 2). Individuals from all other sampled populations formed a second clade with a strongly asymmetrical 'staircase' structure. With the exception of some longer internal branches (e.g. the branches leading to the North American samples and Dersingham) most nodes were separated by relatively short genetic distances. Within this clade (all non-Spanish *L. saxatilis*), individuals from France (Roscoff) were basal to all other samples, followed by samples from sites in the western UK, including Wales (South Stack and Broad Haven), the Isle of Man, and the western coast of Scotland (Oban). Samples from Iceland and the USA formed a clade that was nested within samples from the western UK and Ireland. Samples from the two Norwegian locations and samples from the White Sea (Russia) were nested between samples from the eastern and western sides of the UK. According to the phylogeny, samples from the North Sea, including locations in the eastern UK (St Abbs in Scotland, Thornwick and Ravenscar in England) and Sweden (Tjärnö and Koster area), grouped together and were sister to the samples from Norway and Russia.

Admixture analysis

The Admixture analysis revealed patterns of ancestry that were consistent with our ML tree topology (Fig. 3). Runs with higher values of K had higher validation error (Supporting Information, Table S2), so we only interpret $K = 2$ and $K = 3$. At $K = 2$, the analysis revealed a clear divide between the Spanish *L. saxatilis* and all other samples, but some evidence for shared ancestry between samples from Burela (Spain) and Roscoff (France). At $K = 3$, the Spanish samples show the same results. However, all other samples were assigned to the two additional groups. Samples from the western UK, Ireland, USA, and Iceland all showed clear assignment to one group and samples from the southern reaches of the North Sea (western UK and Sweden) showed clear assignment to the other group, apart from the southernmost sample (Dersingham, brackish ecotype) which was assigned to the western group. Samples from Norway and Russia, which were nested between samples from the eastern and western sides of the UK in our phylogeny, showed substantial assignment to both groups. Evidence for admixture can also be seen in the Scottish samples from the north-eastern UK (St Abbs) and the samples from Iceland.

Patterns of differentiation

Estimates of F_{ST} were calculated within and between major geographic regions: (i) France, Ireland, & west UK (abbreviated to FIwUK), (ii) North Sea (NS), (iii) Norway & Russia (NR), (iv) Dersingham, (v) Iceland, (vi), Spain, and (vi) the USA (Fig. 4). Within northern Europe (i.e. excluding Spain), levels of F_{ST} within regions were generally lower than between regions (Fig. 4). The major exception was for comparisons within the North Sea, where estimates were routinely as high as those between populations from different geographic regions. Focusing on the comparisons between regions, F_{ST} tended to be much lower between France, Ireland, & west UK and Norway & Russia than between either of these regions and the North Sea region, despite the much larger geographic distance between the former two regions. A similar pattern was observed for comparisons with Iceland, in that the difference between Iceland and the North Sea region was much more pronounced than differences both between Iceland and Norway & Russia, and between Iceland and France, Ireland, & west UK. Also consistent with the results of the phylogenetic analysis, F_{ST} estimates involving Dersingham, Spain, and the USA were almost always much higher than those observed between other regions within Europe.

Genetic diversity

Levels of genetic diversity (observed heterozygosity) varied markedly among sample locations, geographic regions and, in two cases, between the ecotypes (Fig. 5). Relatively high heterozygosity was observed for the crab ecotype in Spain (Silleiro), locations in the western UK (Wales, the Isle of Man, Scotland), France, Norway (Trondheim and Varanger), and the north-eastern UK (Scotland). Despite being collected within metres of other another, heterozygosity in the Spanish wave ecotype (Silleiro) was notably lower than the crab ecotype. Lower

