MOLECULAR ECOLOGY

Molecular Ecology (2016) doi: 10.1111/mec.13811

Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*)

LAURA BENESTAN,* BRADY K. QUINN,† HALIM MAAROUFI,‡ MARTIN LAPORTE,* FRASER K. CLARK,§¶ SPENCER J. GREENWOOD,§¶ RÉMY ROCHETTE† and LOUIS BERNATCHEZ*

*Departement de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Pavillon Charles-Eugène-Marchand 1030, Avenue de la Médecine Québec, Québec Canada, G1V 0A6, †Department of Biological Sciences, University of New Brunswick, P.O. Box 5050, Saint John, NB Canada, E2L 4L5, ‡Institut de Biologie Integrative et des Systemes (IBIS), Pavillon Charles-Eugène Marchand, 1030, Avenue de la Médecine Québec, Québec Canada, G1V 0A6, §Department of Biomedical Sciences, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada, C1A 4P3, ¶AVC Lobster Science Centre, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada, C1A 4P3

Abstract

Investigating how environmental features shape the genetic structure of populations is crucial for understanding how they are potentially adapted to their habitats, as well as for sound management. In this study, we assessed the relative importance of spatial distribution, ocean currents and sea surface temperature (SST) on patterns of putatively neutral and adaptive genetic variation among American lobster from 19 locations using population differentiation (PD) approaches combined with environmental association (EA) analyses. First, PD approaches (using BAYESCAN, ARLEQUIN and OUTFLANK) found 28 outlier SNPs putatively under divergent selection and 9770 neutral SNPs in common. Redundancy analysis revealed that spatial distribution, ocean current-mediated larval connectivity and SST explained 31.7% of the neutral genetic differentiation, with ocean currents driving the majority of this relationship (21.0%). After removing the influence of spatial distribution, no SST were significant for putatively neutral genetic variation whereas minimum annual SST still had a significant impact and explained 8.1% of the putatively adaptive genetic variation. Second, EA analyses (using Pearson correlation tests, BAYESCENV and LFMM) jointly identified seven SNPs as candidates for thermal adaptation. Covariation at these SNPs was assessed with a spatial multivariate analysis that highlighted a significant temperature association, after accounting for the influence of spatial distribution. Among the 505 candidate SNPs detected by at least one of the three approaches, we discovered three polymorphisms located in genes previously shown to play a role in thermal adaptation. Our results have implications for the management of the American lobster and provide a foundation on which to predict how this species will cope with climate change.

Keywords: candidate gene, larval dispersal, marine invertebrate, RADseq, seascape genetics, thermal adaptation

Received 24 May 2016; revision received 9 August 2016; accepted 16 August 2016

Introduction

Incorporating environmental information into a population genetics framework is essential to identify the proximal factors that modulate the strength and interactions of evolutionary forces, which ultimately determine the extent and scale of local adaptation of living organisms (Manel & Segelbacher 2009). Towards this end, the field of landscape genetics aims to assess how environmental parameters influence the extent of genetic variation within and among populations (Manel *et al.* 2003).

Correspondence: Laura Benestan, Fax: 418 656 7176; E-mail: laura.benestan@icloud.com

While landscape genetic studies of terrestrial species have been flourishing over the last decade (Manel & Holderegger 2013), the number of studies investigating marine species in a 'seascape genetics' framework has been more limited (Storfer *et al.* 2006; Riginos & Liggins 2013; Kershaw & Rosenbaum 2014).

Marine species are typically characterized by the absence of obvious physical barriers to gene flow over large geographic distances (Palumbi 1994). However, dispersal potential may vary across a fragmented seascape due to patterns and gradients of environmental factors such as ocean currents, temperature and salinity. In particular, over the past 5 years, seascape genetic studies have shown that complex patterns of genetic connectivity are related to larval connectivity estimates based on ocean currents in a wide range of marine species (reviewed in Selkoe *et al.* 2016), including mussels [*Mytilus sp.*: (Gilg & Hilbish 2003)], urchins [*Centrostephanus rodgersii*: (Banks *et al.* 2007)], corals [*Acropora palmata*: (Baums *et al.* 2006)], barnacles [*Balanus glandula*: (Galindo *et al.* 2006)], snails [*Kelletia kelletii*: (White *et al.* 2010)], California spiny lobster [*Panulirus interruptus*: (Iacchei *et al.* 2013)], New Zealand rock lobster [*Jasus edwardsii:* (Thomas & Bell 2013)], crabs [*Carcinus aestuarii*: (Schiavina *et al.* 2014)], reef fish [*Elacatinus lori*: (D'Aloia *et al.* 2013)] and shrimp [*Pandalus borealis*: (Jorde *et al.* 2015)]. However, most of these studies did not consider the potential impacts of environmental factors on adaptive genetic variation (but see Pujolar *et al.* 2014; Tepolt & Palumbi 2015). An 'adaptive' perspective is desirable, given that the key questions of how and where gene flow is constrained are tightly linked to the fitness of individuals in their environment (Lenormand 2002). Therefore, elucidating the environmental determinants of population structure and local adaptation in marine ecosystems is a worthy enterprise that is needed to answer important questions of relevance facing marine conservation and management (Selkoe *et al.* 2008, 2016).

