

## ORIGINAL RESEARCH

# Genome-wide association analyses reveal polygenic genomic architecture underlying divergent shell morphology in Spanish *Littorina saxatilis* ecotypes

Tony Kess  | Elizabeth G. Boulding 

Department of Integrative  
Biology, University of Guelph, Guelph, ON,  
Canada

**Correspondence**

Tony Kess, Department of Integrative  
Biology, University of Guelph, Guelph, ON,  
Canada.  
Email: tonykess@gmail.com

**Present address**

Tony Kess, Fisheries and Oceans Canada, St.  
John's, NL, Canada

**Funding information**

Ministerio de Economía y Competitividad,  
Grant/Award Number: CGL2016-75482-P;  
Canadian Network for Research and  
Innovation in Machining Technology, Natural  
Sciences and Engineering Research Council  
of Canada

**Abstract**

Gene flow between diverging populations experiencing dissimilar ecological conditions can theoretically constrain adaptive evolution. To minimize the effect of gene flow, alleles underlying traits essential for local adaptation are predicted to be located in linked genome regions with reduced recombination. Local reduction in gene flow caused by selection is expected to produce elevated divergence in these regions. The highly divergent crab-adapted and wave-adapted ecotypes of the marine snail *Littorina saxatilis* present a model system to test these predictions. We used genome-wide association (GWA) analysis of geometric morphometric shell traits associated with microgeographic divergence between the two *L. saxatilis* ecotypes within three separate sampling sites. A total of 477 snails that had individual geometric morphometric data and individual genotypes at 4,066 single nucleotide polymorphisms (SNPs) were analyzed using GWA methods that corrected for population structure among the three sites. This approach allowed dissection of the genomic architecture of shell shape divergence between ecotypes across a wide geographic range, spanning two glacial lineages. GWA revealed 216 quantitative trait loci (QTL) associated with shell size or shape differences between ecotypes, with most loci explaining a small proportion of phenotypic variation. We found that QTL were evenly distributed across 17 linkage groups, and exhibited elevated interchromosomal linkage, suggesting a genome-wide response to divergent selection on shell shape between the two ecotypes. Trait-associated loci showed partial overlap with previously identified outlier loci under divergent selection between the two ecotypes, supporting the hypothesis of diversifying selection on these genomic regions. These results suggest that divergence in shell shape between the crab-adapted and wave-adapted ecotypes is produced predominantly by a polygenic genomic architecture with positive linkage disequilibrium among loci of small effect.

**KEYWORDS**

genome-wide association, genomic architecture, geometric morphometrics, *Littorina saxatilis*

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Local adaptation to ecological conditions is the central mechanism driving reproductive isolation during ecological speciation (Schluter, 2001). Important to understanding this process is uncovering the genomic changes underlying evolution of adaptations contributing to reproductive isolation (Coyne & Orr, 2004; Hoban et al., 2016). The identification of speciation genes and their associated genome-wide patterns of differentiation can aid in better understanding how and when local adaptation can give rise to the differentiation of species (Nosil & Schluter, 2011). Research on the “how” of species differentiation at the genetic level has been aided by development of models describing differentiation at a genomic scale, in which initially divergent genomic regions house those genes involved in early local adaptation (Wu, 2001). As divergence progresses, these genomic regions of reduced gene flow increase in size, resulting in a mosaic pattern of heterogeneous differentiation when comparing genomes of speciating lineages (Nosil, Funk, & Ortiz-Barrientos, 2009; Via & West, 2008). The homogenizing effect of ongoing gene flow during this process may select for genomic architectures that minimize recombination and reduce gene flow, through increased establishment probabilities of mutations that produce larger phenotypic effect (Griswold, 2006; Yeaman & Whitlock, 2011). Loci responsible for locally adapted traits are thus predicted to localize primarily in a few linked regions of the genome that exhibit elevated patterns of differentiation relative to the genomic background (Via, 2012; Wu, 2001).

However, genome-wide architectures underlying ecological speciation have been identified in some systems (Keller et al., 2013; Lawniczak et al., 2010), and highly divergent genomic regions may not always correspond to ecologically relevant loci (Cruickshank & Hahn, 2014; Ravinet et al., 2017). Polygenic models of local adaptation with gene flow have also demonstrated that adaptation and speciation may be caused by coordinated action of many loci of small effect (Le Corre & Kremer, 2012; Flaxman, Wacholder, Feder, & Nosil, 2014; Yeaman, 2015). Subtle shifts in allele frequency occurring between trait-associated loci and the rest of the genome are predicted in these scenarios, and models have demonstrated the role of genome-wide linkage in enabling divergence (Flaxman et al., 2014). Identifying the extent that genomic architectures predicted in models of speciation with gene flow correspond to observed patterns poses an ongoing challenge to understanding speciation genomics (Seehausen et al., 2014).

In this study, we use restriction-associated DNA(RAD) sequencing to determine whether the genomic architecture of adaptive traits key in ecological speciation follows patterns predicted in simple or polygenic models of local adaptation. Identification of the genomic architecture underlying ecological speciation can be conducted through tests of genomic divergence, paired with quantitative genetic methods, such as genome-wide association (GWA), that identify associations between ecologically divergent traits and causal loci or loci in close physical linkage with them (Brennan et al., 2018; Stinchcombe & Hoekstra, 2008), revealing

ecologically important quantitative trait loci (QTL). Combining tests of genomic divergence and GWA conducted within species undergoing ecological speciation can help delineate whether speciation with gene flow is initiated by highly divergent genomic regions, or can instead be produced by coordinated action of polygenic loci without pronounced differences in allele frequencies, and can direct further characterization of the molecular mechanisms contributing to locally adapted traits (Barrett & Hoekstra, 2011). Additionally, trait association methods can help disentangle the role of selective, demographic and structural factors in driving genomic differentiation.

GWA methods provide the opportunity to study naturally occurring genomic variation and its association with locally adapted traits. Although GWA shows promise in identifying loci involved in local adaptation, putatively neutral genetic divergence occurring as genomes become structured within populations or ecosystems may produce false-positive associations. Association models that fail to correct for this divergence risk falsely associating traits with neutral loci differentiated following decreases in genetic exchange, requiring methods to account for genetic structuring (Vilhjálmsson & Nordborg, 2013; Voight & Pritchard, 2005). Genomic divergence produced by local adaptation poses greater difficulty in identifying true associations between traits and markers, as correction for patterns of local differentiation may also reduce signals of true association corresponding to locally adapted loci (Atwell et al., 2010; Platt, Vilhjálmsson, & Nordborg, 2010; Johnston et al., 2014; Lotterhos & Whitlock, 2015). There remains debate over appropriate methods of accommodating population structure while detecting true trait-associated loci in populations with complex divergence histories (Forester, Lasky, Wagner, & Urban, 2018; Hoban et al., 2016; Vilhjálmsson & Nordborg, 2013). In this study, we compare results from multiple proposed methods used in correcting for population structure and relatedness, using both single-locus and polygenic detection methods to identify the genomic basis of local adaptation in the marine snail *L. saxatilis*.

*L. saxatilis* is considered a model system for the study of ecological speciation because it forms pairs of locally adapted “crab” and “wave” ecotypes across shared intertidal ranges in rocky beach sites across the Northeast Atlantic (Johannesson et al., 2010; Rolán-Alvarez, Austin, & Boulding, 2015). These ecotypes show repeatable, parallel adaptation to crab predation or wave action in habitats within the intertidal zone (Butlin et al., 2014) (Boulding, Rivas, González-Lavín, Rolán-Alvarez, & Galindo, 2017; Johannesson, Johannesson, & Rolán-Alvarez, 1993). Adaptive shell traits are thought to have evolved in situ despite ongoing gene flow (Butlin et al., 2014; Galindo, Martínez-Fernández, Rodríguez-Ramilo, & Rolán-Alvarez, 2013; Kess, Galindo, & Boulding, 2018; Rolán-Alvarez et al., 2004) and have been shown to mediate partial reproductive isolation between the two ecotypes with respect to shell size (Rolán-Alvarez et al., 2015), implicating shell size divergence as a “magic trait” facilitating both local adaptation and assortative mating (Boulding et al., 2017; Galindo, Casheda, Caballero, & Rolán-Alvarez, 2019; Johannesson et al., 2010; Servedio, Doorn, Kopp, Frame, & Nosil, 2011).

Although the importance of antipredator shell traits in facilitating divergence among ecotypes has been demonstrated in *L. saxatilis* (Boulding et al., 2017), little is known about the genomic architecture underlying shell morphology (Rolán-Alvarez et al., 2015). Studying the genomic basis of shell traits can provide insight into the genomic architecture that has produced local adaptation observed among *L. saxatilis* ecotypes despite ongoing gene flow (Johannesson, Butlin, Panova, & Westram, 2018). Comparisons using a smaller set of dominant markers have found partial overlap with outlier loci, but detection of individual shell shape-associated loci using high-resolution genomic datasets within this system has remained elusive (Westram et al., 2018) despite repeated identification of extensive genome-wide divergence between ecotypes (Kess et al., 2018; Ravinet et al., 2016; Westram et al., 2016), underscoring the importance of investigating the phenotypic contributions of genomic regions of divergence with a variety of GWA methods. Here, we use a set of genome-wide SNPs from three sampling sites, spanning two glacial lineages and paired with individual shell phenotypes. This dataset allowed us to study the genomic architecture underlying shell divergence at a broad geographic scale, across the Galician *L. saxatilis* range.