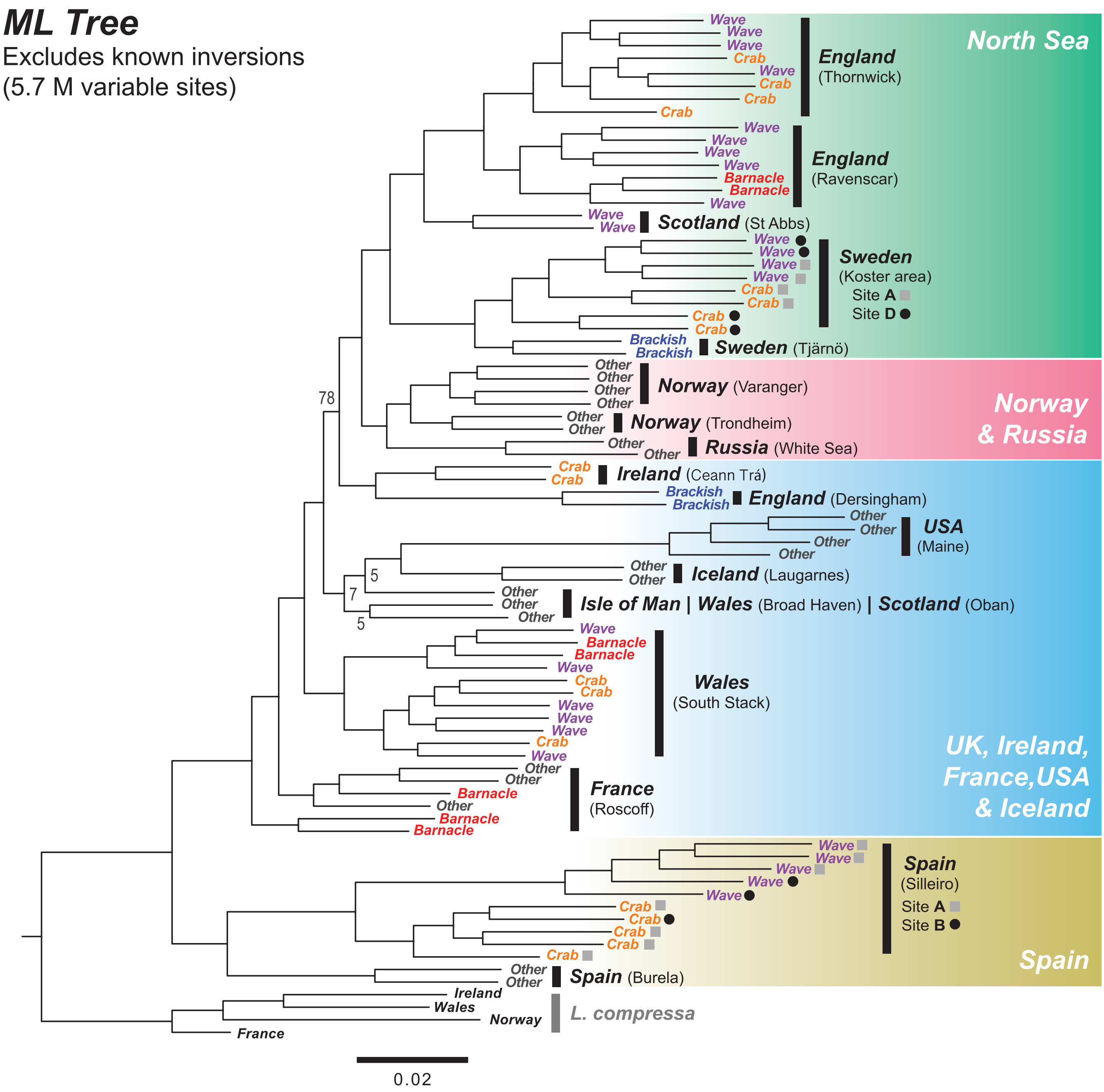

Figure 2. Evolutionary relationships in *L. saxatilis* inferred from whole genome sequences (inversion-free dataset). The maximum likelihood tree was constructed from a concatenated alignment of 5.7 million SNPs and rooted with sequences from *L. compressa*. The coloured tip labels indicate the ecotype of the sequenced sample. The black vertical bars group samples collected from the same locality. The open boxes and closed circles next to the Swedish and Spanish samples show snails collected from two nearby sites in the same area. The coloured boxes enclose samples from broader geographic regions. Bootstrap support for nodes is 100% unless specified. See Supporting Information, Figure S1 for trees of all three datasets.

heterozygosity was also observed for individuals collected from the northern coast of Spain (Burela), Iceland, Dersingham, Russia, and populations on the English and Swedish North Sea coasts. The USA had the lowest heterozygosity of any population. Aside from the difference in heterozygosity between the crab and wave ecotypes in Spain and the lower heterozygosity for the brackish ecotype in Dersingham, there was no consistent difference in the level of genetic diversity between the four ecotypes within or between locations.

Relationships between the ecotypes

Our sequenced individuals included four recognized ecotypes of *L. saxatilis*: the crab, wave, brackish, and barnacle ecotypes, each collected from multiple locations across Europe (Fig. 1). Rather than forming reciprocally monophyletic clades, the four ecotypes, including the lesser known barnacle and brackish ecotypes, all had polyphyletic distributions, clustering by their sampling location instead of by phenotype (Fig. 2). For example,