Investigating putatively adaptive genetic variation along environmental gradients in several populations represents a promising way to screen for evidence of local adaptation over large geographic areas (Nielsen 2005; Savolainen *et al.* 2013). The potential explanatory power of such investigation has been substantially enhanced by the development of increasingly affordable genomic tools for next-generation sequencing (Willette *et al.* 2014). To date, only a few seascape studies have taken advantage of these tools to explore both adaptive and neutral genetic patterns in marine species (Gagnaire *et al.* 2012; Bourret *et al.* 2013, 2014; Hess *et al.* 2013; Guo *et al.* 2015; Tepolt & Palumbi 2015).

The American lobster (*Homarus americanus*) supports the most important fishery in Canada (http://www.dfo-mpo.gc.ca). Consequently, sustainability of this fishery is a major concern for fishers and managers. Implementing sustainable management procedures requires an accurate description of population structure (Reiss *et al.* 2009). This need led to previous studies that documented neutral genetic structure of this species by means of microsatellites (Kenchington *et al.* 2009) and more recently by RAD sequencing (Benestan *et al.* 2015). Both studies detected the existence of two genetic clusters separating northern and southern samples of this species. These genetic clusters coincide with the occurrence of a discontinuity in larval exchange between these two regions, which suggests that ocean currents may promote 'neutral' genetic divergence in this species (see Appendix S1, Supporting information). In addition, this species' range spans a strong thermal gradient (Aiken & Waddy 1986), but the possibility of adaptive differentiation among populations associated with this environmental gradient remains to be tested. Documenting adaptive genetic structure will augment our understanding of conservation units based on neutral genes and may help establish effective conservation strategies (Allendorf *et al.* 2010). In particular, identifying the genetic basis of local adaptation to temperature is a major goal of conservation biology as it could help predict how a species will respond to climate change (Savolainen *et al.* 2013).

Temperature represents a key selective agent that appears to drive adaptive divergence among populations of many marine invertebrate species (Sanford & Kelly 2011). This is likely the case for the American lobster, which has a broad distribution along the Atlantic coast of North America, from 35.25°N in Cape Hatteras, North Carolina, to 51.73°N in the Strait of Belle Isle, Labrador (Lawton & Lavalli 1995). American lobsters are exposed to temperatures as low as −1 °C and as high as 26 °C (Aiken & Waddy 1986; Quinn & Rochette 2015). Temperature has been shown to be an important determinant of metabolism (Qadri *et al.* 2007), behaviour (Crossin *et al.* 1998) and several life history traits of this species (Lawton & Lavalli 1995). In particular, sea surface temperature (SST) during summer months is critically important to lobster larvae, affecting their survival, development and distance dispersed after hatching (MacKenzie 1988; Quinn *et al.* 2013).

Studies that searched for evidence of adaptive genetic variation have mostly used traditional population differentiation (PD) approaches (Jensen *et al.* 2016), which aim to identify loci putatively under selection by comparing the genetic differentiation index (F_{ST}) of each locus to values expected under a null model of neutral evolution (Francois *et al.* 2016; Jensen *et al.* 2016). One advantage of this approach is that it does not require a priori information concerning the environmental forces

that act as selective pressures. Environmental association (EA) analyses represent an alternative and/or complementary avenue to PD approaches that may allow detecting adaptive patterns missed by PD methods (Pritchard *et al.* 2010; Rellstab *et al.* 2015; Francois *et al.* 2016) insofar as the environmental variables investigated are relevant to genetic structure. They identify adaptive genetic variation by seeking correlations between environmental variables and allele frequencies (reviewed in (Rellstab *et al.* 2015). Both PD and EA approaches are prone to false positive associations (Frichot *et al.* 2012; Rellstab *et al.* 2015; Villemereuil & Gaggiotti 2015; Francois *et al.* 2016), but they can each detect loci under selection not identified by the other approach. Combining PD and EA approaches may thus provide an efficient strategy to identify patterns and causes of local adaptation (Gagnaire *et al.* 2015; Rellstab *et al.* 2015) while guarding against false positives (Villemereuil *et al.* 2014; Francois *et al.* 2016).