In this study, we used GWA to uncover the genomic architecture underlying variation in shell morphology between *L. saxatilis* ecotypes. We address three main goals in this study: (a) geometric morphometric characterization of shell morphology variation among crab and wave ecotypes and phenotypically intermediate individuals, to identify shell shape differences reflecting divergence between ecotypes for testing in genome-wide association; (b) genome-wide association analysis of each shell trait using 4,066 SNP markers developed using double-digest restriction-associated DNA sequencing (ddRAD); and (c) testing predictions for the evolution of genomic architecture during adaptation with gene flow, through identifying whether QTL exhibit elevated linkage or correspond to highly divergent genomic regions.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection and genotyping

Genotyped samples used in this study were obtained from a previous ddRAD study of *L. saxatilis* crab ecotypes from the upper intertidal zone, wave ecotypes from the lower intertidal zone, and phenotypically intermediate samples with intermediate banding and riding shell trait combinations from the mid-shore of the intertidal zone. Samples were collected from three sites on the North and Northwest coasts of Spain (Kess et al., 2018): Silleiro (long  $-8.90^\circ$ , lat  $42.13^\circ$ ), Corrubedo (long  $-9.10^\circ$ , lat  $42.60^\circ$ ), and Burela (long  $-7.36^\circ$ , lat  $43.67^\circ$ ). Library preparation for ddRAD was carried out using the protocol described in Kess, Gross, Harper, and Boulding (2016). Using Stacks 1.40 (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011), we generated a panel of 4,066 SNP loci genotyped in a minimum of 70% of 477 sequenced individuals that also passed filtering using VCFtools 0.1.12b (Danecek et al., 2011) to

exclude loci that deviated from Hardy-Weinberg equilibrium in all population samples at a Bonferroni-corrected  $\alpha$  of 0.05/number of tested SNPs.

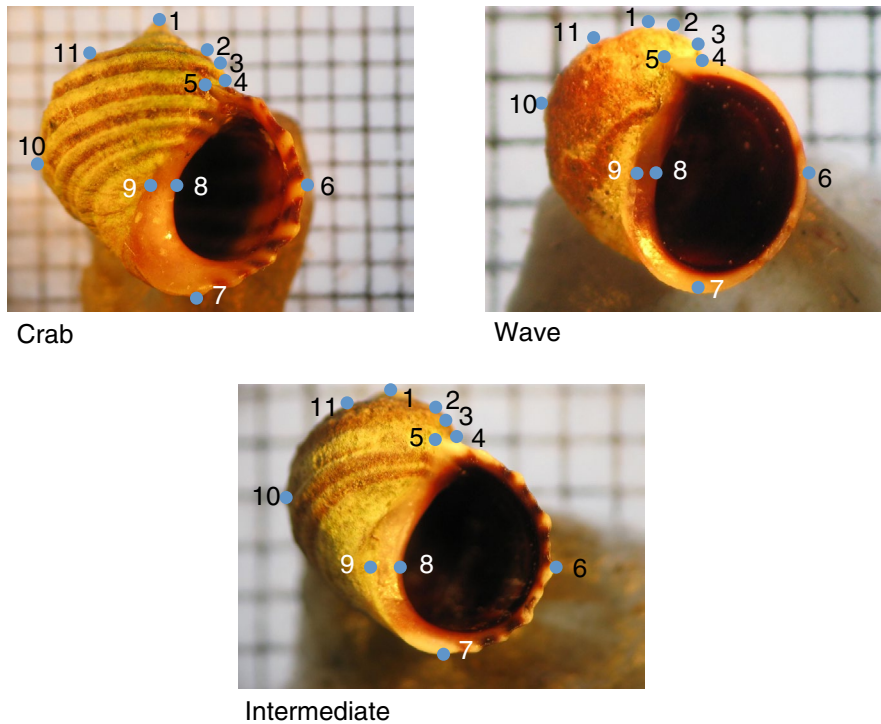
### 2.2 | Shell imaging and weight

We removed all tissue from shells during DNA extraction using forceps and 70% ethanol. A total of 574 empty shells from all collected individuals that were processed for DNA extraction, including the 477 successfully sequenced individuals, were photographed to record shell phenotypes, using a digital camera under magnification using a dissection microscope. Prior to imaging, shells were individually weighed and each weight was recorded. Shells were then positioned on a base of dental wax such that shell aperture was parallel to the camera lens, with the axis of coiling and spire on the same vertical axis in the photograph. Each shell was photographed against a transparent plastic size standard grid with 1mm x 1mm squares to enable size and magnification correction in geometric morphometric analysis.

### 2.3 | Geometric morphometric analysis

To characterize shell morphology of all sampled individuals, we used geometric morphometric landmarks developed by Carvajal-Rodríguez, Conde-Padin, and Rolán-Alvarez (2005) to score  $x$ - and  $y$ -coordinates of 11 of the original 12 landmark locations on each shell (Figure 1) and 1 mm size standards in each photograph, using tpsDig 2.17 (Rohlf, 2013). The tenth landmark described in Carvajal-Rodríguez et al. (2005) was removed due to low repeatability. Shell landmark consistency was checked by repeating scoring of one representative crab ecotype, wave ecotype, and phenotypically intermediate photograph five times by two separate technicians, followed by comparison of landmark point separation in tpsRelw 1.53 (Rohlf, 2015), and adjustment of landmark scoring criteria used across photographs prior to landmark scoring of the total sample. Variation due to shell rotation in photographs was checked by repeating photographs of the same representative ecotypes and visualizing point separation, followed by standardization of photography protocol until landmark separation within scored individuals was minimized.

Aligned landmarks, centroid size, and nonuniform components of shape variation were calculated in tpsRelw. We adjusted raw landmark  $x$  and  $y$  values for each sample using generalized Procrustes analysis to retain only shape information from landmark positions by removing variation due to size, position, and rotation (Rohlf & Slice, 1990). This operation was carried out in tpsRelw by removing variation unrelated to shape through superimposing individual landmark values onto a consensus shape, calculated by minimizing the differences in sums of squared landmark distances between the same landmarks in each sample. The consensus shape was calculated from the complete set of all photographed shells. Centroid size, used as an estimate of shell size, was calculated in tpsRelw during alignment and consensus shape estimation as the square root of the sum of the squared distances between the average  $x$ -coordinate



**FIGURE 1** Positions of 11 geometric morphometric landmarks developed by Carvajal-Rodríguez et al. (2005) on shell features of crab and wave *Littorina saxatilis* ecotypes and intermediate individuals from the mid-shore

and average y-coordinate of all landmarks for each individual snail (the centroid) and the corresponding coordinates of each of its landmarks (Bookstein, 1991). tpsRelw was also used to calculate eight partial warps corresponding to nonuniform shape changes between specimens at different spatial scales, and partial warp scores which describe the orientation and direction of shape change at each partial warp away from the reference for each individual (Zelditch, Swiderski, Sheets, & Fink, 2004). Relative warps, which characterize separate, uncorrelated major axes of shape variation within the sample, were obtained by carrying out principal component analysis on partial warp scores for each individual (Rohlf, 1993). Changes of shell shape corresponding to changes in relative warp scores for the first two principal components were visualized on deformation grids using tpsRelw.

## 2.4 | Estimation of shell morphology divergence between ecotypes

We then characterized the extent that phenotypic variation in shell traits corresponded to differences between ecotypes. We used RSTUDIO 0.99.893 (RStudio Team, 2015) using R version 3.3.3 (R Core Team, 2017) to calculate analysis of covariance (ANCOVA) models to quantify the contribution of ecotype divergence to shape variation between individuals. To distinguish between the role of local selection within ecotypes or shared evolutionary history within populations in explaining different components of shell shape, we ran separate ANCOVAs, each using one of the scores for relative warps 1 to 5 (RW1-5) as dependent variables. Interaction between the centroid size covariate was included to account for shell shape allometry previously identified in the Silleiro population studied by

Carvajal-Rodríguez et al. (2005). Variation attributable to shared selection on ecotypes, or shared differences in evolutionary history were identified by significant  $p$ -values for each of these factors, below a Bonferroni-corrected  $\alpha$  value of 0.0083.

## 2.5 | Genome-wide association of shell morphology

### 2.5.1 | Single-locus models

We used score tests implemented in the R package *GenABEL* 1.8-0 (Aulchenko, Ripke, Isaacs, & Duijn, 2007a) to conduct GWA analysis to identify genotype–trait associations for shell morphology traits across ecotypes and populations, using the panel of 4,066 SNP loci that had passed quality control. Separate GWA analyses were carried out for 11  $x$  and  $y$  aligned landmark coordinates, relative warps 1 and 2, shell weight, and centroid size. Aligned landmarks were used rather than relative warps as the former may better correspond to localized expressed transcripts from homeotic and other genes during development (Boulding et al., 2008). To control for allometry, centroid size was included as a covariate in the linear models for all shape traits. A Bonferroni-corrected  $\alpha$  of 0.05/4066 tested SNPs was used to determine whether an association with a SNP was significant in each test.