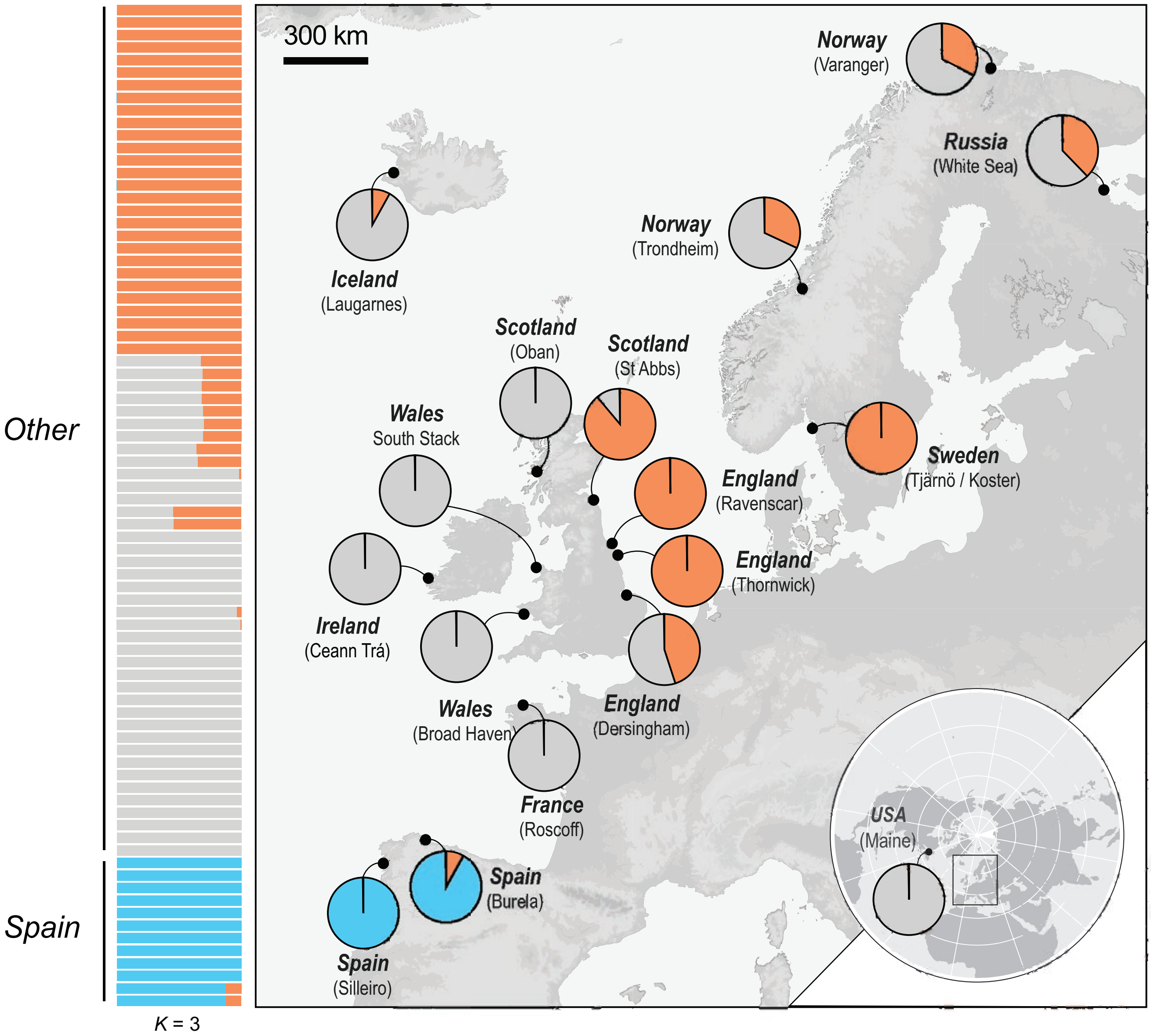

Figure 3. Results of Admixture analysis on all samples at $K = 3$, and geographic patterns of ancestry within the northern *L. saxatilis* at $K = 3$. The coloured bars show the probability of assignment to each of the K clusters. The pies for locations with more than one sample show the average scores calculated across samples from each location.

two individuals of the brackish ecotype collected from a very sheltered bay in Sweden (Tjärnö) clustered with crab and wave individuals collected from nearby locations (Koster area). The two other individuals of the brackish ecotype, collected from a lagoon in the UK (Dersingham), clustered with other samples collected from the UK. Similarly, individuals of the barnacle ecotype, collected in England (Ravenscar), Wales (South Stack), and France (Roscoff), clustered with individuals of other ecotypes collected at the same locations. However, the pattern of clustering for the barnacle ecotype differed slightly in the topology constructed from the full dataset (i.e. with inversions), as some of the barnacle individuals from Wales and France clustered together in a clade that also included individuals of the wave ecotype from Wales (Supporting Information, Fig. S1).

The phylogenetic distributions of the crab and wave ecotypes followed the same general pattern, as individuals sampled from Spain (Silleiro), England (Thornwick), Wales (South Stack), and Sweden (Koster area) clustered by sampling location rather than ecotype (Fig. 2). However, relationships between the crab and wave ecotypes within sites varied among locations and between the three different phylogenetic analyses (Fig. 6). Despite both being collected from two different sites located about 3 km apart, the crab and wave ecotypes on the Spanish west coast formed separate monophyletic clades in the trees for the full dataset and the inversion-free dataset; in the inversion-and-outlier-free analysis, the wave samples formed a monophyletic group, but were nested within the crab samples. In Sweden, the wave samples also formed a monophyletic group that was nested within the crab clade in the inversion-free and inversion-and-outlier-free analyses, but both ecotypes formed monophyletic groups in the full dataset. In England, the crab samples formed a monophyletic group that was nested within wave in the tree for the full

