



# Genomewide evidence of environmentally mediated secondary contact of European green crab (*Carcinus maenas*) lineages in eastern North America

Nicholas W. Jeffery<sup>1,2</sup>  | Ian R. Bradbury<sup>1,2,3</sup>  | Ryan R. E. Stanley<sup>4</sup> |  
Brendan F. Wringe<sup>1</sup> | Mallory Van Wyngaarden<sup>3</sup> | J. Ben Lowen<sup>4</sup> |  
Cynthia H. McKenzie<sup>1</sup> | Kyle Matheson<sup>1</sup> | Philip S. Sargent<sup>1</sup> | Claudio DiBacco<sup>4</sup>

<sup>1</sup>Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St. John's, NL, Canada

<sup>2</sup>Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada

<sup>3</sup>Department of Ocean Sciences, Memorial University of Newfoundland, St. John's, NL, Canada

<sup>4</sup>Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, NS, Canada

## Correspondence

Nicholas W. Jeffery, Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St. John's, NL, Canada.  
Email: nick.jeffery@dfo-mpo.gc.ca

## Funding information

Canadian Healthy Oceans Network; Natural Sciences and Engineering Research Council of Canada

## Abstract

Genetic-environment associations are increasingly revealed through population genomic data and can occur through a number of processes, including secondary contact, divergent natural selection, or isolation by distance. Here, we investigate the influence of the environment, including seasonal temperature and salinity, on the population structure of the invasive European green crab (*Carcinus maenas*) in eastern North America. Green crab populations in eastern North America are associated with two independent invasions, previously shown to consist of distinct northern and southern ecotypes, with a contact zone in southern Nova Scotia, Canada. Using a RAD-seq panel of 9,137 genomewide SNPs, we detected 41 SNPs (0.49%) whose allele frequencies were highly correlated with environmental data. A principal components analysis of 25 environmental variables differentiated populations into northern, southern, and admixed sites in concordance with the observed genomic spatial structure. Furthermore, a spatial principal components analysis conducted on genomic and geographic data revealed a high degree of global structure ( $p < .0001$ ) partitioning a northern and southern ecotype. Redundancy and partial redundancy analyses revealed that among the environmental variables tested, winter sea surface temperature had the strongest association with spatial structuring, suggesting that it is an important factor defining range and expansion limits of each ecotype. Understanding environmental thresholds associated with intraspecific diversity will facilitate the ability to manage current and predict future distributions of this aquatic invasive species.

## KEYWORDS

*Carcinus maenas*, European green crab, invasive species, RAD-seq, seascape genetics

## 1 | INTRODUCTION

The contemporary study of “seascape genetics” (Riginos & Liggins, 2013) is increasingly revealing fine-scale local structuring through the use of genomic data, allowing robust investigations of the forces influencing spatial genetic variation. Fine-scale population structure in marine species (Benestan et al., 2015; Bradbury et al., 2010; Harrison et al., 2017) may be driven by a number of processes, including secondary contact among lineages or closely related species (e.g., Abbott et al., 2013; Hewitt, 1996, 2000; Mayr, 1942), recent adaptation along an environmental gradient (e.g., Endler, 1977; Sanford & Kelly, 2011), or some combination of physical isolation, genetic drift, and selection (Harrison & Larson, 2016). In practice, these processes may be difficult to distinguish, particularly in nonequilibrium systems (Bierne, Welch, Loire, Bonhomme, & David, 2011; Harrison et al., 2017). Nonetheless, secondary contact has a long history of identification as a primary driver of clinal variation, and is increasingly implicated as a cause of fine geographic scale variation, and a source of heterogeneous genomic divergence in marine species (Barney, Munkholm, Walt, & Palumbi, 2017; Berg et al., 2017).

The advent of genomewide association studies, through technologies such as restriction-site associated DNA sequencing (RAD-seq), transcriptomics, and whole genome sequencing provides a novel platform to evaluate genomic variation and divergence associated with allopatric isolation or selection (Benestan et al., 2015; Sherman et al., 2016; Tepolt, 2015; Tepolt & Palumbi, 2015). Although the application of RAD-seq to evaluate genetic variation has recently come under fire (i.e., Hoban et al., 2016; Lowry et al., 2017; but see McKinney, Larson, Seeb, & Seeb, 2017), owing to the sparsity of markers across the genome (hundreds to tens of thousands of markers across genomes which are hundreds of millions to billions of base pairs long), it has been proven as a capable tool to evaluate population structure and environmental associations, provided careful interpretation and acknowledgment of the limitations (Bierne, 2010; Bierne et al., 2011). Fine-scale population structure has been identified in a variety of taxa including kelp, pelagic fish, mollusks, echinoderms, and crustaceans (see Riginos & Liggins, 2013 and references therein) using high-throughput sequencing at a range of patch sizes spanning several meters to thousands of kilometers. Invasive species where hybridization occurs among invasive and native species or waves of invasion represent an unique natural history experiment (e.g., Fitzpatrick et al., 2010; Saarman & Pogson, 2015; Viard, David, & Darling, 2016) and allow the examination of both how secondary contact may influence the distribution of genomic variation as well as how hybridization and introgression maybe be influenced by factors such as environmental variation.

The European green crab (*Carcinus maenas*) is a poignant example of such an invasive species. Green crab are highly tolerant of a wide range of temperatures and salinities, encompassing a 25°C range in sea surface temperatures in their native range and a range in salinity tolerance of 4–52‰ in adults (Klassen & Locke, 2007; Tepolt & Somero, 2014). At least two independent introductions of green crab have occurred since the 1800s (Audet et al., 2003; Blakeslee

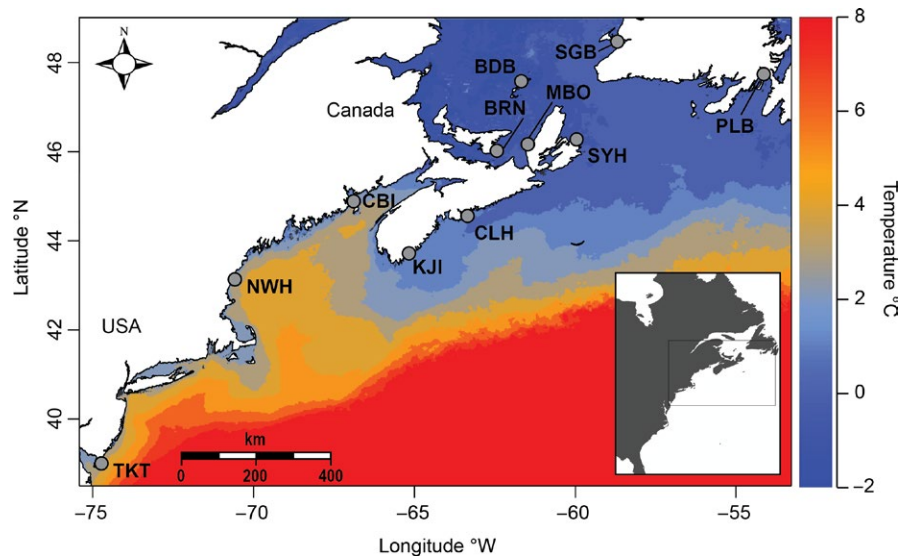
et al., 2010; Roman, 2006), resulting in the establishment of genetically distinct northern and southern ecotypes (Jeffery, DiBacco, Van Wyngaarden et al., 2017; Tepolt & Palumbi, 2015). These ecotypes correspond to northern and southern populations in the native range (Roman, 2006). Audet et al. (2003) revealed that prior to the introduction of the presumably cold-adapted northern European green crab lineage, cold coastal currents appear to have halted the northward expansion of the southern green crab ecotype at approximately the latitude of Halifax, NS (~44.6°N). Similarly, cold winters in the Gulf of Maine have previously been linked to increased adult mortality whereas warm winters have been linked to increased recruitment (Yamada & Kosro, 2010). In both the native and introduced ranges, the mean high temperature at which cardiac function fails in adult crabs is consistently higher in southern populations which are adapted to overall warmer sea surface temperatures (Tepolt & Somero, 2014). This latitudinal pattern in thermal tolerance likely evolved in the native range and has since mediated the success and expansion of populations in eastern North America (Tepolt & Somero, 2014). It is clear that the environment plays an important role in the population structure and distribution of this species, though the specific nature of this relationship remains unknown.

Here, we investigate environmental correlates of secondary contact in this species using genomewide climate-associated SNPs obtained by RAD sequencing of green crabs from locations along their invasive range in eastern North America. We first seek to explore the presence of genomic–environmental-associated variation using latent factor mixed models (LFMM; Frichot, Schoville, Bouchard, & François, 2013) and a Bayesian genome scan method (Villemereuil & Gaggiotti, 2015). Based on physiological evidence from Tepolt and Somero (2014), we predict that seasonal temperature minima are limiting the range expansion of the invasive green crab ecotypes and that environmentally associated genomic regions will track a steep cline in temperature previously reported in Atlantic Canada. This work builds directly on previous studies highlighting the genomic differences (Jeffery, DiBacco, Van Wyngaarden et al., 2017; Tepolt & Palumbi, 2015) and subsequent hybridization of two independent waves of invasion (Darling, Tsai, Blakeslee, & Roman, 2014; Jeffery, DiBacco, Wringe et al., 2017) to better understand the environmental correlates of population specific invasion success and secondary contact in this species. Our results will provide insight into how secondary contact dynamics can be correlated with environmental heterogeneity potentially constraining the success of coastal marine invaders, and will help to understand and predict future range expansions and contractions of green crab ecotypes.