The goal of this study was to perform one of the first seascape genomics studies in a marine invertebrate by assessing the potential role of spatial distribution, ocean currents and temperature in shaping both putatively neutral and adaptive genetic structure in American lobster. We jointly performed PD analyses and EA approaches (see Methods) on samples of egg-bearing female lobsters from 19 locations spanning most of the species' range. We then applied multivariate redundancy analyses to estimate the relative contribution of spatial distribution, ocean currents, and temperature to neutral and adaptive genetic patterns. Finally, we implemented a BLAST search on the best candidate SNPs defined by both PD analyses and EA approaches to identify genes with molecular functions potentially involved in local adaptation to temperature among American lobster inhabiting different locations.

Material and methods

Sampling and genotyping

Between May and August 2012, we sampled a total of 696 egg-bearing female American lobsters from 19 locations spanning most of the species' range along a pronounced gradient of sea surface temperature (SST). We only sampled egg-bearing female as they are thought to display homing behaviour related to spawning and hatching grounds and therefore better informative about actual genetic population structure (Pezzack & Duggan 1986). Of the 19 sampling sites included in this study, 17 were previously analysed in Benestan *et al.* (2015) for other objectives than seascape genomics, namely potential for population assignment. Yet, adding two new sites led us to resume bioinformatics analyses from the beginning. Sampling, DNA extraction, RAD-sequencing library preparation, sequencing with Illumina technology and bio-informatic analyses using STACKS v. 1.09 program (Catchen *et al.* 2013) followed the methods described in Benestan *et al.* (2015). From the data set generated in that study, we developed a set of 13 688 filtered SNPs, which excluded SNPs that were not genotyped in at least 80% of the individuals and 70% of the locations, or did not show a minor allele frequency of at least 0.05 in all locations (see Table 1 and Benestan *et al.* (2015) for justification).

Population differentiation (PD) approach

We searched for loci with a level of population differentiation exceeding neutral expectations using three F_{ST}-based outlier analyses. First, we used the software OUTFLANK (Whitlock & Lotterhos 2015), which calculates a likelihood based on a trimmed distribution of F_{ST} values to infer the distribution of F_{ST} for neutral markers. OUTFLANK was run with default options (LeftTrimFraction = 0.05, RightTrimFraction = 0.05, Hmin = 0.1, 19) and identified outlier SNPs across the 19 sites based on the Q-threshold of 0.05. Second, we detected outlier SNPs with BAYESCAN v. 2.1 (Foll & Gaggiotti 2008), a Bayesian method based on a logistic regression model that separates locus-specific effects of selection ('adaptive' genetic variation) from population-specific effects of demography ('neutral' genetic variation). BAYESCAN runs were implemented using prior model (pr_odds) of 10 000, as recommended by Lotterhos & Whitlock (2015), including a total of 10 000 iterations and burn-in of 200 000 steps. Finally, we also identified outlier SNPs using ARLEQUIN v. 3.5 (Excoffier & Lischer 2010), which was run using 100 000 simulations and 1000 demes.

Table 1 Number of putative SNPs retained following each filtering step

From reads to SNPS	SNP count
Stacks catalog	199 664
Population filters	
Genotyped	
>80% of the samples	74 512
>70% of the populations	
MAF filters	
Global MAF > 0.05	18 034
Local MAF > 0.1	
Coverage filter	
From 10 to 100x	17 831
HWE filters	
F_{IS} between −0.3 and 0.3	
H_{OBS} < 0.5	13 688

ARLEQUIN is based on the infinite island model that integrates heterozygosity and simulates a distribution for neutrally distributed markers. A Q-value of 0.05 was used as threshold for statistical significance for OUTFLANK and BAYESCAN and a P of 0.05 for ARLEQUIN. All outlier analyses were conducted on the entire data set divided according to sampling location. Using the results of these three analyses, we divided our data set in two categories, putatively neutral SNPs and SNPs putatively under divergent selection (SNPs putatively under balancing selection were removed), to then infer demographic and potentially adaptive processes (Beaumont & Balding 2004). A SNP was considered as being putatively under divergent selection if all three PD identified it as an outlier.