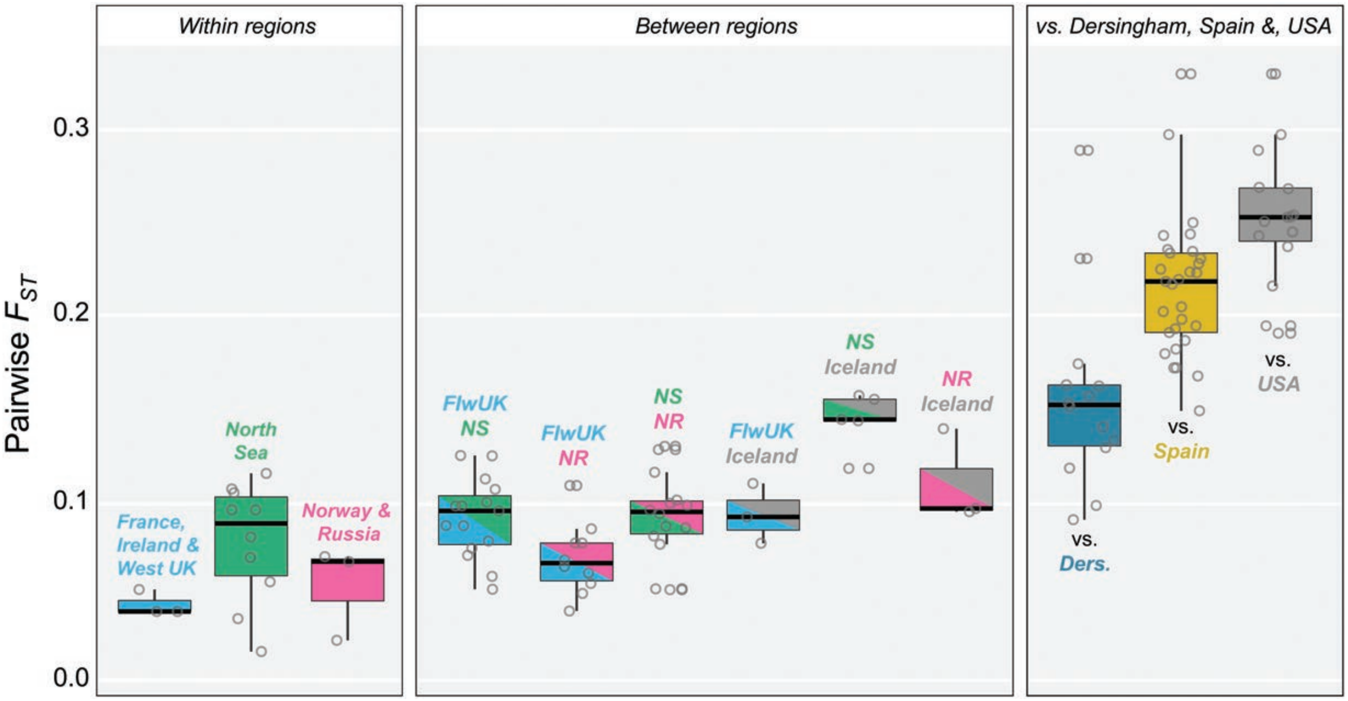

Figure 4. Estimates of pairwise F_{ST} between sample locations. Boxplots show the distributions of F_{ST} calculated for all possible pairs of sites within regions, between regions, and for comparisons that included Spain and the USA. The circles show the values of F_{ST} for each population pair, jittered along the *x*-axis for visibility. FIwUK, France, Ireland, and west UK; NS, North Sea; NR, Norway and Russia. Dersingham, Spain, and the USA were compared to all other sample sites.

dataset, but were interdigitated within wave in the inversion-free and inversion-and-outlier-free trees. At the Welsh site, crab and wave samples were interdigitated in all three analyses.

DISCUSSION

Using whole genome sequences and wide sampling, this study aimed to gain new insight into the phylogeography and origins of ecological diversification of the intertidal snail *L. saxatilis*. The results of our analyses are broadly compatible with the results of the previous studies (Doellman *et al.* 2011, Panova *et al.* 2011), accounting for the limitations of short mtDNA sequences and differences in sampling locations between the studies. However, the increased power of genome-wide data allows us to identify possible glacial refugia more confidently and infer likely routes of colonization within Europe. Also, the phylogenetic placement of samples of four previously recognized ecotypes of *L. saxatilis*, including the well-studied crab and wave ecotypes and lesser known brackish and barnacle ecotypes, provides new insights about diversification of the species at different spatial scales.

Phylogeography of *L. saxatilis*

Although *L. saxatilis* has become an important model organism for eco-evolutionary studies (Johannesson *et al.* 2017), obtaining a clear phylogeography of the species has been difficult. Two previous studies based on short mtDNA sequences yielded complex patterns of haplotype sharing across very broad areas of the postglacial North Atlantic coastline and also between species. Although consistent with the recolonization of heavily glaciated northern areas of Europe from multiple refugia, the broad sharing of haplotypes among populations (and even with other species of *Littorina*) made it difficult to identify refugia or routes of recolonization (Doellman *et al.* 2011, Panova *et al.* 2011). Fortunately, whole-genome sequences give a clearer picture of the geographic relationships, thus providing new insight.

Based on the phylogenetic relationships and levels of genetic diversity and differentiation among the sequenced samples, we identify several geographic regions where large populations may have been maintained during the last ice age. The most apparent of these is the Iberian coast line, which was free of glacial ice. The strong divergence between the Spanish and French/Welsh samples suggests the existence of a phylogeographic discontinuity somewhere between Iberia and France, and is consistent with a gap in the distribution of *L. saxatilis* in the Bay of Biscay (Reid 1996). Indeed, this break has been observed in studies of population structure in many marine organisms, including seahorses (Riquet *et al.* 2019), fish (Larmuseau *et al.* 2009), mammals (Fontaine *et al.* 2007), and other gastropods of the genus *Littorina* (Sotelo *et al.* 2020). This region also coincides with the northern or southern limit of the native ranges of many marine species (Southward *et al.* 1995). Although higher temperature, depth, and associated factors (e.g. nutrients) have been suggested to play a major role in the biogeographic discontinuity in the Bay of Biscay, lack of suitable habitat (i.e. southern French Atlantic shores are mainly sandy) is thought to be another factor limiting dispersal of species inhabiting rocky shores, such as *L. saxatilis*.

Samples from Burela and Silleiro, from northern and western Spanish shores respectively, cluster in two divergent clades, independently of the ecotypes they represent (Fig. 2). This is in agreement with another previously described genetic discontinuity around Burela (Quesada *et al.* 2007, Doellman *et al.* 2011,