## 2 | MATERIALS AND METHODS

### 2.1 | Environmental predictors

We selected 25 environmental variables, specifically annual and seasonally averaged bottom and surface temperatures and salinities, which have been noted as ecologically relevant to *C. maenas* (Klassen & Locke, 2007). A decadal average (2002–2012) was



**FIGURE 1** Map of eastern North America, including the 11 sampling locations where green crabs (*Carcinus maenas*) were collected. The Atlantic Ocean winter sea surface temperature (°C) is shown across this range

taken for each of four three-month seasonal aggregations corresponding with winter (January–March), spring (April–June), summer (July–September), and fall (October–December). Seasonal sea surface temperature (SST) and sea surface salinity (SSS) were assembled at a 1 km<sup>2</sup> resolution from Level 3 SST climatological satellite data (Aqua-MODIS 4-micron nighttime) (Sbrocco & Barber, 2013) and global oceanographic climatological SSS composites (Tyberghein et al., 2012). All climatological data layers were converted to ASCII grid with WGS84 global stereographic projection and a uniform land mask applied. Environmental variables (SST and SSS) corresponding to each geo-referenced sample location were subsequently extracted.

## 2.2 | Detection of environment associated loci

We used previously published RAD data from Jeffery, DiBacco, Van Wyngaarden et al. (2017) which included 9,137 SNPs in 241 individual crabs from 11 sites ranging from New Jersey, USA, to southern Newfoundland, Canada (Figure 1). To detect loci which show a strong association with the environmental data, we implemented latent factor mixed models (LFMM; Frichot et al., 2013) available in the R package *LEA* (Frichot & François, 2015) as the first method of outlier detection. This method detects correlations between environmental variables and genotypic variation to reveal outliers, while controlling for population structure ( $K$ ) and unmeasured environmental variables as latent factors. We set  $K = 2$  based on the results of previous discriminant analyses of principal components (DAPC) and *STRUCTURE* (Pritchard, Stephens, & Donnelly, 2000) results (Jeffery, DiBacco, Van Wyngaarden et al., 2017; Jeffery, DiBacco, Wringe et al., 2017). LFMM was run with a burn-in of 5,000 iterations and 15,000 Markov Chain Monte Carlo (MCMC) iterations which were replicated five times. Z-scores per locus were combined from the five replicates, and false discovery rates were evaluated using Benjamini and Hochberg (1995) adjusted  $p$ -values. We then combined the lists of loci detected as outliers for each of the 25

environmental variables we tested, retaining all loci with an adjusted  $p$ -value < .05.

We compared our results from LFMM to BayeScEnv (Villemereuil & Gaggiotti, 2015), an  $F_{ST}$ -based genome scan approach which accounts for environmental differentiation among populations. This method has a lower false discovery rate relative to purely  $F_{ST}$ -based methods (Villemereuil & Gaggiotti, 2015). We standardized the environmental variables for each population by taking the absolute difference from the mean and dividing each by the standard deviation as recommended by the program authors. We used 20 pilot runs with 50,000 MCMC iterations each and an initial burn-in of 50,000 iterations for each of the 25 environmental parameters. The prior probability that a locus is an outlier was set to 0.1 to minimize the false discovery rate. As with LFMM, we combined all outliers for all environmental parameters and then compared the number of overlapping loci between the two methods. Outliers were subset from our full SNP panel using the R package *genepopedit* (Stanley, Jeffery, Wringe, DiBacco, & Bradbury, 2017) to generate our outlier panel of SNPs. Linkage disequilibrium  $r^2$  values were calculated among all environmental outliers in *PLINK* (Purcell et al., 2007). Our nonoutlier panel of SNPs was then comprised of all SNPs not detected as outliers by either LFMM or BayScEnv, thus excluding any loci that were strongly correlated with the environment.

## 2.3 | Environmental association and redundancy analysis

To determine the extent of global and local population structure of green crabs, we conducted a spatial principal component analysis (SPCA). We used the R function *CartDist* (Stanley & Jeffery, 2017) to first calculate least-cost distances among populations in the R package *marmap* (Pante & Simon-Bouhet, 2013) using land as an impenetrable barrier. We then used a nonmetric multidimensional scaling function in *vegan* (Oksanen, Kindt, Legendre, O'Hara, & Stevens, 2017) to convert these least-cost distances into Cartesian coordinates which accounted

for land barriers in the SPCA. The SPCA was run on the nonoutlier and outlier RAD panels, using the Delaunay triangulation method (Delaunay, 1934) in the R package *adegenet* (Jombart, 2008). We used Monte Carlo randomization tests with 1,000 permutations to assess the significance of global (high autocorrelation) and local (dissimilar neighboring sites) structures. We then extracted the first axis lagged scores from the environmental outlier SPCA and used a multiple linear regression to determine any significantly associated environmental variables, calculating the variance inflation factors (VIF) for each model and removing each variable with the highest VIF until all VIFs were <5, indicating the removal of collinearity among predictors. We additionally ran a logistic generalized linear model of the SPCA scores and latitude to model the observed cline in population structure. To verify that the VIF model selection worked, we additionally conducted elastic net regression on the SPCA lagged scores and environmental data which accounts for collinearity among predictors in the R package *glmnet* (Friedman, Hastie, & Tibshirani, 2010).

Next, we used a principal components analysis (PCA) in *adegenet* to evaluate environmental similarity among our sample locations using all environmental variables. We then used redundancy analyses (RDA) and partial redundancy analyses (pRDA) to account for the geographic autocorrelation of each site for both the outlier and nonoutlier locus allele frequencies using *vegan* to select which environmental variables best explain the genetic structure. Analyses of variance (ANOVA with 1,000 permutations) were used to assess the significance of each environmental parameter within the RDA. The *ordistep* function, a stepwise permutational ordination method, was used with 1,000 permutations to evaluate the environmental variables and create a model which best explained the spatial distribution of our genotype data.

Finally, we compared a simple isolation-by-distance (IBD) relationship as a null hypothesis to a genetic-environment association controlling for geographic distances using multiple regression on distance matrices (MRM; Lichstein, 2007). For the IBD relationship, we used pairwise population linearized  $F_{ST}$  values as the response value and least-cost distances calculated in *marmap* as the explanatory variable using the MRM function with 10,000 permutations in the *ecodist* R package (Goslee & Urban, 2007). We then conducted an MRM using the best-supported environmental variable(s) and geographic distance from the above analyses to determine the significance of the environment on gene flow among populations.

## 2.4 | Cline width and dispersal estimation

We used the R package *HZAR* (Derryberry, Derryberry, Maley, & Brumfield, 2014) on our 41 environmental outliers and 100 randomly sampled loci from the neutral SNP panel to model clines of allele frequency and least-cost distances among sampling sites. Placentia Bay individuals were removed due to their anthropogenic introduction to the northern end of the sampling locations which would skew the cline width, as they are a known hybrid population (Blakeslee et al., 2010; Jeffery, DiBacco, Van Wyngaarden et al., 2017; Tepolt & Palumbi, 2015). We tested seven clinal models per locus, including one null

model and combinations of fixed, free, or no scaling of maximum and minimum allele frequencies, and exponential tails at neither or both ends of the cline. The best model was selected by a corrected Akaike information criterion (AICc), and loci were deemed to be clinal based on those with a minimal negative log-likelihood above the 90% quantile limit for all loci. *HZAR* provided estimates of both cline center and width for the neutral and outlier loci. Average dispersal was then calculated based on Sotka and Palumbi (2006), where dispersal is equal to the square of the cline width multiplied by a product of the maximum linkage disequilibrium ( $R^2$ ) for each locus and recombination rate  $r$  (0.5).

## 2.5 | Gene ontology

We attempted gene ontology using Blast2GO 3.3.5 (Conesa et al., 2005) on 80 bp FASTA sequences of all 41 environmentally associated outlier loci to determine whether any outlier loci were associated with genes that may be related to environmental tolerances (i.e., heatshock or immune system proteins). We applied the *blastn* algorithm and queried multiple databases including SWISSPROT and NCBI nucleotide/protein collections with an *E*-value cutoff of  $1e^{-4}$  and otherwise default parameters.