Spatial structure and environmental factors

Spatial structure was modelled with Cartesian coordinates and distance-based Moran's eigenvector map (dbMEMs) variables obtained through a Euclidian distance matrix. These dbMEMs (hereafter spatial distribution) are independent vectors that summarize the spatial structure associated with the neighbourhood network (the distance matrix) across scales (Borcard & Legendre 2002), thereby representing a spectral decomposition of the spatial relationship among the study sites. A numerical simulation study has shown that analysis using dbMEMs is capable of detecting spatial structure at several scales, which can then be used to control for spatial correlation in tests of y~x relationships (*e.g.* genetic-environment relationships in seascape genetics; (Peres Neto & Legendre 2010). To calculate dbMEMs, we first converted degrees North latitude and West longitude to Cartesian coordinates with the *geoXY* function available in the *SoDa* package of R software v. 3.1.3 (Team R core 2014). Then, we computed a Euclidian distance matrix on the Cartesian coordinates using the *dist* function and we performed the *PCNM* function (permutations = 1000) on this matrix. The *PCNM* function, available in the *PCNM* package, transformed the spatial distances to rectangular data that are suitable for constrained ordination (Borcard & Legendre 2002).

Environmental factors considered in our seascape-genomic analyses were larval connectivity estimates based on ocean currents (see next paragraph for details) and nine estimates of sea surface temperature (SST). We considered only SST (not bottom temperature) because empirical data were not readily available for all our sampling locations at daily intervals over multiple years. In contrast, well-validated bottom temperature data were not available over the spatial and temporal domains needed in the present study. While SST directly impacts planktonic lobster larvae, bottom temperature would be more representative of potential selection acting on benthic juvenile and adult lobsters. However, SST and bottom temperature tend to be correlated over much of the geographic domain studied here (Drinkwater & Gilbert 2004; Brickman & Drozdowski 2012a). As temperature may affect different life history stages of the American lobster at different times of the year (Aiken & Waddy 1986; see also Introduction), we estimated the following nine metrics of SST: maximum, minimum, average SST from April to September (spring and summer), from October to April (fall and winter) and over the entire year. We estimated these nine SST indices for each year between 2002 and 2012 and in analyses used the average value of each index over these 11 years. The SST data for our 19 study locations were generated by the Remote Sensing Laboratory of the Maurice Lamontagne Institute and obtained from Observatoire global du Saint-Laurent-OGSL database (http://ogsl.ca), which contains geo-referenced SST along North America's coastlines with a nominal spatial resolution of 1.1 km and a 24-h update frequency. The nine temperature metrics estimated for each sampling location are included (Table S1, Supporting information).

Larval connectivity values among our sampling sites, which reflect the estimated spread of larvae from a spawning site to a settlement site as a result of ocean currents, were derived from simulations with an individual-based biophysical dispersal model of American lobster larvae (Chasse & Miller 2010) coupled to a three-dimensional physical oceanographic model (CANOPA) of the Atlantic Shelf of eastern North America (longitude: 71.5°–54.9°W; latitude: 38.6°–52.0°N; (Brickman & Drozdowski 2012b). The physical oceanographic model has a spatial resolution of 1/12° N or W (~9 km × 6 km, or 54 km^2), and simulations were run over 8 years, from 2005 to 2012. During each simulation, clusters of larvae were released every 12 h in the months of June–September, when larval release and drift occur in nature (MacKenzie 1988; Quinn *et al.* 2013; Quinn & Rochette 2015). Larvae were released in same quantity and at same time in all cells of our model domain that fell within the lobster's historical range (Pezzack 1992), with a total of 2.16×10^9 larvae released per year per ~54 km^2 cell (Quinn 2014; preliminary values based on those from Chasse & Miller 2010). Larvae drifted passively at the surface, and no mortality was included. Time spent drifting was controlled based on (i) water temperature experienced by larvae, (ii) temperature-dependent development equations derived from laboratory studies on this species (MacKenzie 1988; Quinn *et al.* 2013) and (iii) settlement beginning halfway through stage IV (Cobb *et al.* 1989) and occurring where bottom temperature was ≥10°C (Chiasson

et al. 2015). Positions of larvae within the flow field were tracked at 5-min time steps, which allowed the number and origin of settling larvae for each model cell to be determined. Additional details concerning this model can be found in Quinn (2014) and B. K. Quinn, J. Chassé, and R. Rochette (in prep).