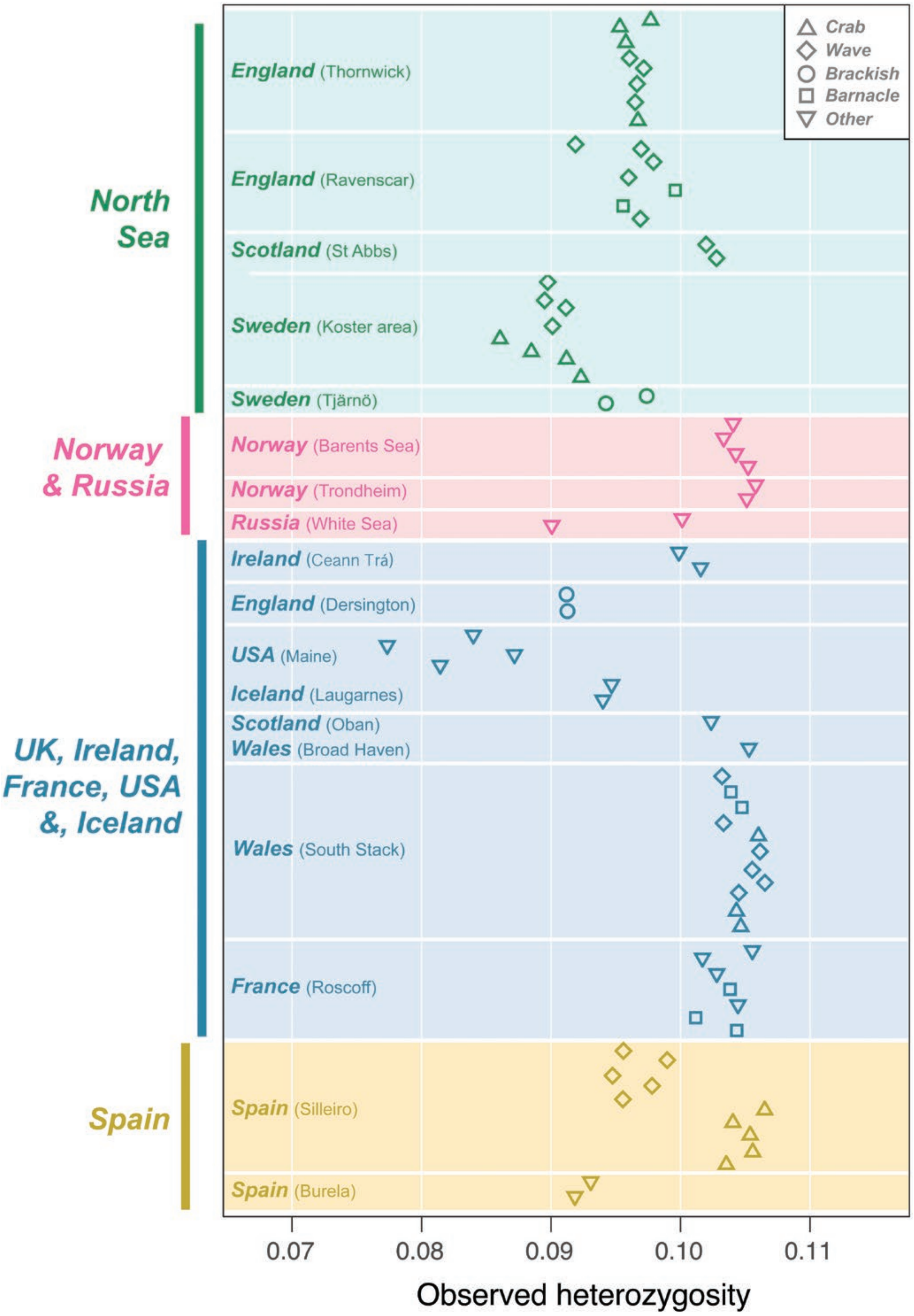

Figure 5. Genetic diversity (proportion of heterozygous sites) for each individual. Individuals are arranged in the same order as in the phylogeny in Figure 2. The different symbols indicate the ecotype of each individual. The four colours coincide with the geographic regions in Figure 2.

Tirado *et al.* 2016). Further sampling in southern France and the north of the Iberian Peninsula may help refine the geographic location of these discontinuities and test for potential contact between these lineages.

The relatively high genetic diversity in French and Welsh samples also suggests that these regions supported large, stable populations of *L. saxatilis* during the last ice age, which is consistent with these areas having been south of, or near the