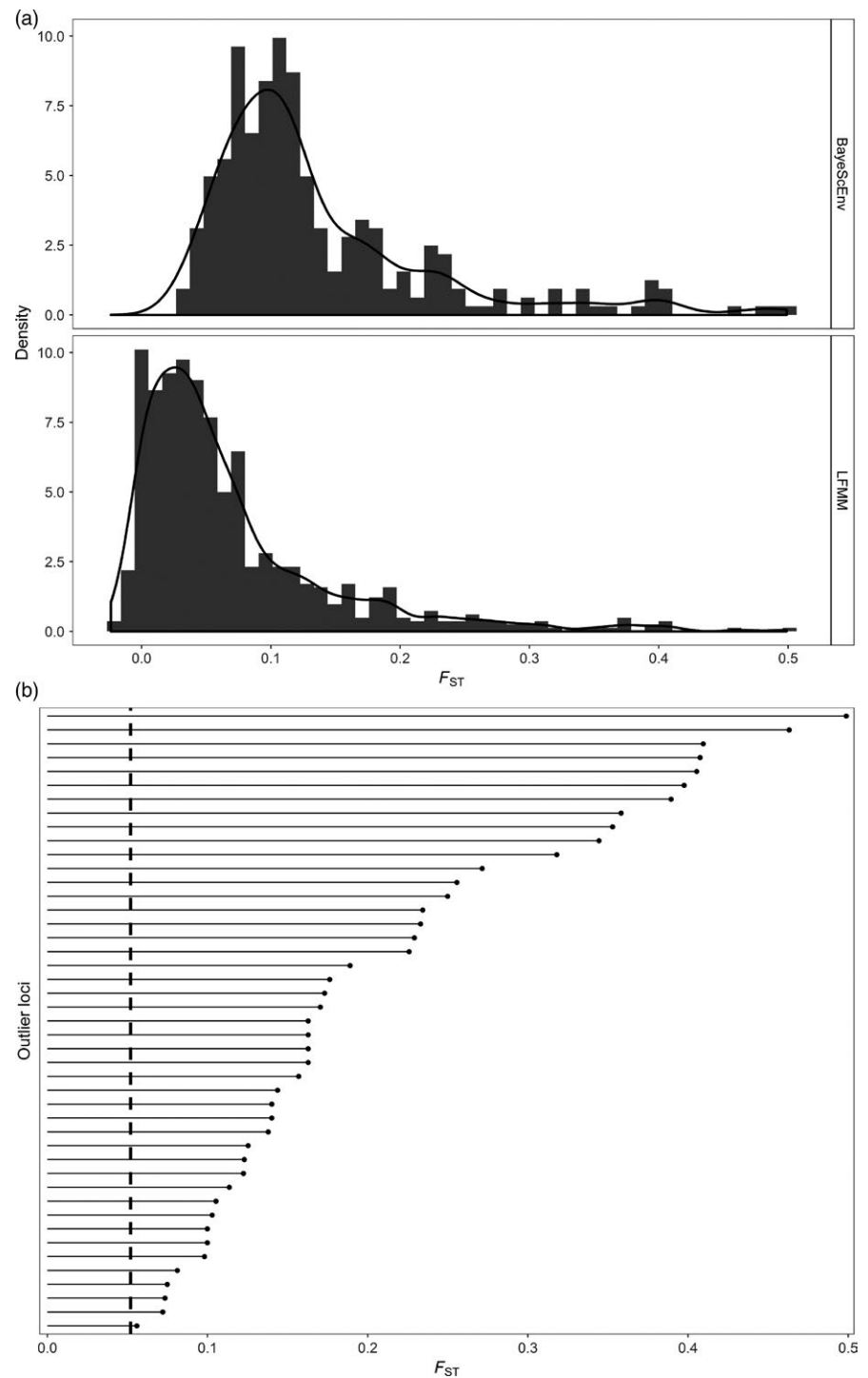
## 3 | RESULTS

### 3.1 | Environmental outlier detection

Of the 9137 SNPs analyzed from Jeffery, DiBacco, Van Wyngaarden et al. (2017), LFMM classified 770 loci as outliers across all environmental variables at a false discovery rate of 0.05 and BayeScEnv detected 302 loci as outliers for at least one environmental variable. Of the outliers reported by LFMM and BayeScEnv, 45 loci (0.49% of all SNPs) were detected by both methods (Figure 2). We then further removed four SNPs that were located on the same RAD-tag as other SNPs, resulting in 41 loci that were used as our environmental outlier panel, while 1,027 loci were removed from the full RAD panel to create a panel of 8,110 nonoutlier SNPs (88.76%). The outlier loci showed a clear north-south division in standardized allele frequencies (Figure 3a), which corresponds to an abrupt break in seasonal temperatures (Figure 3b). Locus-specific  $F_{ST}$  values for the outlier loci ranged from 0.056 to 0.499 (mean  $\pm$  SD:  $0.23 \pm 0.12$ ; Figure 2), and linkage disequilibrium values among environmental outliers were relatively low (mean  $r^2 \pm$  SD:  $0.08 \pm 0.09$ ).

### 3.2 | Principal component analysis of environmental data

A PCA of all environmental parameters at each sampling location resulted in the clustering of sampling locations that mirrored the genetically determined northern, admixed, and southern population clusters found by Jeffery, DiBacco, Wringe et al. (2017) (Figure 4). The top seven seasonal environmental variables based on loading weight on the first principal component axis were used in subsequent redundancy analysis models. A linear regression of



**FIGURE 2** (a)  $F_{ST}$  frequency distributions for 302 loci detected by BayeScEnv and 770 loci detected by latent factor mixed models (LFMM) as putative outliers correlated with at least one environmental variable. (b)  $F_{ST}$  values for each of the 45 outlier loci that overlap between both BayeScEnv and LFMM. The dashed vertical line indicates the mean  $F_{ST}$  for all 9,137 loci (0.052)

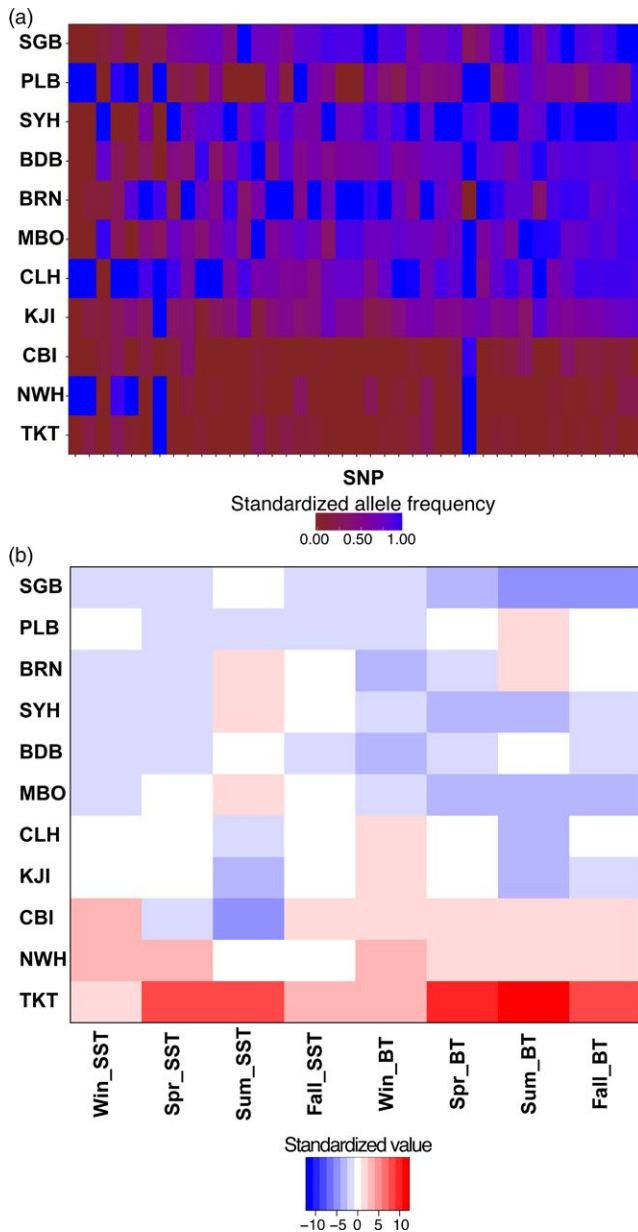
the first principal component axis from the environmental data versus the first principal component axis from outlier allele frequency data yielded a significantly positive relationship ( $r^2 = .635$ ,  $p = .003$ ; Figure S1), indicating a strong association between the environmental structuring and genomic structuring.

### 3.3 | Environmental associations with genotype

Separate SPCAs on both the environmental outlier and nonoutlier datasets revealed a significant degree of global structure ( $p = .001$ )

and little local structure ( $p = 1.0$ ) as noted by the extreme positive eigenvalue that explained 66% of the observed structure (Figure 5a inset). A logistic regression of SPCA lagged scores against latitude showed a clear cline in population structure (Figure 5a). A multiple linear regression of the SPCA axis 1 lagged scores for the outlier loci against each seasonal environmental predictor with a VIF  $< 5$  found that both winter sea surface temperature (WinSST  $R_{adj}^2 = .90$ ,  $p = .002$ ; Figure 5b) and summer bottom temperature ( $p = .04$ ) were significantly associated (Table S1). Similarly, elastic net regression selected a model with only two predictors—WinSST and SumBT





**FIGURE 3** (a) Standardized allele frequency heat map for the 41 loci detected as environmental outliers. A distinct cline is observed between the southern and northern populations of green crab. (b) Heat map of site-specific seasonal sea surface and bottom temperatures

(coefficients 0.64 and 0.08, respectively, all others = 0)—that best correlated with the outlier and nonoutlier SPCA lagged scores.

The relationship between WinSST and observed population-genetic structure was also highlighted by the redundancy analysis, which revealed WinSST to be the most significant environmental variable associated with the outlier ( $F = 13.03$ ,  $p = .005$ ) and non-outlier SNP genotype distribution ( $F = 4.5024$ ,  $p = .02$ ) selected by the *ordistep* function (Figure 6). When accounting for geographic proximity among sites using a partial RDA, WinSST was again the top predictor of both outlier and nonoutlier genotypes based on model selection, although this was marginally nonsignificant for

the nonoutlier SNPs ( $F = 13.03$ ,  $p = .005$  and  $F = 2.41$ ,  $p = .055$ , respectively). Subsequent analyses of variance revealed that WinSST was the only significant predictor of both the nonoutlier and outlier variation ( $F = 11.80$ ,  $p = .002$  and  $F = 10.45$ ,  $p = .007$ , respectively). Variance partitioning based on the outlier RDA models revealed that the environment explains 43.95% and geography explains 9.35% of the variation in genetic structure, with the remaining variance explained by the interaction between environment and geography.