For determination of connectivity, the model's domain was divided into 5400 km^2 geographic blocks ('source-sink areas', n = 338 total) containing 100 oceanic model cells each (see Fig. S1, Supporting information), among which dispersal probabilities were calculated (Fig. S2, Supporting information). One of the 19 study sites (named BON) fell outside the model's geographic domain and could not be used to make pairwise estimates of connectivity (Table 1, Figs. S1 and S2, Supporting information). In each year, the number of larvae released from and settling in each of the remaining 18 sites' blocks was calculated, as was the number of larvae exchanged by each pair of blocks. Larval connectivity between each pair of sites was determined based on whether dispersal probability was 1, they are said to be connected (yes, 1) or not connected (0, no) across all 8 years of model simulations (Fig. S2, Supporting information) between the two sites of a pair, and then used to calculate asymmetric eigenvector maps (AEMs). AEM is a spatial eigenfunction method developed to model multivariate (*e.g.* genetic data) spatial distributions generated by an asymmetric, directional physical process, such as current-driven larval dispersal (Blanchet *et al.* 2011). The nodes-by-edges matrix, which translates the larval connectivity matrix into a vector of weights at each site (based on the absence/presence of connectivity), was constructed with 18 nodes (*i.e.* sites) and 25 edges (*i.e.* connectivity links obtained from our matrix data). From this matrix, the calculation of AEM resulted in thirteen AEM vectors (hereafter ocean currents) reflecting the ocean currents network obtained from our biophysical dispersal model.

Redundancy analysis (RDA): linking genetic structure to environmental factors

We conducted redundancy analyses (RDA) to investigate the relative contribution of spatial distribution, ocean currents and temperature to both neutral and putatively adaptive genetic structure at all study sites except BON (outside of the model range). RDA is a direct extension of multiple regression to model multivariate response data (Legendre & Gallagher 2001). We first attempted to reveal the relationship between spatial distribution (using dbMEMs vectors) and/or ocean currents (using AEM vectors) with the neutral and adaptive genetic structure using a RDA. Then, we implemented a partial RDA, which partitioned the total explained genetic variation among spatial distribution, ocean currents and temperature (SST) to investigate the separate and joint influences of spatial distribution and environmental variables on genetic structure, thus overcoming collinearity issues. Using sampling sites as subjects, we assessed the variability in minor allele frequencies (MAF) of SNPs (response variables) that could be explained by our explanatory variables (spatial distribution, ocean currents, SST). MAF were calculated in Plink v.1.9 using all 13 688 SNPs available, as recommended by Manel & Holderegger (2013). Because the PCNM method performs better on detrended data, that is data from which the broadscale trend has been removed (Borcard & Legendre 2002), we applied a detrending on the raw MAF data using the *decostand* function with the hellinger method available in *vegan* package in R (Oksanen *et al.* 2007). Next, we performed principal component analysis (PCA) of the MAF data and only retained the meaningful principal component factors (PCs) with eigenvalues >1, according to the Kaiser–Guttman criterion (Yeomans & Golder 1982). The independent parameters that best explained variability in the PC factors were selected through a stepwise procedure elimination (forward and backward giving similar results) using the *ordistep* function available in the *vegan* package. The *ordistep* function selects variables to build the 'optimal' model, which is the model with the highest adjusted coefficient of determination (R_{adj}^2). In the case of the partial RDAs, the effect of spatial distribution on genetic structure was 'subtracted' (dbMEM vectors used as covariables) and constrained ordination was performed on the residual variability of the genetic data.

For all four tests (neutral and adaptive data sets: RDA and partial RDA), analyses of variance (ANOVAs; 1000 permutations) were performed to assess the global significance of the RDAs and marginal ANOVAs (1000 permutations) were run to determine which environmental factors were significantly correlated with genetic variation. RDAs were computed using the *rda* function available in the *vegan* package in R software. We performed an RDA and a partial RDA on all the 9770 putatively neutral SNPs and on the 28 SNPs that were identified as putatively under divergent selection by all three tests based on the PD approach (see Results). The remaining SNPs identified as being under balancing selection by BAYESCAN and ARLEQUIN were not considered further.