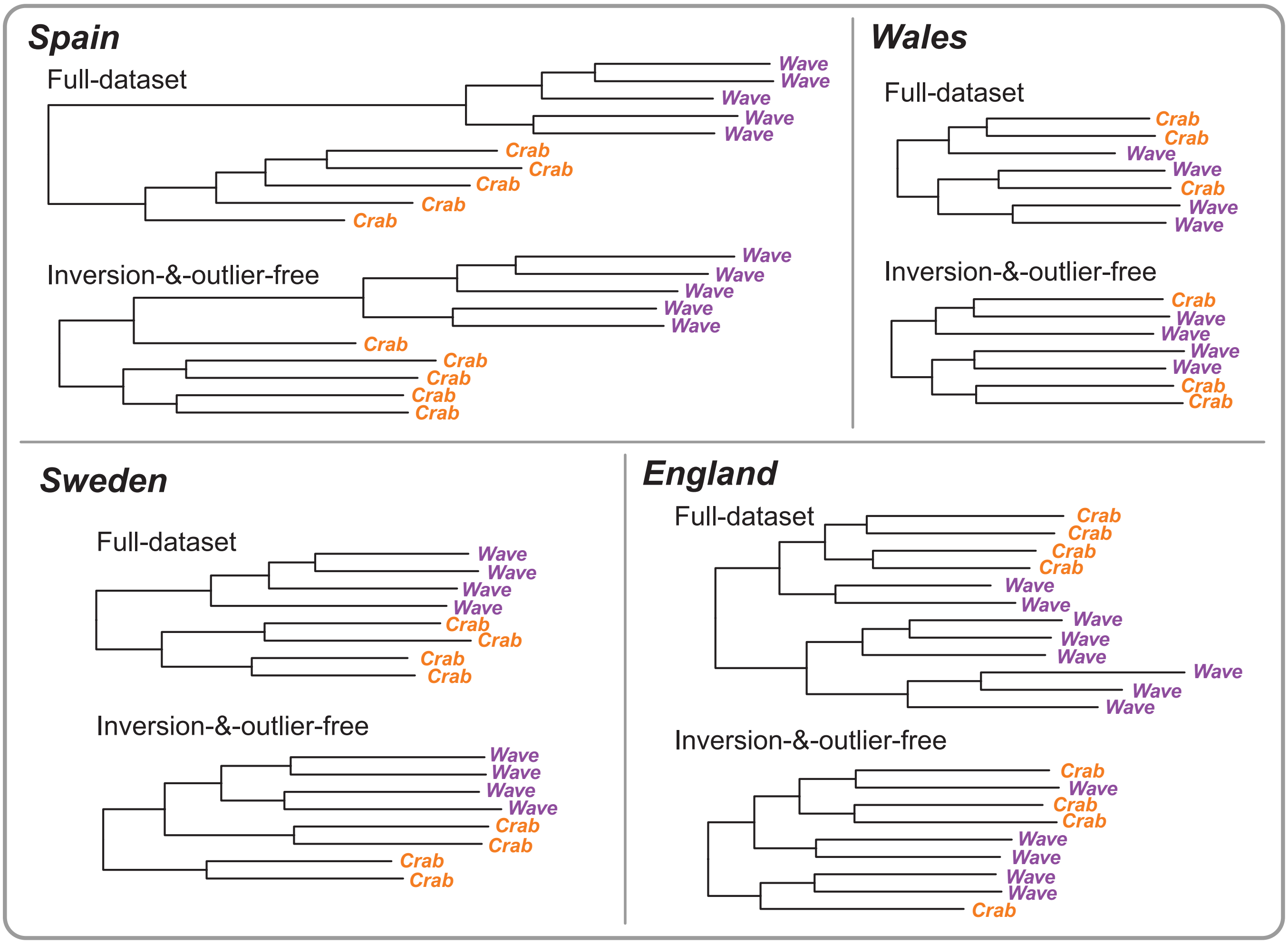

Figure 6. Phylogenetic clustering of crab and wave samples in four different regions and two different datasets. The full dataset includes chromosomal inversions and loci that show exceptional divergence or evidence for selection between the crab and wave ecotypes. The inversion-and-outlier-free dataset is filtered to remove these loci in order to minimize the effect of divergent selection on the phylogenetic inference. For relationships inferred using the inversion-free tree, see Figure 2.

southern limit of glacial ice sheets at the Last Glacial Maximum (LGM). There is evidence that ancestral populations near these locations were the source of colonization of other areas. For example, the phylogenetic position and low genetic diversity of the samples from the USA is fully consistent with the trans-Atlantic colonization of North America from this region, via Iceland, which groups with the USA in our trees. Although this should be interpreted with caution due to the low support for the samples from the USA (Fig. 2), a similar conclusion was also drawn in two previous phylogeographic studies of *L. saxatilis* based on mtDNA sequences (Doellman *et al.* 2011, Panova *et al.* 2011). However, a major difference between this study and the earlier phylogeographic studies based on mtDNA is that previous studies found two divergent mitochondrial clades on the North American coast, suggesting two separate colonizations before the LGM: one on the north-eastern coast and one on the south-eastern coast (Doellman *et al.* 2011, Panova *et al.* 2011). In our study, we were only able to include samples from one site located on the northern part of the coast, so future studies may reveal a second source of trans-Atlantic colonization. More samples from intermediate locations (e.g. Greenland) and more sophisticated analyses may help resolve the phylogenetic position of samples from the USA and the sources and relative timing of the trans-Atlantic colonization of *L. saxatilis*.

Unlike the samples from France and Wales, samples from the North Sea show reduced genetic diversity and elevated among-site genetic differentiation. These results could either reflect the postglacial recolonization of the North Sea, which was part of continental Europe at the LGM, or a substantial reduction in effective population size in the event that a refugial population managed to persist in an area of suitable coastline. These alternatives are very difficult to distinguish given the samples are genomic tools that are currently available for this species, but either way our results demonstrate that these populations have experienced a glacial history.

Our results also suggest that the most northern populations (Norway and Russia and Iceland) may have been recolonized following the secondary contact and admixture of the differentiated eastern and western populations. This result was clearest in our Admixture analysis, which revealed that samples from Norway and Russia showed assignment to the eastern and western clusters (Fig 3). This result is also consistent with the unexpected phylogenetic position of the samples from Norway and Russia, which fell between the samples from the western Atlantic and the samples from the North Sea. In this case, we do not find strong evidence for a reduction in heterozygosity following colonization, probably because the mixing of differentiated groups results in increased levels of genetic variation. Further work is needed to understand how patterns of admixture in these populations vary along the genome, and to test for similar patterns in *Littorina* populations that are distributed even further to the north (e.g. Svalbard, which is above the 77th parallel).

Ecotype formation

In addition to gaining insight into the phylogeography of *L. saxatilis*, we also wanted to gain deeper insight into the phenotypic diversification of the species across Europe. Our phylogenetic analysis included four ecotypes that inhabit very different environments and have highly divergent morphologies. At the broad geographic scale, all of the ecotypes, including the lesser-studied brackish and barnacle ecotypes, clustered by location rather than based on their morphology. Given that these populations are separated by hundreds or thousands of kilometres, the result is explained most simply by the parallel demographic origin of all four ecotypes due to the presence of similar environmental conditions across Europe. It is, however, important to note that a parallel demographic origin does not mean that the ecotypes have a genetically independent origin. For example, in the crab and wave ecotypes, there is strong evidence that the same alleles, often in chromosomal inversions, have fuelled local adaptation in multiple locations, either because they spread through gene flow or were present in the common ancestor to these populations (Johannesson *et al.* 2010, Butlin *et al.* 2014, Morales *et al.* 2019). This may also be the case for the barnacle ecotype, as individuals from France and Wales show stronger clustering by ecotype when chromosomal inversions and crab-wave outlier loci are included in the analysis. If true, this further suggests that loci associated with crab-wave divergence may also contribute to adaptation in other environments.

One possible example where the distribution of ecotypes may reflect recent long-distance colonization is for the brackish ecotype in Dersingham (Fig. 2). This population exists on a long stretch of coast where *L. saxatilis* is otherwise largely absent, and is the only population on the North Sea coast that clusters with samples from the western UK. These facts, coupled with the low genetic diversity and high differentiation (F_{ST}), all point to colonization from an Irish or east UK source. Long-distance colonization might be more likely for the brackish ecotype, possibly through transport by wading birds that use these habitats. Long distance colonization of this kind could also explain sporadic populations of the brackish ecotype of *L. saxatilis* in salt marshes and estuaries on the Atlantic coast of Africa, which show evidence for strong founder effects (Knight *et al.* 1987). A more detailed study of the brackish ecotype is needed to test this hypothesis more thoroughly.