Using multiple regression on distance matrices, we observed a significant isolation-by-distance relationship among populations ( $r^2 = .39$ ,  $p = .004$ ) largely driven by inter-regional comparisons. However, an MRM model including winter sea surface temperature had a stronger relationship with pairwise  $F_{ST}$  values ( $r^2 = .55$ ,  $p = .002$ ) suggestive of this environmental variable's role in possibly constraining gene flow.

### 3.4 | Estimation of dispersal

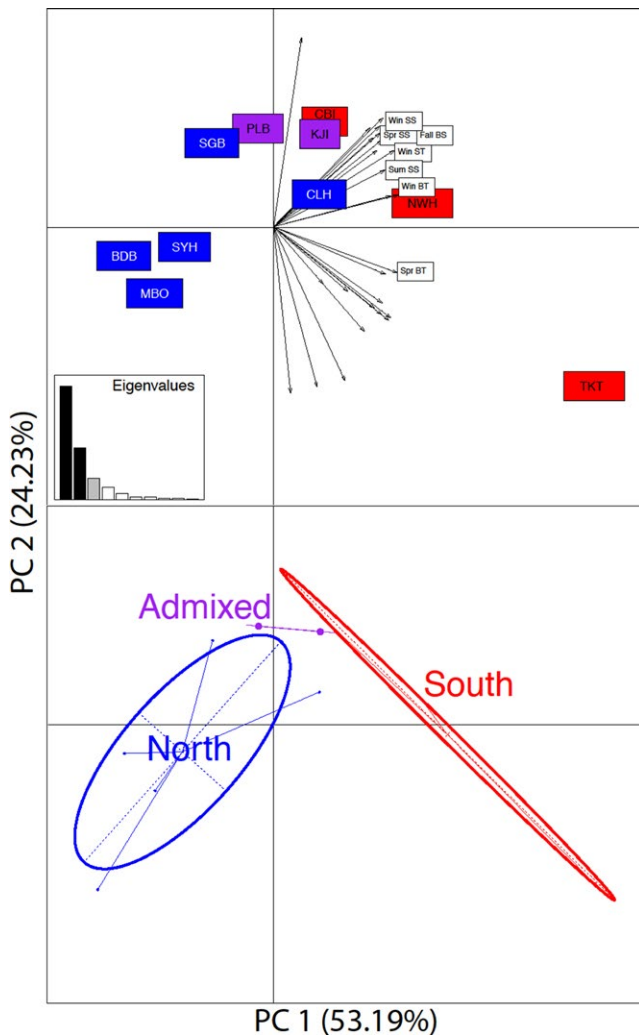
Twenty-four of 41 outliers (58.5%) were considered clinal based on their log-likelihood distribution, while only 10 of the randomly selected 100 nonoutlier loci (10%) were considered clinal (Figure S2). The center of the cline was estimated to be  $994.91 \pm 291.67$  km (mean  $\pm$  SD) from the southernmost sites, while the cline width was estimated to be  $299.22 \pm 666.88$  km based on outlier allele frequencies. The 10 clinal nonoutlier loci estimated the cline center at  $955 \pm 145.83$  km and a cline width of  $477.07 \pm 793.07$  km. Neither the cline centers ( $p = .62$ ) or the cline widths ( $p = .56$ ) were significantly different between nonoutlier and outlier loci when correcting for sample size. Nonoutlier loci estimated the standard deviation of parent-offspring distance to be  $82.31 \pm 127.71$  km (mean  $\pm$  SD). The standard deviation of parent-offspring distance was estimated to be  $138.50 \pm 319.95$  km based on the clinal environmental outliers.

### 3.5 | Gene ontology

Our Blast2GO search of 41 outlier loci yielded only one reliable result, which aligned an outlier sequence to a Toll-like receptor (TLR3) from the decapod *Portunus trituberculatus* at an e-value of  $1.6e^{-20}$ . TLRs are cell-surface receptors which aid in the innate immune system (Li, Li, Wang, & Song, 2015) and as such may have variable alleles that are advantageous in different environments.

## 4 | DISCUSSION

Here, we present results consistent with the environmentally associated distribution of secondary contact punctuated by an environmental gradient among two independent invasions of green crab along the coast of eastern North America. Our multivariate analyses reveal winter sea surface temperature to be the best predictor of spatial structure after accounting for multicollinearity among environmental predictors. Similarly, multiple regression on distance matrices revealed a stronger isolation-by-environment relationship



**FIGURE 4** Principal component analysis on the top seven seasonal environmental variables for our 11 sampling locations based on their loadings shows a clear separation of northern and southern sites along PC1. Win, winter; Spr, spring; Sum, summer; BS, bottom salinity; BT, bottom temperature; SS, surface salinity; ST, surface temperature (top panel). Sites classified as northern, southern, or admixed, based on Jeffery, DiBacco, Wringe et al. (2017) are clearly separated based on the environment alone, with admixed sites being intermediate to the northern and southern sites (bottom panel)

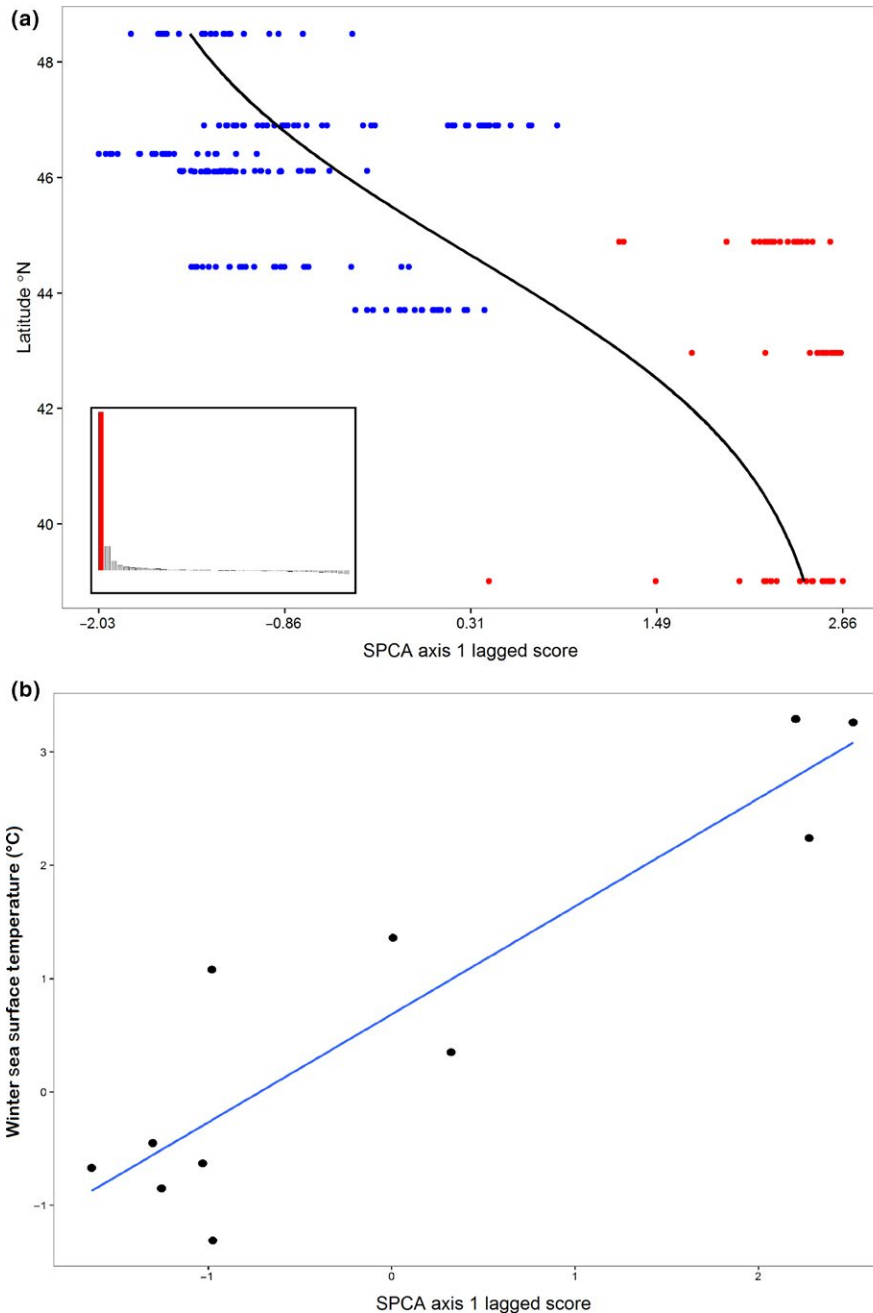
than a simple isolation-by-distance relationship, suggesting that the environment is more of a constraining influence on gene flow than geographic distance among populations (Keller, Van Strien, & Holderegger, 2012). Our results show that in the zone of secondary contact, an abrupt change in winter surface temperature, and to a lesser degree, summer bottom temperature, correlates with a nonlinear shift in allele frequency in both the outlier loci and the nonoutlier loci between green crab ecotypes. These variables represent seasonal extremes (lows and highs), which intersect with the physiological tolerances of each ecotype carried over from their native range, and are likely factors which limit gene flow and range expansion in this species (Tepolt & Somero, 2014). In the native range, green crab show a genetic continuum of mitochondrial lineages

along their latitudinal gradient (Roman & Palumbi, 2004) and show large differences in structure between northern and southern populations (Tepolt & Palumbi, 2015) entirely consistent with the patterns observed here. The broad invasive geographic range of green crab in eastern North America was likely made possible by the two independent invasions originating from very different thermal regimes in the native range, a fact that is supported by evidence that the initial southern invasion stalled in the Halifax region (~44.5°N) in the 1960s, perhaps limited by the cold seasonal temperatures north of this region. This system is unique in this respect, where two genetically distinct lineages underwent independent invasions, allowing for a broad colonization and the study of environmental associations in a newly invasive and highly dynamic system.