To quantify population structure for correction in GWA models, we used principal coordinate analysis of genetic distances between individuals using 2405 putatively neutral SNPs. To produce this dataset, we used a conservative approach to remove SNPs exhibiting signatures of potential selection, removing all SNPs identified as potential outliers in at least one outlier test in Kess et al. (2018), or exceeding the 95th percentile of the genome-wide  $F_{ST}$  distribution,

calculated between ecotypes in each site in plink 1.9 (Chang et al., 2015). We then identified trait associations with SNPs using two models frequently used for controlling for population structure in GWA: principal coordinate and genomic kinship matrix corrections, using combinations of *GenABEL* functions described below (Hecht, Campbell, Holecek, & Narum, 2013; Lawson & Petren, 2017; McKown et al., 2014). These models identify correlations between each SNP and each phenotypic variable, after variation due to population structure has been controlled for by inclusion of either a genomic kinship matrix or principal components or coordinates that explain a significant proportion of population structure.

We used the *ibs* function in *GenABEL* to calculate a genomic kinship matrix (K) containing pairwise identity-by-state (IBS) values estimated using all of the 2405 putatively neutral SNPs that were successfully genotyped for that pair of individuals (Aulchenko et al., 2007a). We then produced a distance matrix describing pairwise measures of genetic distance between individuals by subtracting  $0.5 - K$  from each element. We then used classic multidimensional scaling on this matrix to calculate principal coordinates of genetic distance values between individuals, using the *cmdscale* function in R. Scree plots (Cattell, 1966) were used to visualize variation explained by each principal coordinate. We then included the top three principal coordinates which each explained > 5% of variation in genomic differentiation between individuals ( $\beta_{PC1,PC2,PC3}$ ) as covariates in a linear model predicting phenotype ( $Y_i$ ) with intercept  $\mu$  and random error term,  $e$  (PC model), similar to the model used in Price et al. (2006). Associations were detected between each shell trait and each SNP using a score test carried out by the *qtscore* function in *GenABEL*, which approximates the model below:

$$Y_i = \mu + \beta_{SNP} * SNP + \beta_{PC1} * PC1 + \beta_{PC2} * PC2 + \beta_{PC3} * PC3 + \beta_{CS} * CS + e_i \tag{1}$$

We also conducted GWA using a two-step linear mixed model using the genomic kinship matrix (K) to estimate the proportion of phenotypic similarity between individuals that can be attributed to their extent of shared common ancestry. This model follows the family-based association model described in Chen and Abecasis (2007) and Aulchenko, Ripke, Isaacs, and Duijn (2007b), with genomic identity by state between individuals calculated directly from allele frequency, as in Kang et al. (2010). Individuals are presumed to have a phenotype value that is determined by population mean phenotype  $\mu$ , plus the contribution of additive genetic effects  $G_i$  and  $j$  covariates  $X$  with slope  $\beta_X$ . For the shape variables, we included centroid size as a covariate. The random additive polygenic effect  $G_i$  is distributed with variance  $K \sigma_G^2$  where  $K$  is the genomic kinship matrix calculated from 4066 SNP loci and  $\sigma_G^2$  is the additive genetic variance. In the first step, the maximum likelihood estimated trait variation in an individual explained by additive genetic contributions of shared alleles,  $\hat{e}_i$ , and the covariate(s) is estimated using the *polygenic* function, calculated using the linear mixed model:

$$Y_i = \mu + \sum_j \beta_X X + G_i + \hat{e}_i \tag{2}$$

A score test for each genotyped SNP was then performed using *mmscore* to approximate the fit of a second model. This model fits calculated residual values representing trait variation not explained by genome-wide relatedness between individuals, or included covariates, as the dependent variable, using the following equation:

$$\hat{e}_i = \mu + \beta_{SNP} * SNP + e_i \tag{3}$$

We tested the second *mmscore* model using structured association (Pritchard, Stephens, Rosenberg, & Donnelly, 2000) (K:SA) with the dependent variable being residuals calculated from the *polygenic* model in Equation (4) including only centroid size as a covariate, as shown in the equation:

$$\hat{e}_i = \mu + \beta_{SNP} (\text{population}) * SNP + e_i \tag{4}$$

The coefficients of  $\beta$  were estimated separately for all populations (Silleiro, Corrubedo, Burela) and were then combined using a weighted average with the weight being the squared inverse of the standard error of  $\beta$ .

Score test values distribute approximately as a  $\chi^2$  distribution (Chen & Abecasis, 2007), allowing SNP effect size ( $r^2$ ) calculation from one-degree-of-freedom  $\chi^2$  test statistics obtained from the score test using the following formula:

$$r^2 = \frac{\chi^2}{n - 2 + \chi^2} \tag{5}$$

where  $n$  is the number of individuals genotyped per tested SNP.

## 2.6 | Redundancy analysis

Multivariate methods have recently been utilized in identifying QTL in genotype–environment associations (Forester et al., 2018; Harrisson et al., 2017). These methods have the advantage of simultaneously accounting for multiple drivers of phenotypic or environmental divergence, and identifying polygenic trait architectures. Here, we used redundancy analysis in the R package *vegan* 2.4-5 (Oksanen et al., 2017) to identify trait-associated loci by using morphometric variables as predictors, and identify SNPs with the greatest constrained variance explained by predictors as QTL. In the RDA model, we used all traits used in single-locus GWA methods as predictors, excluding those that exhibited correlations greater than 0.7, and a matrix of complete, imputed genotypes for all 4066 SNP loci in each site as response variables. We imputed missing loci using *LinkImpute* (Money et al., 2015). Following the methodology of Forester et al. (2018), we identified QTL as those exhibiting RDA loadings greater than three standard deviations from the mean. SNP loadings were characterized for the first three canonical axes, which were selected based on visual inspection of Scree plots. Outlier SNPs were then correlated with each shell trait variable, and the strongest correlation for each SNP was used to group SNPs by traits driving their differentiation.

## 2.7 | Characterization of linkage and differentiation

During adaptation with gene flow, loci contributing to phenotypic differentiation are predicted to exhibit elevated linkage disequilibrium (LD) and divergence (Flaxman et al., 2014; Schilling et al., 2018). We used plink 1.90b3.46 (Chang et al., 2015) to assess pairwise LD between all 4066 SNPs. We then compared pairwise  $r^2$  among identified QTL with putatively neutral SNPs that were not significantly associated with shell traits, did not exhibit elevated differentiation based on  $F_{ST}$  values greater than the 95% quantile as calculated in plink, or were not previously identified as  $F_{ST}$  outliers using formal outlier tests in Kess et al. (2018). We compared  $r^2$  for each class of SNP within high and low shore locations in each site and between ecotypes in each site, as in Schilling et al. (2018), and used a Mann-Whitney U test on  $r^2$  values for QTL and putatively neutral SNPs to assess the significance of LD differences.

To directly test the genomic distribution of QTL, we aligned our RAD loci containing QTL to the recently released draft genome from the Swedish crab ecotype of *L. saxatilis* (Westram et al., 2018). We extracted RAD loci with QTL from the Stacks consensus sequences using the *filterbyname* scripts in BMap (Bushnell, 2016). Alignments were carried out with Burrows-Wheeler Aligner, using the *bwa mem* algorithm (Li, 2013). We then binned aligned contigs to their corresponding linkage group information (Westram et al., 2018). To test whether QTL exhibit nonrandom distribution among linkage groups, the proportion of QTL aligning to each linkage group was compared using a  $\chi^2$  test.

To identify whether trait-associated loci demonstrate strong signals of differentiation potentially due to divergent selection, we compared identified QTL with genomic regions exhibiting elevated differentiation based on  $F_{ST}$  values greater than the 95% quantile as calculated in plink, or previously identified as  $F_{ST}$  outliers using formal outlier tests in Kess et al. (2018). This significance threshold was selected to ensure all divergent loci or those significantly associated with population structure were included for comparison with trait-associated SNPs. Following these analyses, we quantified the number of trait-associated loci also identified as divergent loci.

## 3 | RESULTS

### 3.1 | Geometric morphometric analysis and sources of shape variation

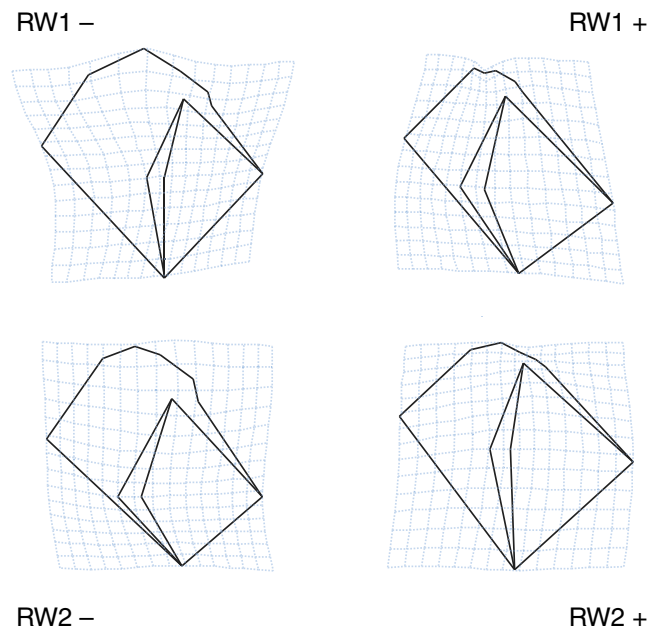
We found that the majority (62.79%) of shape variation was explained by the first two relative warps which reflected differences in shell aperture and spire height. The first relative warp accounted for 43.16% (RW1) and the second accounted for 19.63% (RW2) of shell shape variation. We found five relative warps that individually explained greater than 5% of shell shape variation and cumulatively explained 84.30% of variation between individuals, of which RW3, RW4, and RW5 each respectively explained 8.79, 7.8, and 5.31% of total shape variation. Visualizing changes in shell shape with changing relative warp values using deformation grids revealed increases in RW1 values corresponding with increase in shell aperture size and reduction

in spire height (Figure 2). We found a similar pattern of shape change for RW2, with increases in RW2 value commensurate with reduction in upper whorl size, and upward expansion of the shell aperture.