Environmental association (EA) approach

We used three approaches to identify a set of best candidate SNPs for local adaptation. First, we used the

Pearson test *via* the *cor.test* function available in R software and identified all SNPs that showed statistically significant associations ($P < 0.001$ and $r > 0.70$ or $r < -0.70$) between their allele frequencies and at least one of the nine temperature parameters (called COR in Fig. 2B). Then, we searched for SNP-environment associations using two environmental association programs that take into account population genetic structure: BAYESCENV (Villemereuil & Gaggiotti 2015) and LFMM v.1.4 (Frichot *et al.* 2013). BAYESCENV detects genetic signature of selection by identifying loci that show large positive F_{ST} (outside of the neutral model F_{ST} distribution) values that are significantly correlated with environmental variables. We set the neutral model of F_{ST} distribution with $P < 0.05$. LFMM uses a hierarchical Bayesian mixed model based on the residuals of PCA to take into account population genetic structure. We applied a $P < 0.05$ with a Bonferroni correction, which corresponded to SNPs that showed a z-score higher than 5 (p = 2*pnorm(−abs(z)) as recommended by Frichot *et al.* 2013). The number of genetic clusters ($k = 2$) needed for running LFMM analyses was determined by a discriminant analysis of principal component (DAPC) described in Benestan *et al.* (2015).

We defined the set of best candidate SNPs for local adaptation to temperature as those SNPs that were found to be associated with at least one of the nine temperature parameters in all three of the analyses (COR, LFMM and BAYESCENV). We then performed a spatial principal component analysis (sPCA) on these best candidate SNPs with the R package *adegenet* (Jombart 2008), which accounted for spatial distribution and allowed us to assess whether variation in our best candidate SNPs was associated with environmental variables beyond what would be expected based on proximity of the different sites alone. For the sPCA, the spatial proximity network among localities was built using the neighbourhood-by-distance method based on latitude and longitude data. Then, we extracted the 'locality scores' of this sPCA, which reflect genetic variability linked to spatial structure among sites, and used these to transform the genetic variation of candidate SNPs into a multilocus geographic cline.

Gene ontology

We attempted to detect whether any of the 505 candidate SNPs (belonging to 432 candidate sequences) identified as potentially adaptive matched any of those listed in the SWISS-PROT database (Bairoch & Apweiler 2000). First, we BLASTed all 432 candidate sequences containing the 505 candidate SNPs against the complete transcriptome of the American lobster (Fraser Clark and Spencer Greenwood, University of Prince Edward Island, *personal communication*). As the RAD candidate sequences were only 90 bp in length, this step helped increase the length of the RAD candidate sequences and hence reduced the number of false positives found when performing a BLAST search of these query sequences on a gene database. After this initial BLAST-screen, we found a total of 122 contigs (extracted from the complete transcriptome data) that contained the 432 candidate sequences. These contigs were used as query sequences in a more stringent BLAST search conducted on the well-annotated SWISS-PROT database. Minimal *E*-value threshold of 1×10^{-6} and a homology of sequences of more than 70% were required for our BLAST analysis. This yielded a set of candidate SNPs successfully identified as belonging to known genes, giving in the SWISS-PROT database. The codon containing the SNP was identified in the contig sequence translated in the six reading frames. To ascertain whether a given mutation was synonymous or nonsynonymous, the codon containing the SNP variants was translated into an amino acid according to the location of the start codon. Gene ontology (GO) annotation terms were then associated with the synonymous and nonsynonymous candidate SNPs.