This work builds directly on a large body of literature on secondary contact both in native and among native and invasive species (Fitzpatrick et al., 2010; Saarman & Pogson, 2015; Viard et al., 2016). For example, in the *Mytilus* mussel complex, three species exhibit strong barriers to reproduction, although hybridization still occurs (Saarman & Pogson, 2015). A hybrid zone exists between the invasive *M. galloprovincialis* and native *M. trossulus* in California, a system which has existed for several hundred years; Saarman and Pogson (2015) suggest that the transfer of adaptive genes from *M. edulis* to *M. galloprovincialis* in the species' native European range may have contributed to the invasive success of *M. galloprovincialis*. Similarly, Barred Tiger salamanders (*Ambystoma tigrinum mavortium*) recently introduced (<60 years) into the native habitat of the California Tiger salamander (*A. californiense*), form a strong hybrid zone with  $F_1$  hybrids showing increased fitness relative to either pure parent species (Fitzpatrick et al., 2010). However, unlike in species which show both endogenous and exogenous barriers to gene flow (Saarman & Pogson, 2015), green crab ecotypes interbreed successfully (Jeffery, DiBacco, Wringe et al., 2017), although regions of secondary contact are environmentally intermediate relative to the habitat of northern or southern ecotypes. Collectively, these studies reinforce the need for knowledge of the demographic history and proper interpretation of the adaptive or neutral processes that lead to the creation of hybrid zones, particularly in recent introductions which may have not yet reached equilibrium.

#### 4.1 | Seascape genomics of marine animals

Studies that directly relate environmental data to population-level genomic data in aquatic species ("seascape genomics") are rare, and to our knowledge, no studies exist that integrate genomic data and fine-scale environmental data in an invasive species with multiple invasions (Sherman et al., 2016). Recent investigations of seascape genomics for North American Atlantic species have detected significant associations between the environment and structure in Atlantic cod (*Gadus morhua*) (Bradbury et al., 2010), American lobster (*Homarus americanus*) (Benestan et al., 2016), and sea scallop (*Placopecten magellanicus*) (Van Wyngaarden et al., 2017). Sea bottom temperature was revealed to be correlated with outlier loci on both sides of the Atlantic within cod, suggestive of parallel evolution



**FIGURE 5** (a) The observed logistic relationship between the lagged scores of a spatial principal component analysis for our environmental outlier loci and latitude, showing a transition from the southern to northern zones at approximately 44°N. The inset shows the spatial principal component analysis (SPCA) eigenvalues showing a high degree of global structure (extreme positive eigenvalue) but little local structure (negative eigenvalues). (b) A significant relationship is observed ( $r^2 = .90$ ,  $p = .04$ ) between the axis 1 lagged scores from a spatial principal component analysis on environmental outlier genotypes and winter sea surface temperature

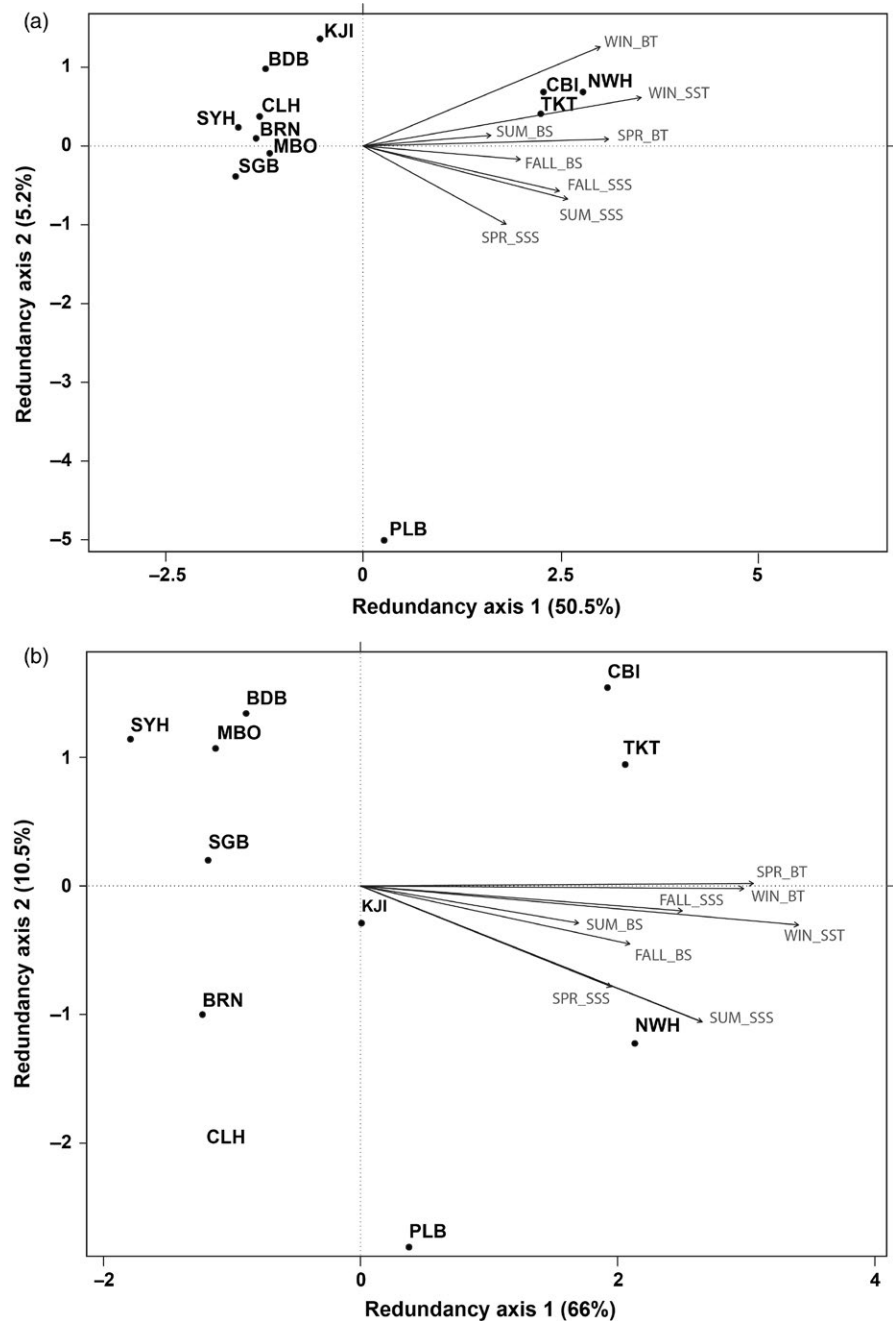
in distinct cod lineages (Bradbury et al., 2010). For American lobster, Benestan et al. (2016) revealed that ocean currents best explained neutral population structure, but that minimum annual sea surface temperature (SST) was likely driving local adaptation. Because adult lobsters are benthic, it was hypothesized that this minimum SST is experienced by pelagic larval lobsters directly affects their mortality, thus leading to local adaptation in warmer or colder regions. Van Wyngaarden et al. (2017) concluded that cold temperatures, including minimum annual and winter temperatures, in addition to differences in salinities among sites, potentially restrict gene flow among northern and southern populations, leading to strong latitudinal structuring in this species. Similar to green crab, the native cod, scallop, and lobster show strong north-south clustering in the

Northwest Atlantic, and are primarily structured by periods of extreme cold in the winter which limits gene flow between northern and southern populations and may impose a selective pressure. Whereas both American lobster and sea scallop are native to this region, it seems reasonable to hypothesize that invasive green crab are likely adapted to different temperatures from their native range and that this may be limiting range expansion and secondary contact in eastern North America (Tepolt & Somero, 2014).

#### 4.2 | Green crab environmental limits

Our results suggest that winter sea surface temperature primarily maintains the spatial structure of green crab ecotypes in eastern





**FIGURE 6** (a) A redundancy analysis on the seasonal temperature environmental variables and neutral loci allele frequencies suggests winter sea surface temperature as the most significant predictor of genetic differentiation. (b) Redundancy analysis on the outlier SNPs additionally suggests winter sea surface temperature as the most significant predictor genetic differentiation

North America. It is this cold seasonal temperature that likely constrains the range and gene flow of two genetically distinct invasions, showing an abrupt break in population structure within their northwestern Atlantic range, except for a zone of secondary contact in southern Nova Scotia. The broad-scale environmental correlations observed in this study are most likely carried over from structure in the native range, as both our outlier and nonoutlier loci showed similar relationships to temperature, potentially reflective of the long-term divergence between ecotypes in Europe (Jeffery, DiBacco, Wringe et al., 2017). Because the observed environmental associations are present both in outlier and nonoutlier loci (although to a lesser extent), a simplistic interpretation of the factors responsible for these associations remains elusive. The environmental

associations observed within outlier loci may represent a genome-wide signal that is the result of several historical processes, including selection, drift, or both, in the native range, rather than true genomic islands of divergence that are associated with temperature-dependent selection. These loci may also represent alleles that are intrinsically incompatible between ecotypes that have become associated with the environment along a climatic gradient via the endogenous-exogenous barriers “coupling hypothesis” (sensu Bierne et al., 2011). Without genomic data from the native range and a comprehensive green crab genome, these hypotheses cannot be tested directly here. Interestingly, some outlier loci in NWH, PLB, and CLH showed the opposite pattern of allele frequency than is expected based on the majority of outliers. Although these loci showed no

obvious trend with environmental variables, this pattern suggests the presence of genomic heterogeneity within the green crab genome and variation in evolutionary processes involved, such as selection or introgression between ecotypes across the genome.