We found that interactive effects between ecotype and population factors were observed in all traits excluding RW3, indicating nonshared, population-specific differences in ecotype shell shape (Table 1). Interaction between these factors and centroid size also made a significant contribution to explaining variation in the first two relative warps, indicating population-specific allometric relationships for these shape variables, and suggesting ecotype-specific developmental trajectories of shape change vary between populations.

### 3.2 | Genome-wide association analysis and genomic architecture of shell morphology

We observed polygenic bases of shell trait variation through identification of 216 QTL in genome-wide association tests performed using single-locus models in *GenABEL* and polygenic association using RDA, though associated locus identity varied depending on the method used to correct for population structure. We identified 97 QTL in the K:SA model, 57 in the PC model, and 82 using RDA (58 from RDA1, 18 from RDA2, 6 from RDA3). A small proportion of associations were common across pairwise comparisons between models: 15 QTL (0.108) were shared between the PCA and K:SA models, and four QTL (0.023) were shared between the RDA and K:SA models. However, only one QTL was shared between the RDA and PC models, and no QTL were shared among all three comparisons. The different



**FIGURE 2** Deformation grids for relative warp 1 (RW1) and relative warp 2 (RW2) values calculated from 11 geometric morphometric landmarks, depicting shell shape change with change of relative warp value, and ecotype labels identifying shape change specific to each ecotype

**TABLE 1** Significant factors (ecotype [E] and population [P]), covariate (centroid size [C]), and interaction effects in order of significance, and adjusted  $R^2$  values associated with measures of shell shape in ANCOVA of relative warp and centroid size obtained from geometric morphometric analysis

Trait	Significant factors, covariates, and interactions	$R^2$
RW1	E × P × C ( $p = 0.0008$ )	0.699
RW2	E × P × C ( $p = 8.8 \times 10^{-6}$ )	0.3147
RW3	E ( $p = 6.2 \times 10^{-10}$ ), C ( $p = 1.35 \times 10^{-5}$ ), P ( $p = 0.008$ )	0.1388
RW4	E × P ( $p = 0.0001$ )	0.1001
RW5	E × P × C ( $p = 0.007$ )	0.1139
Centroid	E × P ( $p = 0.0003$ )	0.6982

methods of population structure correction used in each method likely account for these differences, as we found that individuals grouped by RW1 scores also clustered by PC3, whereas population differences were explained by PC1 and PC2 (Figure 3).

Consistent with a polygenic basis of shell variation, we found that the majority of identified loci in single-locus models exhibited small effect sizes (~5%) for each trait (Table 2). However, a subset of large-effect alleles were observed associated with X7, X9, and Y2 landmarks, RW1, all exceeding 10% variation explained; a single locus identified in the K:SA model explained 24% of centroid size variation.

We found that the majority of SNPs identified using RDA exhibited strongest association with centroid size ( $n = 59$  SNPs), and landmarks Y3 ( $n = 10$  SNPs) and X7 ( $n = 5$  SNPs), supporting a role for aperture shape differences and shell shape facilitating divergence between ecotypes. In contrast to the PC model, we observed separation of ecotypes followed by glacial lineages on the first two RDA axes when constrained by shell morphology, indicating shell morphology variation plays a role in genomic differentiation between ecotypes across sites (Figure 4).

### 3.3 | Linkage disequilibrium

We observed elevation of statistical LD among QTL, but identified a genome-wide distribution of QTL overall, indicating genome-wide selection on shell shape. We found that pairwise LD among all SNPs was low even between ecotypes (mean  $r^2$  BR = 0.023, CO = 0.026, SI = 0.022; Figure 5) in each site, indicating low physical linkage of SNPs used in this study. To test for evidence of long-distance linkage between trait-associated loci, we compared levels of LD ( $r^2$ ) within sampling locations and within ecotypes within locations (Table 3). We found that QTL exhibited significantly elevated LD within locations compared to the genome-wide average for putatively neutral SNPs. Similarly, we found that within ecotypes within all sites, at least one ecotype exhibited signatures of long-distance LD of QTL, indicated by slightly elevated  $r^2$  values and shifted  $r^2$  distributions of QTL relative to putatively neutral loci (Table 3).

Alignment of RAD loci containing QTL to the *L. saxatilis* genome revealed that these loci were distributed across 196 contigs; 61 RAD loci aligned to contigs with linkage information. These 61 loci were found across 17 linkage groups, suggesting low physical linkage of QTL. Consistent with low physical LD between QTL, we did not find higher density of QTL across any aligned linkage group ( $\chi^2 = 14.66$ ,  $df = 16$ ,  $p = 0.5497$ ).

### 3.4 | Differentiation of QTL between ecotypes

Consistent with predictions from models of divergence with gene flow, we found partial overlap of divergent SNPs and QTL identified in this study.  $F_{ST}$  values of QTL were elevated relative to the genome-wide distribution for all loci (Figure 6). Of the total set of identified QTL, 64 (0.296) exhibited overlap with loci in divergent genomic regions. This pattern of elevated divergence of QTL was driven by loci identified using RDA; 53 (0.646) QTL identified in RDA exhibited overlap with divergent regions, whereas only 11 (0.1134) QTL identified in the K:SA model overlapped with divergent regions, and only 3 QTL (0.052) identified in the PC model exhibited elevated divergence.

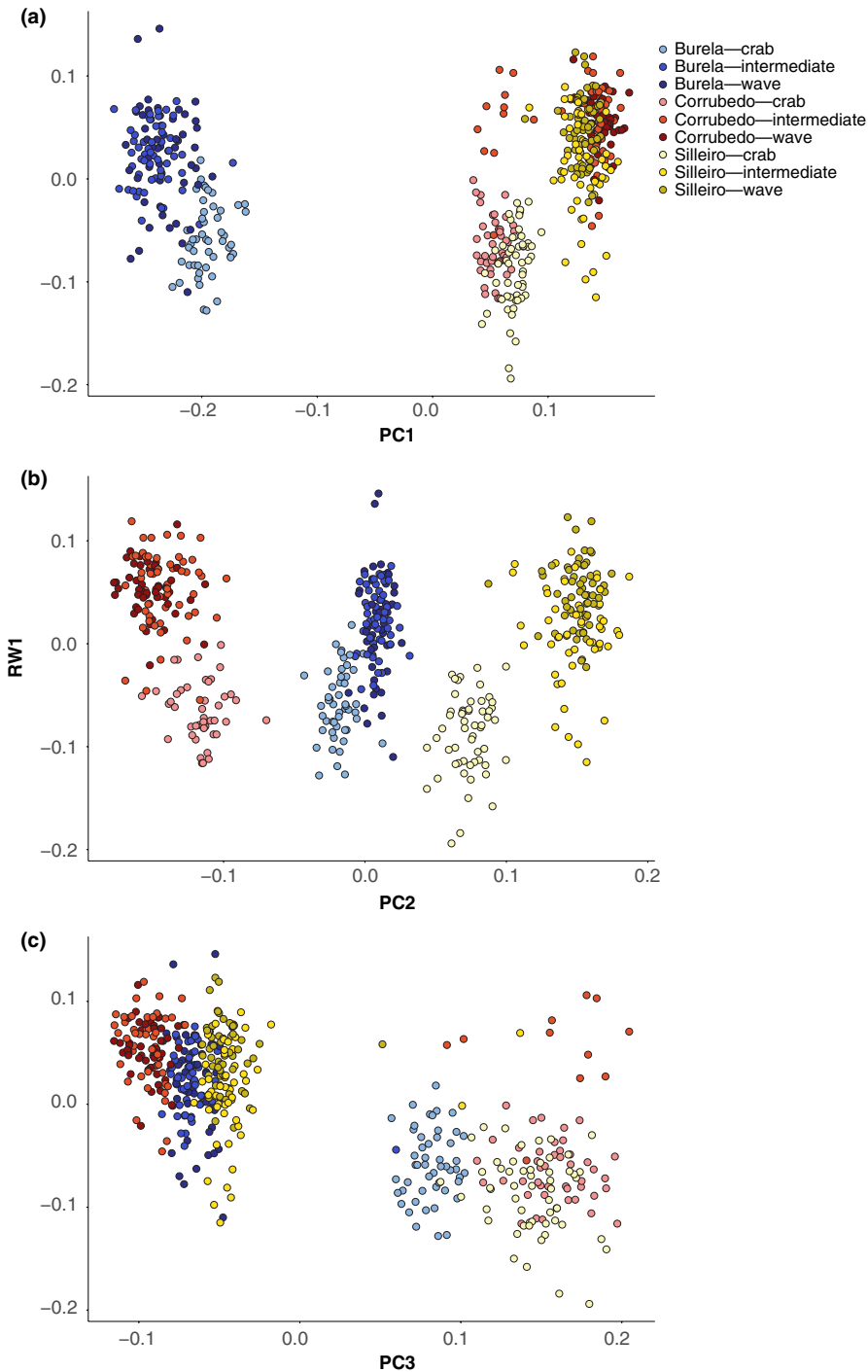
## 4 | DISCUSSION

### 4.1 | Polygenic size and shape divergence

Uncertainty remains over the extent that observed genomic architectures match those predicted to evolve during speciation with gene flow (Seehausen et al., 2014). Here, we used genome-wide association of SNP loci with shell shape variation observed between *L. saxatilis* ecotypes to identify a polygenic, genome-wide basis of ecotype divergence. Our dataset, sampled from three sites spanning two glacial lineages, enabled investigation of the genomic architecture underlying shell divergence between ecotypes across the geographic range of Galician *L. saxatilis* crab and wave ecotypes. We find evidence of: (a) polygenic architecture of differentiated shell shape and size traits, (b) elevated linkage of QTL associated with shell variation despite a genome-wide distribution, and (c) partial overlap of divergent  $F_{ST}$  outlier regions associated with shape variation detected during genome-wide association, suggesting increased rates of divergence at QTL underlying shell variation. Together these results from an empirical study of an emerging model system (Ravinet et al., 2017; Galindo & Grahame, 2014) present a complex illustration of genomic architecture evolution during ecological speciation, revealing variable patterns of allelic differentiation and genome-wide linkage during adaptation.