Results

Data set definition: neutral versus putatively adaptive markers

A total of 13 688 filtered and informative SNPs within 8094 sequences were successfully genotyped from 562 egg-bearing female American lobsters (Table S1, Supporting information). The number of SNPs per sequence ranged from 1 to 7, with about 48.5% of the sequences containing 1 or 2 SNPs. Missing genotype data per SNP averaged 7.2%. BAYESCAN detected 10 544 SNPs (77.0%) putatively neutral, 3119 SNPs (22.8%) putatively under balancing selection and 35 SNPs (0.2%) putatively under divergent selection, at the 5% significance level. Based on the q-value model, we found 22 SNPs showing decisive evidence for selection with a Bayes factor >100 (Fig. 1). ARLEQUIN identified 12 275 putatively neutral SNPs (89.7%), 164 SNPs putatively under divergent selection (1.2%) and excluded 1249 SNPs (9.1%) due to too much missing genotype data. At the same significance level ($P < 0.05$), OUTFLANK identified 41 SNPs under divergent selection. BAYESCAN, OUTFLANK and ARLEQUIN analyses shared 28 SNPs identified as being putatively under divergent selection (Fig. 2A) for which F_{ST} values varied between 0.0321 and 0.1780 among the 19 sampling sites compared to an average F_{ST} value of 0.0018 over all markers. These 28 candidate SNPs

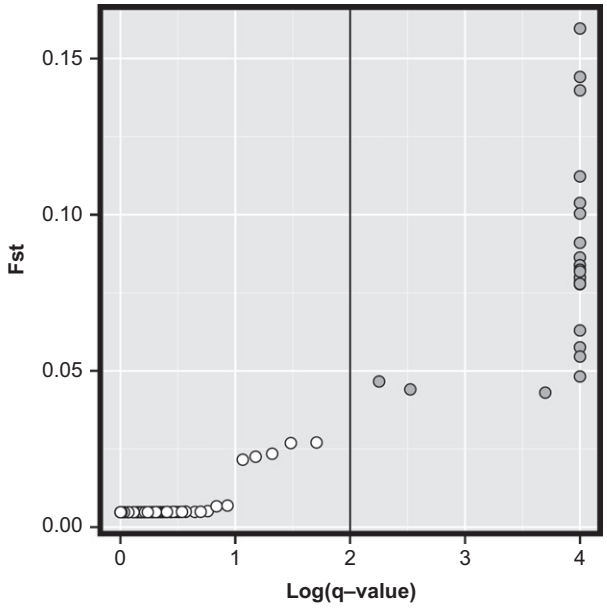

Fig. 1 Bayesian test for selection on individual SNPs in BAYESCAN v. 1.21. SNPs to the right of the vertical black line represent outliers with a Bayes factor >100 (Log (Q-value) > 2).

were used for downstream analyses of adaptive genetic structure. Similarly, we used the 9770 putatively neutral SNPs detected by both BAYESCAN and ARLEQUIN for downstream analyses of neutral genetic structure.

Environmental factors shaping neutral and adaptive genetic structure

Based on the Kaiser–Guttman criterion, 10 PCs were meaningful and kept for the 9770 putatively neutral SNPs, which accounted for more than 70.0% of the total putatively neutral genetic variation. For this putatively neutral genetic variation, one temperature descriptor (maximum annual winter temperature), two geographic vectors (dbMEM-1 and dbMEM-3; Table 2) and five vectors representing a network of ocean currents (AEM-1, AEM-2, AEM-4, AEM-7 and AEM-9; Table 2) were selected by the *ordistep* function and included in the RDA framework. The RDA was globally significant ($P = 0.001$) with an adjusted coefficient of determination (R_{adj}^2) of 0.317. The first two axes of the RDA accounted for 16.7% and 10.8% of the genetic variation, respectively. By considering the most explanatory independent parameters selected by the *ordistep* function, the marginal ANOVA showed that one geographic vector (dbMEM-1) and four vectors representing ocean current networks (AEM-1, AEM-4, AEM-7 and AEM-9) were all significant predictors of the putatively neutral genetic variation ($P < 0.05$; Table 2). When partitioning the relative importance of spatial distribution and ocean currents on neutral genetic variation (partial RDA), spatial distribution (dbMEM-1) and ocean currents (AEM-1, AEM-4, AEM-7 and AEM-9) were both still significant, but variation explained by ocean currents was three times (21.0%) that explained by spatial distribution (7.6%) (Table 2).