WinSST represents the extreme low temperatures these crabs face annually, although it remains to be seen which life stages (postsettled juveniles or adults) are most affected by this minimum temperature. Reproductive timing varies across the invasive range, with mating typically occurring in the summer or early fall (Best, McKenzie, & Couturier, 2017; Klassen & Locke, 2007), and as larvae spend several weeks suspended in the water column, it may be that postsettled juveniles of southern ecotype crabs are unable to survive the decrease in sea surface temperature in the northern region (deRivera et al., 2007). Our finding that WinSST had the most robust relationship with ecotype distributions is consistent with the known environmental tolerances of adult green crab and their ability to grow and survive at different latitudes (Lowen, McKindsey, Therriault, & DiBacco, 2016). Additionally, previous studies have noted that cold winter water temperatures are linked to higher adult mortality and lower recruitment (Berrill, 1982; Beukema, 1991), while warmer winters may be linked to stronger year-classes in green crab populations in western North America (Yamada & Kosro, 2010). A recent review of environmental factors influencing the local (<10 km) green crab population structure suggests that water depth and biotic interactions were the primary variables influencing distribution (Cosham, Beazley, & McCarthy, 2016). Interestingly, Cosham et al. (2016) found that temperature did not directly influence local distribution of green crab across its native European range, although this was only at the fine-scale within regions and ultimately not at the scale where the variation in temperature might evoke a selective pressure or impose significant constraints on gene flow.

Our principal component analysis on the environmental data showed distinct northern and southern groupings that correlate strongly with previously detected genetic spatial structure. Known hybrid sites—Kejimikujik, NS and Placentia Bay, NL—appear intermediate to the northern and southern sites in both the environmental and genotypic PCAs (Blakeslee et al., 2010; Jeffery, DiBacco, Wringe et al., 2017). This suggests that both northern and southern ecotypes are able to interbreed in these zones of secondary contact, although the physiological tolerances of hybrid individuals require testing to determine why hybrids are not found in high abundance in other parts of their invasive range (Jeffery, DiBacco, Wringe et al., 2017). Despite its northern location, our PCA suggests that Placentia Bay is similar to Kejimikujik in terms of its environment, and indeed Placentia Bay has warmer surface waters than most of southern Newfoundland, but is still colder than the Maritimes overall (Craig & Colbourne, 2004). The hybrid crabs' ability to survive in this environment likely depends on their ability to move deeper into the bay to avoid the cold winter surface temperatures, which they cannot do in the more exposed St. George's Bay region of Newfoundland. Similar results of interspecific hybridization in anemonefish in the Arabian Peninsula showed an admixture zone in a region associated with intermediate environmental conditions, suggesting differential local

adaptation and intermediate physiological tolerances in hybrid individuals (Saenz-Agudelo et al., 2015). However, it remains to be seen whether hybrid green crabs have a greater or intermediate environmental tolerance relative to the pure northern or southern individuals, as the successful establishment of hybrid populations requires that hybrids persist in either intermediate or broader niches relative to their parental populations or species to avoid being outcompeted (Pruvost, Hollinger, & Reyer, 2013; Seehausen, 2004).

Collectively as a species, green crab have a wide tolerance of a range of temperatures and salinities (Klassen & Locke, 2007). Their broad thermal tolerance, however, is partitioned into the separate ecotypes when the underlying genetic structure is taken into consideration (Tepolt & Somero, 2014). In both their native and introduced ranges, southern populations of green crab have a higher cardiac critical temperature before heart failure than northern populations, while northern populations had a lower critical temperature before heart failure, indicative of potential temperature adaptation and/or acclimatory plasticity (Tepolt & Somero, 2014). Differences in thermal tolerance found in the eastern North America invasive range are likely carried over through their initial introduction, leading to the observed clinal structure of this species. Tepolt and Somero (2014) admit that their study cannot differentiate plasticity from long-term adaptation, but that both processes have likely played a role in the global invasions of green crab. While cardiac critical temperature is just an acute metric of thermal tolerance, these physiological differences corroborate our overall result that winter sea surface temperature is an important factor influencing population structure and limiting gene flow of northern and southern ecotypes in eastern North America, which has been maintained since the separate introductions of these ecotypes from their native range.

### 4.3 | Clines and estimates of dispersal

Estimating larval dispersal rates has historically been achieved using oceanographic models and physical or experimental observation of propagule movement, as well as field observations of the rate of spread of invasive species (Selkoe & Toonen, 2011; Shanks, Grantham, & Carr, 2003). We used a model of cline width calculated from allele frequencies proposed by Sotka and Palumbi (2006) and based on clinal models of gene flow (Barton & Gale, 1993) to estimate per-generation dispersal in eastern North America. Our approximations of the center of clines in allele frequency using both putative outliers and nonoutliers was consistent with the "KJI" sample location in southeastern coastal Nova Scotia, a known zone of secondary contact (Blakeslee et al., 2010; Darling, Bagley, Roman, Tepolt, & Geller, 2008; Jeffery, DiBacco, Wringe et al., 2017). Nonoutlier and outlier clinal loci yielded estimates of dispersal on similar orders of magnitude (~80–140 km/generation) and are similar to estimates reported by Grosholz (1996); Klassen and Locke (2007) and Shanks et al. (2003) for green crab, and are consistent with the high degree of genomewide divergence detected between ecotypes (Jeffery, DiBacco, Van Wyngaarden et al., 2017).

The same model was recently implemented by Van Wyngaarden et al. (2017) to study effective dispersal in the sea scallop (*P. magellanicus*) along a similar range in eastern North America. However, this clinal model assumes balance between natural selection and dispersal rate, which may not be at equilibrium in the highly dynamic green crab system. Similarly, changes in linkage disequilibrium and allele frequencies over time will inevitably alter the estimated cline width, and therefore, we consider our model an approximation of contemporary rates of dispersal.

The single clinal outlier locus that aligned to a reference sequence corresponded to a Toll-like receptor (TLR) gene which is part of the innate immune system (Li et al., 2015). While speculative, the fact that this locus correlates with the environment suggests a potential relationship between different TLR alleles and latitude in green crabs, possibly due to differences in marine parasite species richness at different latitudes (Rohde, 2005; Rohde & Heap, 1998). For example, unique and highly conserved TLRs were revealed within the Atlantic cod genome (Star et al., 2011), a species which has also been shown to show genomic differences associated with temperature across a latitudinal range (Bradbury et al., 2010). Marine pathogens are sensitive to temperature, and disease impacts are predicted to increase in severity as the ocean warms (Harvell et al., 2002). As such, aquatic species with latitudinal variation in alleles associated with immune function, such as we suggest for green crab and has been shown for Atlantic salmon (Dionne, Miller, Dodson, Caron, & Bernatchez, 2007), may show different responses to a changing climate. Without data from the crabs' native range and a comprehensive genome sequence, we cannot determine whether this outlier-associated locus is the result of balancing or divergent selection, a possible adaptation carried over from the crabs' native range, or perhaps the result of a transient epizootic event leading to intermittent selection for alleles of this locus which happen to correspond to different environments (Agrawal, 2011). The fact that only one match was detected by Blast2GO suggests a lack of genomic resources available within crustaceans, or perhaps the complex demographic history of green crab complicates the identification of genomic regions currently experiencing selection, and could result in poor initial outlier detection. Nevertheless, additional genomic data from green crab and other decapods would undoubtedly assist in understanding the genes and genomic architecture involved in the delineation of the two invasive ecotypes.

#### 4.4 | Implications

Understanding the broad- and fine-scale environmental factors that influence an invasive species' distribution on a temporal scale and at different life stages is necessary to model the range expansion and ecological or economic impacts of that species over time (Cosham et al., 2016). Overall, our results suggest that winter sea surface temperature could be an important environmental variable to predict any future expansion of green crab in eastern North America. While the southern ecotype of green crab is largely limited to south of and including the Bay of Fundy, climatic predictions of warming sea

surface temperatures in the region could allow for displacement of the northern ecotype. At the same time, the northern ecotype is currently found as far north as the north coast of Newfoundland, and further northward spread is either limited by dispersal ability, including less frequent anthropogenic transport north of Newfoundland and the natural southwestern current flow (Pringle, Blakeslee, Byers, & Roman, 2011), or sea surface temperature. To date, green crab have been treated as a single population or lineage within Canada for monitoring strategies (Therriault, Herborg, Locke, & McKindsey, 2008), but evidence here suggests that different ecotypes should be modeled separately to determine potential for range expansion and impacts on coastal ecosystems. While this study only considered annual and seasonal temperature and salinity, other factors including ocean currents (e.g., Benestan et al., 2016), water depth, and fine-scale biotic interactions (Cosham et al., 2016; Rosson et al., 2012) could influence selection but were not considered as the focus of this study was on the factors that may influence broad-scale population structuring of northern and southern ecotypes.