### 4.2 | Divergence of trait-associated loci

Though initially thought to house genes integral to speciation with gene flow, recent debate has arisen over alternative evolutionary and demographic mechanisms that also explain high genomic



**FIGURE 3** Plot of the first three principal coordinates of genetic distance between individuals from 2405 putatively neutral SNP loci, (a) PC1, (b) PC2, and (c) PC3, and relative warp 1 (RW1) describing shell shape differences calculated from 11 geometric morphometric landmarks

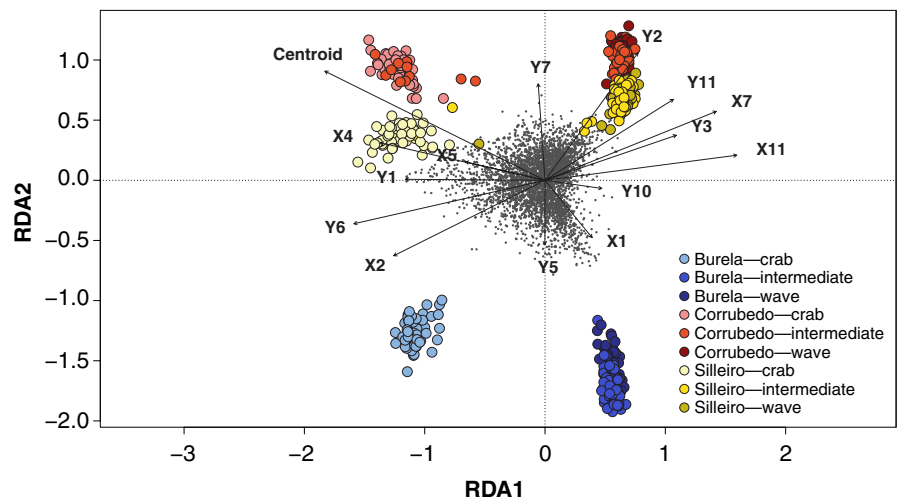
divergence (Cruickshank & Hahn, 2014; Ravinet et al., 2017). By using GWA tests for phenotypic traits with known roles in adaptive divergence between *L. saxatilis* ecotypes, we find evidence suggesting loci associated with divergence in shape between ecotypes show partial overlap with divergent genomic regions with elevated  $F_{ST}$  identified in Kess et al. (2018). These results indicate that selection on divergent shell morphology has likely contributed to genomic divergence between ecotypes in these regions, consistent with the hypothesis of genome-wide divergence driven by ecological divergence (Via, 2012; Wu, 2001). These findings also

reinforce the hypothesis that divergent selection rather than phenotypic plasticity underlies divergent shell morphology in *L. saxatilis* ecotypes inhabiting different intertidal regions (Hollander & Butlin, 2010). Our results differ from the findings of another recent genome-wide association study of shell morphology conducted in Swedish *L. saxatilis* ecotypes, in which no individual significant loci at the genome-wide level were identified (Westram et al., 2018). However, partitioning of morphological variance across chromosomes in the study by Westram et al. (2018) did uncover substantial phenotypic variation overrepresented in regions associated with



**TABLE 2** Number of SNP loci significantly associated with shell morphology traits and mean phenotypic variation explained ( $r^2$ ) per SNP in a linear model with the first three principal coordinates of genotypic variation (PC model), and structured association conducted with a mixed linear model adjusted for population structure using the genomic kinship matrix (K:SA model), or exhibiting highest correlation with a shell morphology trait in redundancy analysis (RDA)

Trait	Associated SNPs		Max percentage phenotypic variation explained ( $r^2$ ) by associated SNP		Median percentage phenotypic variation explained ( $r^2$ ) / associated SNP		SNPs exhibiting highest correlation with trait identified in RDA
	K:SA model	PC model	K:SA model	PC model	K:SA model	PC model	
X1	0	0	0	0	0	0	2
X2	4	0	5.9	0	5.1	0	0
X5	0	0	0	0	0	0	1
X7	23	21	10.2	8.0	5.9	4.9	5
X8	16	2	8.4	4.8	4.6	4.4	0
X9	30	26	15.1	8.8	6.0	5.5	0
X10	20	0	9.3	0	6.0	0	0
X11	6	0	6.6	0	5.5	0	0
Y1	0	4	0	5.9	0	5.1	0
Y2	17	10	10.7	5.3	5.8	4.9	1
Y3	0	0	0	0	0	0	10
Y4	1	9	4.9	9.0	4.9	4.8	0
Y5	0	3	0	8.2	0	5.4	0
Y6	9	0	8.0	0	5.5	0	2
Y7	0	7	0	5.1	0	4.7	0
Y10	0	0	0	0	0	0	2
Y11	8	0	8.2	0	6.6	0	0
RW1	20	0	11.1	0	6.4	0	0
RW2	3	15	6.4	7.9	5.0	5.2	0
Centroid	78	31	24.7	8.0	6.2	5.2	59

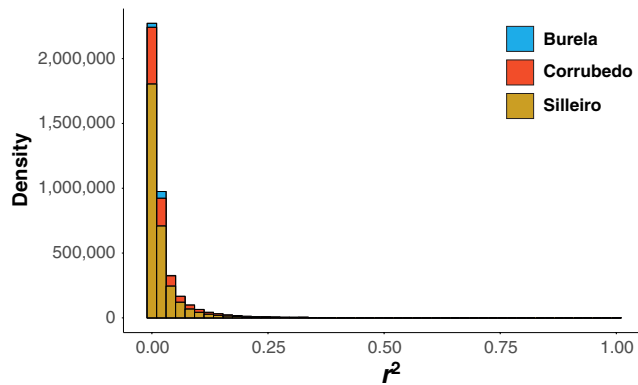


**FIGURE 4** Plot of individual scores on the first two canonical axes (RDA1 and RDA2) obtained from a constrained ordination of 4,066 SNP loci with 15 shell trait variables

structural variation (inversions) exhibiting elevated differentiation within the genome, consistent with our findings of differentiated QTL in the current study. These differences may arise due to methods of identifying associations between phenotypic and genomic variation, discussed below.

### 4.3 | Genome-wide association in populations with multiple sources of structure

We found more highly divergent SNPs underlying shell morphology divergence using polygenic RDA compared to single-locus GWA



**FIGURE 5** Distribution of linkage disequilibrium measured as pairwise  $r^2$  between 4066 SNP loci, calculated separately in samples from Burela, Corrubedo, and Silleiro

models, which we attribute to differences in population structure corrections used in each method. Detection of trait-associated loci using linear models that include principal components or a kinship matrix as covariates of population structure may be conservative when traits analyzed are strongly correlated with population structure (Atwell et al., 2010; Frichot, Schoville, Bouchard, & Francois, 2013). This problem may be more pronounced in scenarios of isolation-by-adaptation, in which local adaptation also reduces gene flow from other populations (Nosil, Egan, & Funk, 2008; Lotterhos & Whitlock, 2015).

Recent applications of multivariate ordination methods such as redundancy analysis have shown promise in finding multilocus adaptation while accounting for both population structure and polygenic interactions among loci (Capblancq, Luu, Blum, & Bazin, 2018; Forester et al., 2018). These methods have recently been shown to outperform mixed-model-based methods, as well as Random Forest,