Patterns of clustering and genetic diversity also shed some light on the local origins of the crab and wave ecotypes at the different locations. For instance, in Sweden it has been suggested that the wave ecotype may be the ancestral form, and the crab ecotype the derived form (e.g. Butlin *et al.* 2014). However, in the inversion-free and inversion-and-outlier-free analyses, wave samples are nested within the clade of crab samples. The same pattern was observed in Spain, where the wave samples were nested within crab and show substantially reduced diversity. However, in England, the crab ecotype appears to have arisen from a wave ancestor, suggesting that the derived ecotype may vary among locations—probably due to factors like the history of colonization and changes in the distribution of habitats over time. Although phylogenetic nesting and lower diversity are expected for the derived form, it is important to note that these signals are not conclusive because population size changes and gene flow since the time of origin can modify both patterns. Demographic modelling will be needed to test these hypotheses in the future.

Our study design also enabled us to test for evidence of parallel evolution of the crab and wave ecotypes at a finer geographic scale. Specifically, we sequenced both ecotypes, collected within metres of one another, but at two different sites located a few kilometres apart, in Spain and in Sweden (~3 km in Spain and ~4 km in Sweden). Considering the ecotypes at this geographic scale, we see different patterns of clustering within Spain and Sweden. The simplest explanation for this result is that the ecotypes arose once in each region and colonized similar habitats nearby. The support for this hypothesis is greatest in Spain; the reduced genetic diversity in the wave ecotype suggests that its origin coincided with a population bottleneck, the genetic signature of which is still visible following its dispersal to multiple locations. However, to explain the clustering of samples by ecotype rather than location in Sweden and Spain, this hypothesis also requires that there is a substantial barrier to gene flow between the ecotypes that allows them to remain genetically differentiated despite evidence for local hybridization. A recent study by Kess *et al.* (2018) provides some evidence for the presence of strong reproductive isolation between the crab and wave ecotypes in Spain, as reduced-representation sequencing of both morphs and phenotypic intermediates (i.e. putative hybrids) revealed very few individuals showing evidence of mixed crab-wave ancestry. In Sweden, there is also evidence for a genome-wide barrier to gene flow, as a large proportion of loci across the genome show clinal change between ecotypes (Westram *et al.* 2018); however, there is clearly more extensive admixture, indicated by high frequencies of hybrids with a wide range of hybrid indices in the contact zone. More sophisticated model-based analyses will be needed to test these hypotheses in the future.

SUPPLEMENTARY DATA

Supplementary data are available at *Evolutionary Journal of the Linnean Society* Journal online.

Figure S1. Evolutionary relationships in *L. saxatilis* inferred from full and inversion-and-outlier-free datasets.

Table S1. Table of sampling locations with latitude and longitude.

Table S2. Estimates of cross validation error calculated in Admixture.

ACKNOWLEDGEMENTS

Isobel Eyres, Richard Turney, Graciela Sotelo, Jenny Larson, and Stéphane Loisel helped with the collection and processing of samples. Petri Kemppainen kindly provided samples from Trondheim Fjord. Mark Dunning helped with the development of bioinformatic pipelines. The analysis of genomic data was conducted on the University of Sheffield high-performance computing cluster, ShARC. Funding was provided by the Natural Environment Research Council (NERC) and the European Research Council (ERC). J.G. was funded by a Juntas Industriales y Navales (JIN) project (Ministerio de Ciencia, Innovación y Universidades, code RTI2018-101274-J-I00).

AUTHOR CONTRIBUTIONS

Sean Stankowski (Conceptualization, Formal Analysis, Writing—original draft, Writing—review & editing with input from the other authors), Anja M. Westram (Conceptualization), Roger K. Butlin (Conceptualization), Kerstin Johannesson (Conceptualization), Zuzanna B. Zagrodzka (Formal Analysis), all other authors provided samples, advice, and stimulating discussion and comments on the manuscript

DATA ACCESSIBILITY

Raw sequence data has been deposited on the NCBI short read archive (Bioproject ID: PRJNA626520). Scripts used to perform these steps are available at https://github.com/seanstankowski/Littorina_phylogeography.

REFERENCES

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 2009;19:1655–64.

Avise JC. *Molecular Markers, Natural History and Evolution*. New York: Chapman and Hall, 1994.

Avise, JC. *Phylogeography: The History and Formation of Species*. Harvard: Harvard University Press, 2000.

Ballard JWO, Whitlock MC. The incomplete natural history of mitochondria. *Molecular Ecology* 2004;13:729–44.

Barry Cox C, Moore PD, Ladle RJ. *Biogeography: An Ecological and Evolutionary Approach*. Chichester: John Wiley & Sons, 2016.

Berv JS, Campagna L, Feo TJ et al. Genomic phylogeography of the white-crowned manakin *Pseudopipra pipra* (Aves: Pipridae) illuminates a continental-scale radiation out of the Andes. *Molecular Phylogenetics and Evolution* 2021;164:107205.

Blakeslee AMH, Miller AW, Ruiz GM et al. Population structure and phylogeography of two North Atlantic *Littorina* species with contrasting larval development. *Marine Biology* 2021;168:1–16. https://doi.org/10.1007/s00227-021-03918-8

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.

Butlin RK, Saura M, Charrier G et al. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* 2014;68:935–49.

Chang CC, Chow CC, Tellier LC et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* 2015;4:7.

Chase MA, Stankowski S, Streisfeld MA. Genomewide variation provides insight into evolutionary relationships in a monkeyflower species complex (*Mimulus* sect. *Diplacus*). *American Journal of Botany* 2017;104:1510–21.

Danecek P, Auton A, Abecasis G et al.; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156–8.

Doellman MM, Trussell GC, Grahame JW et al. Phylogeographic analysis reveals a deep lineage split within North Atlantic *Littorina saxatilis*. *Proceedings Biological Sciences* 2011;278:3175–83.

Faria R, Chaube P, Morales HE et al. Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Molecular Ecology* 2019;28:1375–93.

Fontaine MC, Baird SJE, Piry S et al. Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *BMC Biology* 2007;5:30.

Gautier M. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* 2015;201:1555–79.

Hominick WM. Biogeography. *Entomopathogenic Nematology* 2002;1:115–43.