#### 4.5 | Limitations

Although we were able to detect genetic-environment associations between a small proportion of outlier loci and season temperatures, we cannot conclude which processes have led to this observation without data from the crabs' native range, a well-annotated genome, and a comprehensive understanding of the temporal dynamics of the green crab system. As mentioned, it is likely green crab in their native range are adapted to a range of ocean temperatures from North Africa to Norway, and may show both genetic variation and plasticity in thermal tolerances. The recent introductions of different lineages to North America, and even more recent secondary contact between lineages, appear to simply reflect the native population structure, and a clinal gradient in structure is maintained by a corresponding gradient in temperature, in particular winter bottom temperature. Data from populations in the native range at different latitudes could aid in determining how seasonal minima, such as winter bottom temperature, might also constrain population structure in a naturally replicated system (see Schmidt et al., 2008). Finally, additional genomic resources from green crab, other decapods, or crustaceans in general, including annotated genomes or transcriptomic data, could lead to the identification of other gene regions that show relationships between allele frequency and environment, shedding light on the processes which have led to this observed population structure and any reproductive barriers associated with constraints in gene flow.

#### 4.6 | Summary

The population structure of two green crab ecotypes and hybrid populations in the Northwest Atlantic Ocean is driven on a broad scale by a gradient in seasonal temperature minima, leading to reduced gene flow and constraints in the ranges of two distinct ecotypes which evolved along a latitudinal gradient in Europe.

Winter sea surface temperature was observed to show the strongest relationship with crab genetic spatial structure, suggesting that the continued expansion of the separate ecotypes is limited by the minimum temperature each year. This broad understanding of how the environment limits the genetic spatial structure of invasive green crab can be implemented in models predicting the spread and impacts of this species over time, and allow for projections of range expansion under climate change scenarios.

## ACKNOWLEDGEMENTS

The authors thank staff at Fisheries and Oceans Canada for their assistance with sample collection, in particular Jennifer FitzGerald, Renée Bernier, Erica Watson, Sophie Boudreau, Roland Hagan, and Dr. April Blakeslee (East Carolina University). They also wish to thank the staff of the Aquatic Biotechnology Lab at the Bedford Institute of Oceanography for assistance with DNA extractions and RAD library preparation. Funding was provided to IRB and CDB by the Genomic Resource Development Initiative, Fisheries and Oceans Canada. This work was supported by a National Sciences and Engineering Research Council of Canada (NSERC) visiting fellowship to NWJ and funding provided by the Canadian Healthy Oceans Network (CHONe) to IRB and CDB.

## DATA ARCHIVING

Data for this study are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.bq27d>.

## ORCID

Nicholas W. Jeffery  <http://orcid.org/0000-0003-4242-5712>

Ian R. Bradbury  <http://orcid.org/0000-0002-8152-4943>

## REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Agrawal, A. A. (2011). Current trends in the evolutionary ecology of plant defence. *Functional Ecology*, *25*, 420–432. <https://doi.org/10.1111/j.1365-2435.2010.01796.x>
- Audet, D., Davis, D. S., Miron, G., Moriyasu, M., Benhalima, K., & Campbell, R. (2003). Geographical expansion of a nonindigenous crab, *Carcinus maenas* (L.), along the Nova Scotian shore into the southeastern Gulf of St. Lawrence, Canada. *Journal of Shellfish Research*, *22*, 255–262.
- Barney, B. T., Munkholm, C., Walt, D. R., & Palumbi, S. R. (2017). Highly localized divergence within supergenes in Atlantic cod (*Gadus morhua*) within the Gulf of Maine. *BMC Genomics*, *18*, 271. <https://doi.org/10.1186/s12864-017-3660-3>
- Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), *Hybrid zones and the evolutionary process* (pp. 13–45). New York, NY: Oxford University Press.
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, *24*, 3299–3315. <https://doi.org/10.1111/mec.13245>
- Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., ... Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, *25*, 5073–5092. <https://doi.org/10.1111/mec.13811>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*, 289–300.
- Berg, P., Star, B., Pampoulie, C., Bradbury, I., Bentzen, P., Hutchings, J., ... Jakobsen, K. (2017). Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity*, *119*, 418. <https://doi.org/10.1038/hdy.2017.54>
- Berrill, M. (1982). The life cycle of the green crab *Carcinus maenas* at the northern end of its range. *Journal of Crustacean Biology*, *2*, 31–39. <https://doi.org/10.2307/1548108>
- Best, K., McKenzie, C. H., & Couturier, C. (2017). Reproductive biology of an invasive population of European green crab, *Carcinus maenas*, in Placentia Bay, Newfoundland. *Management of Biological Invasions*, *8*, 247–255. <https://doi.org/10.3391/mbi>
- Beukema, J. (1991). The abundance of shore crabs *Carcinus maenas* (L.) on a tidal flat in the Wadden Sea after cold and mild winters. *Journal of Experimental Marine Biology and Ecology*, *153*, 97–113. [https://doi.org/10.1016/S0022-0981\(05\)80009-7](https://doi.org/10.1016/S0022-0981(05)80009-7)
- Bierne, N. (2010). The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution*, *64*, 3254–3272. <https://doi.org/10.1111/j.1558-5646.2010.01050.x>
- Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*, 2044–2072. <https://doi.org/10.1111/j.1365-294X.2011.05080.x>
- Blakeslee, A. M. H., McKenzie, C. H., Darling, J. A., Byers, J. E., Pringle, J. M., & Roman, J. (2010). A hitchhiker's guide to the Maritimes: Anthropogenic transport facilitates long-distance dispersal of an invasive marine crab to Newfoundland. *Diversity and Distributions*, *16*, 879–891. <https://doi.org/10.1111/j.1472-4642.2010.00703.x>
- Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., ... Hardie, D. C. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society of London B: Biological Sciences*, *277*, 3725–3734. <https://doi.org/10.1098/rspb.2010.0985>
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Cosham, J., Beazley, K. F., & McCarthy, C. (2016). Environmental factors influencing local distributions of European green crab (*Carcinus maenas*) for modeling and management applications. *Environmental Reviews*, *24*, 1–9.
- Craig, J., & Colbourne, E. B. (2004). *Temperature variability in the coastal waters of eastern Newfoundland: Fisheries & Oceans Canada, Science, Canadian Science Advisory Secretariat*.
- Darling, J. A., Bagley, M. J., Roman, J., Tepolt, C. K., & Geller, J. B. (2008). Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Molecular Ecology*, *17*, 4992–5007. <https://doi.org/10.1111/j.1365-294X.2008.03978.x>
- Darling, J. A., Tsai, Y. H., Blakeslee, A. M., & Roman, J. (2014). Are genes faster than crabs? Mitochondrial introgression exceeds larval dispersal during population expansion of the invasive crab *Carcinus maenas*. *Royal Society Open Science*, *1*, 140202. <https://doi.org/10.1098/rsos.140202>

- Delaunay, B. (1934). Sur la sphere vide. *Izv. Akad Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7, 1–2.
- Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, 14, 652–663. <https://doi.org/10.1111/1755-0998.12209>
- Dionne, M., Miller, K. M., Dodson, J. J., Caron, F., & Bernatchez, L. (2007). Clinal variation in MHC diversity with temperature: Evidence for the role of host–pathogen interaction on local adaptation in Atlantic salmon. *Evolution*, 61, 2154–2164. <https://doi.org/10.1111/j.1558-5646.2007.00178.x>
- Endler, J. A. (1977). *Geographic variation, speciation, and clines*. Princeton, NJ: Princeton University Press.
- Fitzpatrick, B. M., Johnson, J. R., Kump, D. K., Smith, J. J., Voss, S. R., & Shaffer, H. B. (2010). Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences*, 107, 3606–3610. <https://doi.org/10.1073/pnas.0911802107>
- Frichot, E., & François, O. (2015). LEA: An R package for Landscape and Ecological Association studies. *Methods in Ecology and Evolution*, 6, 925–929. <https://doi.org/10.1111/2041-210X.12382>
- Frichot, E., Schoville, S. D., Bouchard, G., & François, 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>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33, 1.
- Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22, 1–19.
- Grosholz, E. D. (1996). Contrasting rates of spread for introduced species in terrestrial and marine systems. *Ecology*, 77, 1680–1686. <https://doi.org/10.2307/2265773>
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, 25, 2454–2466. <https://doi.org/10.1111/mec.13582>
- Harrisson, K., Amish, S., Pavlova, A., Narum, S., Telonis-Scott, M., Rourke, M., ... Ingram, B. (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>
- Harvell, C. D., Mitchell, C. E., Ward, J. R., Altizer, S., Dobson, A. P., Ostfeld, R. S., & Samuel, M. D. (2002). Climate warming and disease risks for terrestrial and marine biota. *Science*, 296, 2158–2162. <https://doi.org/10.1126/science.1063699>
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247–276. <https://doi.org/10.1111/j.1095-8312.1996.tb01434.x>
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–913. <https://doi.org/10.1038/35016000>
- 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: Pitfalls, practical solutions, and future directions. *The American Naturalist*, 188, 379–397. <https://doi.org/10.1086/688018>
- Jeffery, N. W., DiBacco, C., Van Wyngaarden, M., Hamilton, L. C., Stanley, R. R., Bernier, R., ... Bradbury, I. R. (2017). RAD sequencing reveals genomewide divergence between independent invasions of the European green crab (*Carcinus maenas*) in the Northwest Atlantic. *Ecology and Evolution*, 7, 2513–2524. <https://doi.org/10.1002/ece3.2872>
- Jeffery, N., DiBacco, C., Wringe, B., Stanley, R., Hamilton, L., Ravindran, P., & Bradbury, I. (2017). Genomic evidence of hybridization between two independent invasions of European green crab (*Carcinus maenas*) in the Northwest Atlantic. *Heredity*, 119, 154–165. <https://doi.org/10.1038/hdy.2017.22>
- Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Keller, D., Van Strien, M. J., & Holderegger, R. (2012). Do landscape barriers affect functional connectivity of populations of an endangered damselfly? *Freshwater Biology*, 57, 1373–1384. <https://doi.org/10.1111/j.1365-2427.2012.02797.x>
- Klassen, G., & Locke, A. (2007). A biological synopsis of the European green crab, *Carcinus maenas*. Canadian Manuscript Report of Fisheries and Aquatic Sciences no. 2818: vii+75 pp.
- Li, M., Li, C., Wang, J., & Song, S. (2015). Molecular characterization and expression of a novel Toll gene from the swimming crab *Portunus trituberculatus*. *Molecular Immunology*, 67, 388–397. <https://doi.org/10.1016/j.molimm.2015.06.028>
- Lichstein, J. W. (2007). Multiple regression on distance matrices: A multivariate spatial analysis tool. *Plant Ecology*, 188, 117–131. <https://doi.org/10.1007/s11258-006-9126-3>
- Lowen, J., McKindsey, C., Therriault, T., & DiBacco, C. (2016). Effects of spatial resolution on predicting the distribution of aquatic invasive species in nearshore marine environments. *Marine Ecology Progress Series*, 556, 17–30. <https://doi.org/10.3354/meps11765>
- 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>
- Mayr, E. (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Cambridge, MA: Harvard University Press.
- 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, 356–361. <https://doi.org/10.1111/1755-0998.12649>
- Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., & Stevens, M. H. H. (2017). *Vegan: community ecology package*. R package version 2.4-2.
- Pante, E., & Simon-Bouhet, B. (2013). Marmap: A package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS ONE*, 8, e73051. <https://doi.org/10.1371/journal.pone.0073051>
- Pringle, J. M., Blakeslee, A. M., Byers, J. E., & Roman, J. (2011). Asymmetric dispersal allows an upstream region to control population structure throughout a species' range. *Proceedings of the National Academy of Sciences*, 108, 15288–15293. <https://doi.org/10.1073/pnas.1100473108>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Pruvost, N., Hollinger, D., & Reyer, H. U. (2013). Genotype-temperature interactions on larval performance shape population structure in hybridogenetic water frogs (*Pelophylax esculentus* complex). *Functional Ecology*, 27, 459–471. <https://doi.org/10.1111/1365-2435.12049>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... Daly, M. J. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81, 559–575. <https://doi.org/10.1086/519795>
- Riginos, C., & Liggins, L. (2013). Seascape genetics: Populations, individuals, and genes marooned and adrift. *Geography Compass*, 7, 197–216. <https://doi.org/10.1111/gec3.12032>
- deRivera, C., Hitchcock, N. G., Teck, S. J., Steves, B. P., Hines, A. H., & Ruiz, G. M. (2007). Larval development rate predicts range expansion of an introduced crab. *Marine Biology*, 150, 1275–1288. <https://doi.org/10.1007/s00227-006-0451-9>