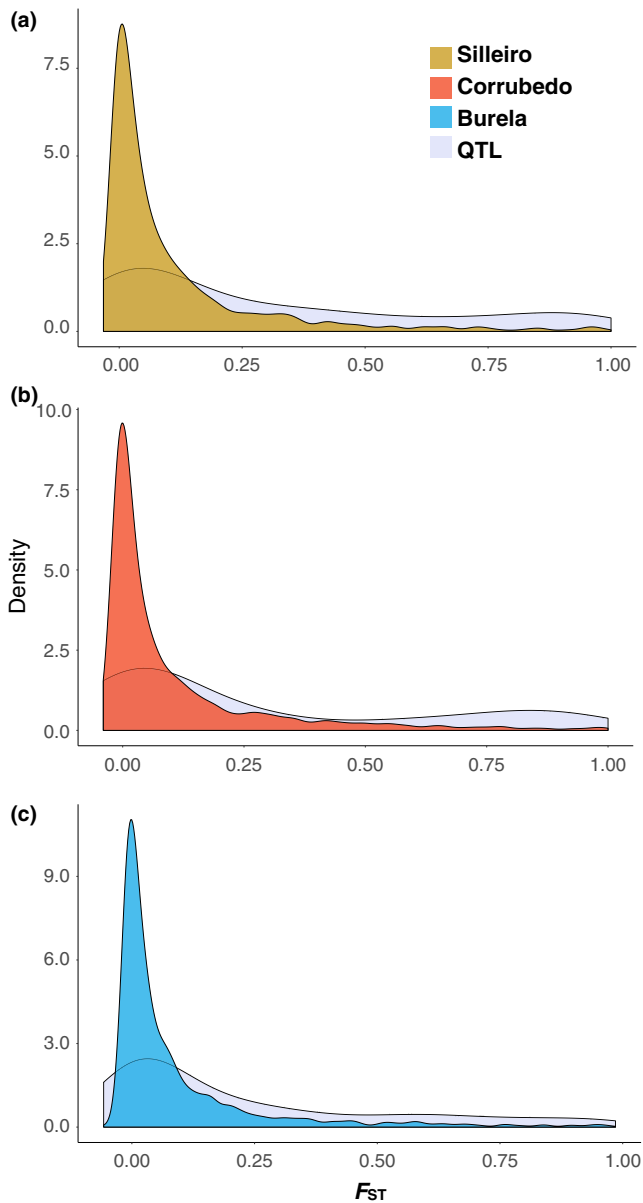
a machine-learning-based method, in uncovering loci associated with environmental variation (Capblancq et al., 2018; Forester et al., 2018). We acknowledge that although RDA has performed better in simulation studies and appeared here to better match expectations from theory, the identification of a larger number of highly divergent trait-associated loci does not guarantee that all these associations are true. Additionally, it is also possible that conservative correction for highly divergent loci in our single-locus models also enabled the detection of polygenic loci that exhibit small shifts in allele frequency. Recent clinal analysis of loci exhibiting signatures of selection across Swedish ecotypes of *L. saxatilis* also exhibited variation in levels of differentiation, suggesting that there is a continuum of divergence observed among polygenic loci in this system (Westram et al., 2018). All significant loci should be treated only as hypotheses of trait association to be further investigated through gene annotation and crosses (Atwell et al., 2010; Savolainen, Lascoux, & Merilä, 2013).

#### 4.4 | Genomic architecture of trait-associated loci

We find evidence suggesting that *L. saxatilis* shell shape exhibits linked, polygenic architecture predicted during trait divergence driven by genome-wide differentiation (Feder, Egan, & Nosil, 2012), evidenced by identification of many candidate QTL with elevated LD. Observation of polygenic architecture is consistent with the previously identified quantitative genetic inheritance of shell shape in another *Littorina* species using half-siblings reared at two different growth rates (Boulding & Hay, 1993). In comparing predicted genomic architectures expected under divergence with gene flow, observation of polygenic control of shape traits appears consistent with a scenario of multifarious selection resulting in the formation of statistically linked complexes of adaptive loci (Le Corre & Kremer,

Site	Comparison	Putatively neutral $r^2$	QTL $r^2$	Mann-Whitney U test
SI	HL	0.016	0.102	$W = 8,254,500,000$ , $p < 2.2 \times 10^{-16}$
	H	0.0253	0.0286	$W = 12,113,000,000$ , $p < 2.2 \times 10^{-16}$
	L	0.0215	0.0355	$W = 5,243,000,000$ , $p = 0.704$
CO	HL	0.0193	0.102	$W = 13,481,000,000$ , $p < 2.2 \times 10^{-16}$
	H	0.0311	0.0314	$W = 11,612,000,000$ , $p = 0.0213$
	L	0.0288	0.0284	$W = 551,200,000$ , $p = 7.852 \times 10^{-8}$
BR	HL	0.0185	0.0707	$W = 13,542,000,000$ , $p = 8.47 \times 10^{-9}$
	H	0.031	0.0322	$W = 8,843,600,000$ , $p = 0.00384$
	L	0.0313	0.0302	$W = 5,272,200,000$ , $p = 0.056$

**TABLE 3** Linkage relationships of neutral loci and QTL in within-site and within-site-and-ecotype calculation of pairwise  $r^2$  per SNP, and pairwise significance of compared SNPs evaluated using a Mann-Whitney U test



**FIGURE 6** Comparison of  $F_{ST}$  distributions between 216 QTL and all 4066 SNPs in (a) Burela, (b) Corrubedo, and (c) Silleiro

2012; Flaxman et al., 2014). Consistent with this hypothesis, we observed elevated statistical linkage of QTL associated with shell morphology and exceeding levels observed in the neutral genome, indicating coupling of QTL both within and between ecotypes within each population. This pattern of coupling between loci associated with divergence is also predicted during adaptation and speciation with gene flow, resulting in elevated genome-wide linkage during the buildup of reproductive barriers (Schilling et al., 2018). Buildup of genome-wide LD during adaptive divergence has previously been identified in another model system for ecological speciation, threespine stickleback (Hohenlohe, Bassham, Currey, & Cresko, 2012). In models specific to *littorinid* evolution, polygenic clines in shell thickness along a steep predation gradient were facilitated by positive LD among physically unlinked loci of small effect (Boulding et al., 2007).

We observe elevated LD specifically at QTL, whereas the remainder of SNPs did not exhibit high LD between ecotypes (Table 3), suggesting that infrequent gene flow may still occur, despite a high level of overall neutral differentiation (Kess et al., 2018; Rolán-Alvarez et al., 2004). We also uncovered more subtle patterns of within-ecotype coupling of QTL, as predicted during evolution of locally adapted phenotypes (Schilling et al., 2018). We observed this pattern within at least one ecotype for each sampling location. Similar coupling of divergent loci within demes has also been identified across sympatric *Heliconius* host races and has been suggested to be important in enabling speciation (Schilling et al., 2018).

We did not find evidence of elevated physical linkage among QTL, as nearly all of these loci aligned to independent contigs across 17 linkage groups. Our finding is inconsistent with the recent identification of polymorphic inversions associated with formation of ecotypes across ecological gradients in *L. saxatilis* ecotypes (Faria, Johannesson, Butlin, & Westram, 2019; Morales et al., 2018), suggesting that these inversions may not be integral to shell shape divergence between ecotypes. Consistent with this observation, comparisons of whole-genome divergence between pools of ecotypes have revealed a polygenic basis of ecotype divergence, although inversion regions did demonstrate elevated frequency of highly differentiated polymorphisms (Morales et al., 2018). The sparseness of RAD SNP loci ( $N = 4,066$ ) in our dataset may have given us low coverage in inversion regions. Coupled with the level of fragmentation of the currently available reference genome, this limitation may explain why we did not find physical linkage among our QTL. The type II inversion hypothesis that balancing selection could maintain heterokaryotypes (Faria et al., 2019) should be further explored with whole-genome resequencing data from individual snails matched with measurements of each individual's shells.

Also consistent with polygenic divergence of ecotype differentiation, we found predominantly loci of small effect contributing to shell morphology variation, but also identified a subset of large-effect alleles in shell traits corresponding to ecotype survival and assortative mating. These larger effect size alleles (>10% phenotypic variation explained) were found in association with landmark locations related to shell aperture differences, consistent with the identification of variation in these traits between crab and wave ecotypes in the present study, as well as in past comparisons (Butlin et al., 2014; Carvajal-Rodríguez et al., 2005). Large-effect loci explaining more than 10% of the phenotypic variance in size were also observed. Size has traditionally been considered to possess a polygenic genetic architecture in other systems, though variation in effect size between associated loci has also been observed (Berenos et al., 2015; Visscher et al., 2007). Shell size has also been considered a “magic” trait in *L. saxatilis*, facilitating both reproductive isolation and adaptation to local predation and small biogenic refuges, respectively (Boulding et al., 2017; Servedio et al., 2011). Larger effect size of alleles underlying these traits may suggest stronger selection on these shell shape features, resulting in stronger selection for gene-flow-resistant trait architectures as well (Griswold, 2006; Albert et

al., 2008). Given the role of shell size in controlling both local ecotype survival and assortative mating, identification of alleles of larger effect is consistent with models that predict large-effect alleles incorporated into genetic architectures for traits underlying local adaptation with gene flow (Griswold, 2006; Yeaman & Whitlock, 2011).

Despite the discovery of loci of both large and small effect, sample size and genomic sampling method may both play a role in biasing estimates of allele number and effect size uncovered in this study. Our small sample sizes in both individuals and markers relative to massive human GWA studies may have prevented detecting the small contributions made to shell variation by many alleles dispersed across the genome (Boyle, Yang, & Pritchard, 2017) and may also bias effect size estimates of alleles detected (Otto & Jones, 2000). Additionally, RADseq samples a small subset of total genomic variation, potentially missing regions important to local adaptation (Lowry et al., 2017), although empirical results have largely supported the utility of RADseq in finding loci important to local adaptation (Catchen et al., 2017; McKinney, Larson, Seeb, & Seeb, 2017). Here, predicted genome size of *L. saxatilis* is 1.32 Gb (Vitturi, Libertini, Panozzo, & Mezzapelle, 1995), and we identify many QTL (216), from a small proportion of the surveyed genome (~0.01%). The nonrandom distribution of RAD loci within the genome (DaCosta & Sorenson, 2014), our selection of restriction enzymes that do not cut within known repetitive regions, and our selection of a quantitative trait associated with ecotype divergence may account for this observation. Our identification of ~5.3% loci exhibiting trait association is similar to the proportion of divergent loci identified across other ecological speciation systems (Nosil et al., 2009) and, although not all QTL were identified as divergent outliers, is likely influenced by our selection of a trait crucial in ecotype divergence. Additionally, elevated LD between ecotypes and specifically at trait-associated loci likely enabled detection of many QTL despite modest genomic coverage. Future studies that both employ larger samples and utilize whole-genome sequences to sample the entirety of genomic variation provide a way to resolve this limitation.