Janson K, Ward RD. Microgeographic variation in allozyme and shell characters in *Littorina saxatilis* Olivi (Prosobranchia: Littorinidae). *Biological Journal of the Linnean Society* 1984;22:289–307.

Jaunsproge M. Cross-media 3D cartography of ‘Europe at the last ice age’ based on initial data compilations. *Master’s Thesis*, Institute for Cartography, Technische Universität Dresden, 2013.

Johannesson K, Butlin RK, Panova M, et al. Mechanisms of adaptive divergence and speciation in *Littorina saxatilis*: integrating knowledge from ecology and genetics with new data emerging from genomic studies. In: Oleksiak M, Rajora O (eds), *Population genomics: marine organisms*. Population Genomics. Springer, Cham, 2017, 277–301.

Johannesson K, Johannesson B. Genetic variation within *Littorina saxatilis* (Olivi) and *Littorina neglecta* Bean: Is *L. neglecta* a good species? *Hydrobiologia* 1990;193:89–97.

Johannesson K, Panova M, Kemppainen P et al. Repeated evolution of reproductive isolation in a marine snail: Unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society London B: Biological Sciences* 2010;365:1735–47.

Kess T, Galindo J, Boulding EG. Genomic divergence between Spanish *Littorina saxatilis* ecotypes unravels limited admixture and extensive parallelism associated with population history. *Ecology and Evolution* 2018;8:8311–27.

Knight AJ, Hughes RN, Ward RD. A striking example of the founder effect in the mollusc *Littorina saxatilis*. *Biological Journal of the Linnean Society* 1987;32:417–26.

Kumar R, Kumar V. A review of phylogeography: biotic and abiotic factors. *Geology, Ecology, and Landscapes* 2018;2:268–74.

Ladle RJ, Whittaker RJ. *Conservation Biogeography*. Chichester: John Wiley & Sons, 2011.

Larmuseau MHD, Van Houdt JKJ, Guelinckx J et al. Distributional and demographic consequences of Pleistocene climate fluctuations for a marine demersal fish in the north-eastern Atlantic. *Journal of Biogeography* 2009;36:1138–51.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, https://arxiv.org/abs/1303.3997, 16 March 2013, preprint: not peer reviewed.

McKenna A, Hanna M, Banks E et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 2010;20:1297–303.

Morales HE, Faria R, Johannesson K et al. Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. *Science Advances* 2019;5:eaav9963.

Nei M, Maruyama T, Chakraborty R. The bottleneck effect and genetic variability in populations. *Evolution* 1975;29:1–10.

Olivi G. *Zoologia adriatica: ossia catalogo ragionato degli animali del Golfo e delle Lagune di Venezia; preceduto da una dissertazione sull storia fisica e naturale del Golfo; e accompagnato da memorie, ed osservazioni di fisica storia naturale ed economia.* G. Remondini e fl., 1792.

Panova M, Aronsson H, Cameron RA et al. DNA extraction protocols for whole-genome sequencing in marine organisms. *Methods in Molecular Biology* 2016;1452:13–44.

Panova M, Blakeslee AMH, Miller AW et al. Glacial history of the North Atlantic marine snail, *Littorina saxatilis*, inferred from distribution of mitochondrial DNA lineages. *PLoS One* 2011;6:e17511.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–59.

Quesada H, Posada D, Caballero A et al. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* 2007;61:1600–12.

Rambaut A. *FigTree v1. 3.1*, 2009. http://tree.bio.ed.ac.uk/software/figtree/

Reid DG. *Systematics and Evolution of Littorina*. London: The Ray Society, 1996.

Riquet F, Liautard-Haag C, Woodall L et al. Parallel pattern of differentiation at a genomic island shared between clinal and mosaic hybrid zones in a complex of cryptic seahorse lineages. *Evolution* 2019;73:817–35.

Rosser NL, Thomas L, Stankowski S et al. Phylogenomics provides new insight into evolutionary relationships and genealogical discordance in the reef-building coral genus *Acropora*. *Proceedings of the Royal Society B: Biological Sciences* 2017;284:20162182. https://doi.org/10.1098/rspb.2016.2182

Sotelo G, Duvetorp M, Costa D et al. Phylogeographic history of flat periwinkles, *Littorina fabalis* and *L. obtusata*. *BMC Evolutionary Biology* 2020;20:23.

Southward AJ, Hawkins SJ, Burrows MT. Seventy years' observations of changes in distribution and abundance of zooplankton and intertidal organisms in the western English Channel in relation to rising sea temperature. *Journal of Thermal Biology* 1995;20:127–55.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–3.

Stankowski S, Johnson MS. Biogeographic discordance of molecular phylogenetic and phenotypic variation in a continental archipelago radiation of land snails. *BMC Evolutionary Biology* 2014;14:2.

Tirado T, Saura M, Rolán-Alvarez E et al. Historical biogeography of the marine snail *Littorina saxatilis* inferred from haplotype and shell morphology evolution in NW Spain. *PLoS One* 2016;11:e0161287.

Tischler G, Leonard S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code for Biology and Medicine* 2014;9:13.

Wang P, Rosati JD, Vallee M. Coastal Sediments 2019. In: *Proceedings of the 9th International Conference on Coastal Sediments 2019*, Tampa/St. Petersburg, Florida, USA, 27–31. World Scientific Publishing Company Pte Limited, Singapore, 2019.

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 1984;38:1358–70.

Westram AM, Faria R, Johannesson K et al. Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular Ecology* 2021;30:3797–814.

Westram AM, Rafajlović M, Chaube P et al. Clines on the seashore: the genomic architecture underlying rapid divergence in the face of gene flow. *Evolution Letters* 2018;2:297–309.

Whittaker RJ, Araújo MB, Jepson P et al. Conservation biogeography: assessment and prospect. *Diversity and Distributions* 2005;11:3–23.