- Rohde, K. (2005). Latitudinal, longitudinal and depth gradients. In K. Rhode (Ed.), *Marine Parasitology* (pp. 347–348). Collingwood, VIC: CSIRO Publishing.
- Rohde, K., & Heap, M. (1998). Latitudinal differences in species and community richness and in community structure of meta-zoan endo-and ectoparasites of marine teleost fish. *International Journal for Parasitology*, *28*, 461–474. [https://doi.org/10.1016/S0020-7519\(97\)00209-9](https://doi.org/10.1016/S0020-7519(97)00209-9)
- Roman, J. (2006). Diluting the founder effect: Cryptic invasions expand a marine invader's range. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*, 2453–2459. <https://doi.org/10.1098/rspb.2006.3597>
- Roman, J., & Palumbi, S. R. (2004). A global invader at home: Population structure of the green crab, *Carcinus maenas*, in Europe. *Molecular Ecology*, *13*, 2891–2898. <https://doi.org/10.1111/j.1365-294X.2004.02255.x>
- Rosson, M. A., Quijón, P. A., Snelgrove, P. V., Barrett, T. J., McKenzie, C. H., & Locke, A. (2012). Regional differences in foraging behaviour of invasive green crab (*Carcinus maenas*) populations in Atlantic Canada. *Biological Invasions*, *14*, 659–669. <https://doi.org/10.1007/s10530-011-0107-7>
- Saarman, N. P., & Pogson, G. H. (2015). Introgression between invasive and native blue mussels (genus *Mytilus*) in the central California hybrid zone. *Molecular Ecology*, *24*, 4723–4738. <https://doi.org/10.1111/mec.13340>
- Saenz-Agudelo, P., Dibattista, J. D., Piatek, M. J., Gaither, M. R., Harrison, H. B., Nanninga, G. B., & Berumen, M. L. (2015). Seascape genetics along environmental gradients in the Arabian Peninsula: Insights from ddRAD sequencing of anemonefishes. *Molecular Ecology*, *24*, 6241–6255. <https://doi.org/10.1111/mec.13471>
- Sanford, E., & Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Annual Review of Marine Science*, *3*, 509–535. <https://doi.org/10.1146/annurev-marine-120709-142756>
- Sbrocco, E. J., & Barber, P. H. (2013). MARSPEC: Ocean climate layers for marine spatial ecology. *Ecology*, *94*, 979. <https://doi.org/10.1890/12-1358.1>
- Schmidt, P. S., Serrão, E. A., Pearson, G. A., Riginos, C., Rawson, P. D., Hilbish, T. J., ... Wetthey, D. S. (2008). Ecological genetics in the North Atlantic: Environmental gradients and adaptation at specific loci. *Ecology*, *89*, S91–S107.
- Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*, 198–207. <https://doi.org/10.1016/j.tree.2004.01.003>
- Selkoe, K., & Toonen, R. J. (2011). Marine connectivity: A new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, *436*, 291–305. <https://doi.org/10.3354/meps09238>
- Shanks, A. L., Grantham, B. A., & Carr, M. H. (2003). Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*, *13*, S159–S169. [https://doi.org/10.1890/1051-0761\(2003\)013\[0159:PDDATS\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2003)013[0159:PDDATS]2.0.CO;2)
- Sherman, C., Lotterhos, K., Richardson, M., Tepolt, C., Rollins, L., Palumbi, S., & Miller, A. (2016). What are we missing about marine invasions? Filling in the gaps with evolutionary genomics. *Marine Biology*, *163*, 198. <https://doi.org/10.1007/s00227-016-2961-4>
- Sotka, E. E., & Palumbi, S. R. (2006). The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology*, *87*, 1094–1103. [https://doi.org/10.1890/0012-9658\(2006\)87\[1094:TUOGCT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[1094:TUOGCT]2.0.CO;2)
- Stanley, R., & Jeffery, N. W. (2017). *CartDist: code to re-project marine least-cost distances into Cartesian coordinates*.
- Stanley, R. R., Jeffery, N. W., Wringe, B. F., DiBacco, C., & Bradbury, I. R. (2017). Genepopedit: A simple and flexible tool for manipulating multilocus molecular data in R. *Molecular Ecology Resources*, *17*, 12–18. <https://doi.org/10.1111/1755-0998.12569>
- Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., ... Sharma, A. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, *477*, 207–210. <https://doi.org/10.1038/nature10342>
- Tepolt, C. (2015). Adaptation in marine invasion: A genetic perspective. *Biological Invasions*, *17*, 887–903. <https://doi.org/10.1007/s10530-014-0825-8>
- Tepolt, C. K., & Palumbi, S. R. (2015). Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Molecular Ecology*, *24*, 4145–4158. <https://doi.org/10.1111/mec.13294>
- Tepolt, C. K., & Somero, G. N. (2014). Master of all trades: Thermal acclimation and adaptation of cardiac function in a broadly distributed marine invasive species, the European green crab, *Carcinus maenas*. *Journal of Experimental Biology*, *217*, 1129–1138. <https://doi.org/10.1242/jeb.093849>
- Therriault, T., Herborg, L., Locke, A., McKindsey, C., & Department of Fisheries and Oceans, Ottawa, ON(Canada); Canadian Science Advisory Secretariat, Ottawa, ON(Canada) (2008). *Risk assessment for European green crab (Carcinus maenas) in Canadian waters*. DFO: Ottawa, ON.
- Tyberghein, L., Verbrugge, H., Pauly, K., Troupin, C., Mineur, F., & De Clerck, O. (2012). Bio-ORACLE: A global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, *21*, 272–281. <https://doi.org/10.1111/j.1466-8238.2011.00656.x>
- Van Wyngaarden, M., Snelgrove, P. V., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., ... Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*, using RADseq-derived SNPs. *Evolutionary Applications*, *10*, 102–117. <https://doi.org/10.1111/eva.12432>
- Viard, F., David, P., & Darling, J. A. (2016). Marine invasions enter the genomic era: Three lessons from the past, and the way forward. *Current Zoology*, *62*, 629–642. <https://doi.org/10.1093/cz/zow053>
- Villemerieul, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, *6*, 1248–1258. <https://doi.org/10.1111/2041-210X.12418>
- Yamada, S. B., & Kosro, P. M. (2010). Linking ocean conditions to year class strength of the invasive European green crab, *Carcinus maenas*. *Biological Invasions*, *12*, 1791–1804. <https://doi.org/10.1007/s10530-009-9589-y>

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Jeffery NW, Bradbury IR, Stanley RRE, et al. Genomewide evidence of environmentally mediated secondary contact of European green crab (*Carcinus maenas*) lineages in eastern North America. *Evol Appl*. 2018;11:869–882. <https://doi.org/10.1111/eva.12601>