## 5 | CONCLUSIONS AND IMPLICATIONS

Our work here, using a genome-wide SNP dataset and individual shell traits, paired with sampling across two glacial lineages, allowed us to uncover the genomic architecture of shell divergence in *L. saxatilis* crab and wave ecotypes across their range in north-west Spain. We used multiple models to discover the SNP markers associated with shell traits reflecting adaptations to differences in crab predation and wave action between regions of the intertidal zone. We detected polygenic trait architectures that overlapped regions of high genomic divergence as identified previously using  $F_{ST}$  outlier tests. We also identify elevated statistical linkage of identified genomic regions associated with shell traits despite no evidence of physical linkage, indicating that coordination of polygenic loci underlies ecotype formation in *L. saxatilis*. Future research on this system would benefit from the availability of a less-fragmented

genome that would improve the identification of physical locations of QTL. It would also be interesting to associate gene expression data (e.g., Rivas et al., 2018) with the genome to facilitate identification of the causal trait-associated SNP loci identified in this study.

## ACKNOWLEDGMENTS

Funding for snail collection, DNA size selection, DNA library preparation, technical assistance, and partial support for T.K. was from an NSERC Discovery Grant to EGB. Partial support for T.K. was from an Ontario Graduate Scholarship and an Integrative Biology PhD Award. Funding for DNA sequencing was provided by the Ministerio de Economía y Competitividad (CGL2016-75482-P) to E. Rolán-Alvarez. We especially thank J. Galindo and M.J. Rivas for collecting snails; J. Wood, M. Garcia, W. Jarvis, C. Moorhouse, and J. Goldstein for digitizing snail images and assisting with photography; and A. Heyland for use of camera equipment. We also thank K. Cottenie, M. Ferguson, R. Gregory, M. Holborn, K. Johannesson, S. Kremer, M. MacLeod-Bigley, and E. Rolán-Alvarez for comments on the manuscript.

## CONFLICT OF INTEREST

None declared.

## AUTHOR CONTRIBUTIONS

T.K. designed the project, conducted laboratory work and statistical analysis, and wrote the manuscript. E.G.B. helped design the project, supervised the laboratory work, assisted with interpretation of the statistical analysis, and helped write and edit the manuscript.

## DATA AVAILABILITY STATEMENT

Plink input files and shell phenotype data used for GWAS have been uploaded to Dryad (<https://doi.org/10.5061/dryad.p3123d8>). Reads will be uploaded to the NCBI Short Read Archive (SUB4134175).

## ORCID

Tony Kess  <https://orcid.org/0000-0002-1079-3791>

Elizabeth G. Boulding  <https://orcid.org/0000-0002-6852-7728>

## REFERENCES

- Albert, A. Y. K., Sawaya, S., Vines, T. H., Knecht, A. K., Miller, C. T., Summers, B. R., ... Schluter, D. (2008). The genetics of adaptive shape shift in stickleback: Pleiotropy and effect size. *Evolution*, *62*, 76–85.
- Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., ... Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, *465*, 627–631. <https://doi.org/10.1038/nature08800>

- Aulchenko, Y. S., Ripke, S., Isaacs, A., & van Duijn, C. M. (2007a). GenABEL: An R package for genome-wide association analysis. *Bioinformatics*, 23(10), 1294–1296.
- Aulchenko, Y., Ripke, S., Isaacs, A., & van Duijn, C. (2007b). Genomewide rapid association using mixed model and regression: A fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*, 177, 577–585. <https://doi.org/10.1534/genetics.107.075614>
- Barrett, R. D. H., & Hoekstra, H. (2011). Molecular spandrels: Tests of adaptation at the genetic level. *Nature Reviews Genetics*, 12, 767–780. <https://doi.org/10.1038/nrg3015>
- Berenos, C., Ellis, P. A., Pilkington, J. G., Lee, S. H., Gratton, J., & Pemberton, J. M. (2015). Heterogeneity of genetic architecture of body size traits in a free-living population. *Molecular Ecology*, 24, 1810–1830. <https://doi.org/10.1111/mec.13146>
- Bookstein, F. L. (1991). *Morphometric tools for landmark data*. New York, NY: Cambridge University Press.
- Boulding, E. G., Culling, M., Glebe, B., Berg, P. R., Lien, S., & Moen, T. (2008). Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses. *Heredity*, 101(4), 381–391. <https://doi.org/10.1038/hdy.2008.67>
- Boulding, E. G., & Hay, T. K. (1993). Quantitative genetics of shell form of an intertidal snail: Constraints on short-term response to selection. *Evolution*, 47, 576–592.
- Boulding, E. G., Hay, T., Holst, M., Kamel, S., Pakes, D., & Tie, A. D. (2007). Modelling the genetics and demography of step cline formation: Gastropod populations preyed on by experimentally introduced crabs. *Journal of Evolutionary Biology*, 20(5), 1976–1987. <https://doi.org/10.1111/j.1420-9101.2007.01369.x>
- Boulding, E. G., Rivas, M. J., González-Lavín, N., Rolán-Alvarez, E., & Galindo, J. (2017). Size selection by a gape-limited predator of a marine snail: Insights into magic traits for speciation. *Ecology and Evolution*, 7(2), 674–688. <https://doi.org/10.1002/ece3.2659>
- Boyle, E. A., Yang, I. L., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169, 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- Brennan, R., Healy, T. M., Bryant, H. J., La, M. V., Schulte, P., & Whitehead, A. (2018). Integrative population and physiological genomics reveals mechanisms of adaptation in killifish. *Molecular Biology and Evolution*, 35(11), 2639–2653. <https://doi.org/10.1093/molbev/msy154>
- Bushnell, B. (2016) BMAP short read aligner, University of California, Berkeley, California. Retrieved from <http://sourceforge.net/projects/bbmap/>
- Butlin, R. K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., ... Rolán-Alvarez, E. (2014). Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution*, 68, 935–949. <https://doi.org/10.1111/evo.12329>
- Capblancq, T., Luu, K., Blum, M. G., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, 18, 1223–1233. <https://doi.org/10.1111/1755-0998.12906>
- Carvajal-Rodríguez, A., Conde-Padin, P., & Rolán-Alvarez, E. (2005). Decomposing shell form into size and shape by geometric morphometric methods in two sympatric ecotypes of *Littorina saxatilis*. *Journal of Molluscan Studies*, 71, 313–318. <https://doi.org/10.1093/mollus/eyi037>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1, 171–182.
- Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, 38, 42–49. <https://doi.org/10.1111/1755-0998.12669>
- Cattell, R. B. (1966). The scree test for the number of factors. *Journal of Multivariate Behavioral Research*, 1, 245–276. [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10)
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7. <https://doi.org/10.1186/s13742-015-0047-8>
- Chen, W. M., & Abecasis, G. R. (2007). Family-based association tests for genomewide association scans. *The American Journal of Human Genetics*, 81, 913–926. <https://doi.org/10.1086/521580>
- Coyne, J., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer.
- Cruikshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23, 3133–3157. <https://doi.org/10.1111/mec.12796>
- DaCosta, J. M., & Sorenson, M. D. (2014). Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS ONE*, 9, e106713.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., De Pisto, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., ... Westram, A. M. (2019). Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Molecular Ecology*, 28, 1375–1393. <https://onlinelibrary.wiley.com/doi/epdf/10.1111/mec.14972>
- Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in Ecology & Evolution*, 34(3), 239–248. <https://doi.org/10.1016/j.tree.2018.12.005>
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28, 342–350. <https://doi.org/10.1016/j.tig.2012.03.009>
- Flaxman, S. M., Wacholder, A. C., Feder, J. L., & Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, 23, 4074–4088. <https://doi.org/10.1111/mec.12750>
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, 27, 2215–2233. <https://doi.org/10.1111/mec.14584>
- Frichot, E., Schoville, S. D., Bouchard, G., & Francois, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, 30, 1687–1699. <https://doi.org/10.1093/molbev/mst063>
- 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., & Grahame, J. W. (2014). Ecological speciation and the intertidal snail *Littorina saxatilis*. *Advances in Ecology*, 1–9, Article ID 239251.
- 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, 1472–1487. <https://doi.org/10.1111/jeb.12152>
- Griswold, C. K. (2006). Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity*, 96, 445–453. <https://doi.org/10.1038/sj.hdy.6800822>
- Harrisson, K. A., Amish, S. J., Pavlova, A., Narum, S. R., Telonis-Scott, M., Rourke, M. L., ... Sunnucks, P. (2017). Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Molecular Ecology*, 26, 6253–6269. <https://doi.org/10.1111/mec.14368>

- Hecht, B. C., Campbell, N. R., Holecek, D. E., & Narum, S. R. (2013). Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, 22, 3061–3076. <https://doi.org/10.1111/mec.12082>
- Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., ... Whitlock, M. C. (2016). Finding the genomic basis of local adaptation in non-model organisms: Pitfalls, practical solutions and future directions. *The American Naturalist*, 188, 379–397.
- Hohenlohe, P. A., Bassham, S., Currey, M., & Cresko, W. A. (2012). Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 367, 395–408. <https://doi.org/10.1098/rstb.2011.0245>
- Hollander, J., & Butlin, R. K. (2010). The adaptive value of phenotypic plasticity in two ecotypes of a marine gastropod. *BMC Evolutionary Biology*, 10, 333. <https://doi.org/10.1186/1471-2148-10-333>
- Johannesson, K., Butlin, R. K., Panova, M., & Westram, A. M. (2018). Mechanisms of adaptive divergence and speciation in *Littorina saxatilis*: Integrating knowledge from ecology and genetics with new data emerging from genomic studies. In O. P. Rajora (Ed.), *Population Genomics: Concepts, Strategies, and Applications* (pp. 1–25). Cham, Switzerland: Springer.
- Johannesson, K., Johannesson, B., & Rolán-Alvarez, E. (1993). Morphological differentiation and genetic cohesiveness over a microenvironmental gradient in the marine snail *Littorina saxatilis*. *Evolution*, 47, 1770–1787.
- Johannesson, K., Panova, M., Kemppainen, P., André, C., Rolán-Alvarez, E., & Butlin, R. (2010). Repeated evolution of reproductive isolation in a marine snail: Unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 1735–1747. <https://doi.org/10.1098/rstb.2009.0256>
- Johnston, S. E., Orell, P., Pritchard, V. L., Orell, P., Pritchard, V. L., Kent, M. P., ... Primmer, C. R. (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, 23, 3452–3468.
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., ... Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42, 348–354. <https://doi.org/10.1038/ng.548>
- Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., ... Seehausen, O. (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, 22, 2848–2863. <https://doi.org/10.1111/mec.12083>
- 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*, 16, 8311–8327.
- Kess, T., Gross, J., Harper, F., & Boulding, E. G. (2016). Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. *Journal of Molluscan Studies*, 82, 104–109.
- Lawniczak, M. K. N., Emrich, S. J., Holloway, A. K., Regier, A. P., Olson, M., White, B., ... Besansky, N. J. (2010). Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, 330, 512–514. <https://doi.org/10.1126/science.1195755>
- Lawson, L. P., & Petren, K. (2017). The adaptive genomic landscape of beak morphology in Darwin's finches. *Molecular Ecology*, 26, 4978–4989. <https://doi.org/10.1111/mec.14166>
- Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, 21, 1548–1566. <https://doi.org/10.1111/j.1365-294X.2012.05479.x>
- Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM, arXiv:1303.3997.
- Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24, 1031–1046.
- Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, 17, 142–152. <https://doi.org/10.1111/1755-0998.12635>
- McKinney, G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017). RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: Comment on Breaking RAD by Lowry et al. (2016). *Molecular Ecology Resources*, 17(3), 356–361.
- McKown, A. D., Klápště, J., Guy, R. D., Galdes, A., Porth, I., Hannemann, J., ... Douglas, C. J. (2014). Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist*, 203, 535–553.
- Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G. Y., & Miles, S. (2015). LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3: Genes, Genomes, Genetics*, 5, 2383–2390.
- Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., & Butlin, R. (2018). Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. *bioRxiv*, 447854. <https://www.biorxiv.org/content/10.1101/447854v1.abstract>
- Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: “isolation by adaptation” and multiple roles for divergent selection. *Evolution*, 62, 316–336. <https://doi.org/10.1111/j.1558-5646.2007.00299.x>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18, 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, 26, 160–167. <https://doi.org/10.1016/j.tree.2011.01.001>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., ... Wagner, H. (2017). *vegan: Community Ecology Package*. R package version 2.4-5. Retrieved from <https://CRAN.R-project.org/package=vegan>
- Otto, S. P., & Jones, C. D. (2000). Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. *Genetics*, 156, 2093–2107.
- Platt, A., Vilhjálmsson, B. J., & Nordborg, M. (2010). Conditions under which genome-wide association studies will be positively misleading. *Genetics*, 186, 1045–1052. <https://doi.org/10.1534/genetics.110.121665>
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38, 904–909. <https://doi.org/10.1038/ng1847>
- Pritchard, J. K., Stephens, M., Rosenberg, N. A., & Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics*, 67, 170–181. <https://doi.org/10.1086/302959>
- R, Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape of speciation. *Journal of Evolutionary Biology*, 30, 1450–1477.
- Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C., & Panova, M. (2016). Shared and non-shared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, 25, 287–305.
- Rivas, M. J., Saura, M., Pérez-Figueroa, A., Panova, M., Johansson, T., André, C., ... Quesada, H. (2018). Population genomics of parallel

- evolution in gene expression and gene sequence during ecological adaptation. *Scientific Reports*, 8(1), 16147. <https://doi.org/10.1038/s41598-018-33897-8>
- Rohlf, F. J. (1993). Relative warp analysis and an example of its application to mosquito wings. *Contributions to Morphometrics*, 8, 131–159.
- Rohlf, F. J. (2013). tpsDig 2.17. Distributed by Author. Department of Ecology and Evolution, State University of New York. New York, NY: Stony Brook.
- Rohlf, F. J. (2015). *tpsRelw. Relative warps analysis*. Stony Brook, NY: Department of Ecology and Evolution. State University of New York at Stony Brook.
- Rohlf, F. J., & Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 3, 40–59.
- Rolán-Alvarez, E., Austin, C., & Boulding, E. G. (2015). The contribution of the genus *Littorina* to the field of evolutionary ecology. In R. N. Hughes, D. J. Hughes, I. P. Smith & A. C. Dale (Eds.), *Oceanography and Marine Biology: An Annual Review*, Vol. 53 (pp. 157–214). London, UK: Taylor & Francis.
- Rolán-Alvarez, E., Carballo, M., Galindo, J., Morán, P., Fernández, B., Caballero, A., ... Johannesson, K. (2004). Nonallopatric and parallel origin of local reproductive barriers between two snail ecotypes. *Molecular Ecology*, 13, 3415–3424. <https://doi.org/10.1111/j.1365-294X.2004.02330.x>
- RStudio Team (2015). RStudio: Integrated development for R. Boston, MA: RStudio, Inc. Retrieved from <http://www.rstudio.com/>
- Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptations. *Nature Reviews Genetics*, 14, 807–820. <https://doi.org/10.1038/nrg3522>
- Schilling, M., Mullen, S., Kronforst, M., Safran, R., Nosil, P., Feder, J., ... Flaxman, S. (2018). Transitions from single- to multi-locus processes during speciation. *Genes*, 9, 274.
- Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution*, 16, 372–380. [https://doi.org/10.1016/S0169-5347\(01\)02198-X](https://doi.org/10.1016/S0169-5347(01)02198-X)
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15, 176–192. <https://doi.org/10.1038/nrg3644>
- Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M., & Nosi, P. (2011). Magic traits in speciation: 'magic' but not rare? *Trends in Ecology and Evolution*, 26, 389–397. <https://doi.org/10.1016/j.tree.2011.04.005>
- Stinchcombe, J. R., & Hoekstra, H. E. (2008). Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity*, 100, 158–170. <https://doi.org/10.1038/sj.hdy.6800937>
- Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367, 451–460. <https://doi.org/10.1098/rstb.2011.0260>
- Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, 17, 4334–4345. <https://doi.org/10.1111/j.1365-294X.2008.03921.x>
- Vilhjálmsdóttir, B. J., & Nordborg, M. (2013). The nature of confounding in genome-wide association studies. *Nature Reviews Genetics*, 14, 1–2. <https://doi.org/10.1038/nrg3382>
- Visscher, P. M., Macgregor, S., Benyamin, B., Zhu, G., Gordon, S., Medland, S., ... Martin, N. G. (2007). Genome partitioning of genetic variation for height from 11,214 sibling pairs. *American Journal of Human Genetics*, 81, 1104–1110. <https://doi.org/10.1086/522934>
- Vitturi, R., Libertini, A., Panozzo, M., & Mezzapelle, G. (1995). Karyotype analysis and genome size in three Mediterranean species of periwinkles (Prosobranchia: Mesogastropoda). *Malacologia*, 37, 123–132.
- Voight, B. F., & Pritchard, J. K. (2005). Confounding from cryptic relatedness in case-control association studies. *PLoS Genetics*, 1, e32.
- Westram, A. M., Panova, M., Galindo, J., & Butlin, R. K. (2016). Targeted resequencing reveals geographical patterns of differentiation for loci implicated in parallel evolution. *Molecular Ecology*, 25, 3169–3186. <https://doi.org/10.1111/mec.13640>
- Westram, A., Rafajlović, M., Chaube, P., Faria, R., Larsson, L., Panova, M., ... Butlin, B. (2018). Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. *Evolution Letters*, 2, 297–309.
- Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14, 851–865. <https://doi.org/10.1046/j.1420-9101.2001.00335.x>
- Yeaman, S. (2015). Local adaptation by small-effect alleles. *American Naturalist*, 186, S74–S89.
- Yeaman, S., & Whitlock, M. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65, 1897–1911. <https://doi.org/10.1111/j.1558-5646.2011.01269.x>
- Zelditch, M. L., Swiderski, D. L., Sheets, H. D., & Fink, W. L. (2004). *Geometric morphometrics for biologists: A primer*. San Diego, CA: Elsevier.

**How to cite this article:** Kess T, Boulding EG. Genome-wide association analyses reveal polygenic genomic architecture underlying divergent shell morphology in Spanish *Littorina saxatilis* ecotypes. *Ecol Evol*. 2019;9:9427–9441. <https://doi.org/10.1002/ece3.5378>