For the analysis based on the 28 SNPs putatively under divergent selection, we retained five PCs based on the Kaiser–Guttman criterion, which together accounted for 78.5% of the putatively adaptive genetic variation. Here, three temperature descriptors (mean summer, minimum annual and maximum annual SST) and one geographic vector (dbMEM-1) were selected by the *ordistep* function and included in the RDA framework. The RDA was globally significant ($P = 0.004$) and revealed an adjusted coefficient of determination of 0.301. The first two axes of the RDA accounted for 35.9% and 6.5% of the genetic variation, respectively (Fig. 3). The marginal ANOVAs for the RDA indicated that minimum annual, mean summer and maximum annual SST were the most significant predictors of the

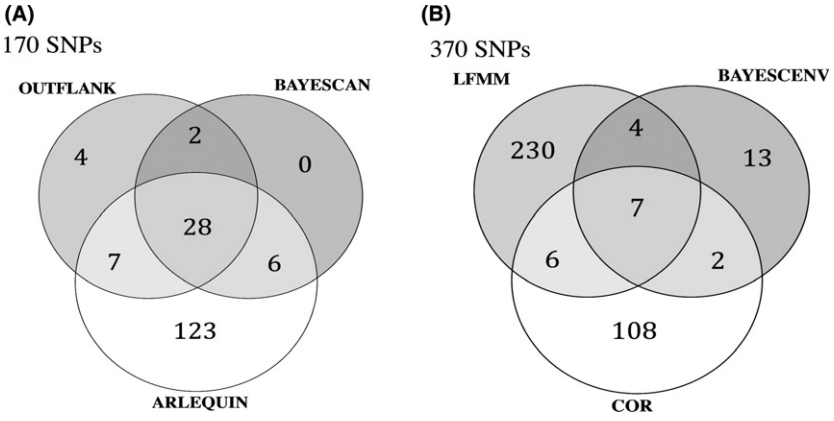

Fig. 2 Number of SNPs identified as putatively under selection using (A) three genome scan methods and (B) three environment association analyses. The total number of SNPs is reported in the upper left corner of each panel.

Table 2 RDA and partial RDA result for each response variable ('neutral' or 'adaptive' genetic variation) in relation to the explanatory variables included in the model

		Selected variables (ordistep function)				
SNPs	Analyses	Environmental	Spatial	Connectivity	*P* model	R_{adj}^2
9770 neutral	RDA	Maximum winter SST			0.001	0.317
			dbMEM-1*			
			dbMEM-3			
				AEM-1*		
				AEM-2		
				AEM-4***		
				AEM-7***		
				AEM-9*		
	Partial RDA		dbMEM-1**		0.005	0.076
				AEM-1**	0.003	0.21
				AEM-4**		
				AEM-9**		
28 outliers	RDA	Minimum annual SST**			0.004	0.301
		Mean summer SST				
		Maximum annual SST				
			dbMEM-1			
	Partial RDA	Minimum annual SST***			0.001	0.081

Significant explanatory variables are indicated with the following symbols.
*$P < 0.05$.
**$P < 0.01$.
***$P = 0.001$.

putatively adaptive genetic variation ($P < 0.05$; Table 2). However, the ANOVA for the partial RDA showed that minimum annual SST was the only significant predictor of the putatively adaptive genetic variation ($R_{adj}^2 = 0.081$, $P = 0.001$) when spatial distribution was taken into account (Table 2).

Population differentiation (PD) approaches versus environmental association (EA) analyses: overlapping SNPs

The LFMM analysis identified a total of 248 SNPs showing at least one significant association with the nine temperature parameters (Table 2). BAYESCENV was markedly more conservative and identified only 26 SNPs potentially linked to temperature. Correlation tests between minor allelic frequencies (MAF) and the nine temperature parameters revealed a set of 123 SNPs showing significant associations (81 positive: $r > 0.70$; 42 negative: $r < -0.70$) with at least one of the nine temperature parameters ($P < 0.001$). We identified seven overlapping SNPs (Fig. 2B) among these three EA analyses based on different models and assumptions (LFMM, BAYESCENV and Pearson correlation test), six of which were also among the 28 common SNPs detected by the three PD programs.

Clines in allele frequency

For the sPCA at the seven putatively adaptive SNPs identified by all EA analyses, we retained only the first positive eigenvalue as an abrupt decrease in eigenvalues was observed after it (Fig. 4B), which may indicate the boundary between true patterns and noninterpretable structures. The linear regression of the genetic locality scores extracted from the sPCA against spatial distribution and environmental factors revealed that the best predictors of locality scores were minimum annual SST ($R^2 = 0.382$, $P = 0.002$) and mean winter SST ($R^2 = 0.306$, $P = 0.008$). dbMEM vectors were not significantly related to genetic locality scores ($P > 0.05$), whereas latitude and longitude were ($R^2 = 0.178$ and 0.157, $P = 0.040$ and 0.052, respectively), albeit less strongly so than the temperature (SST) parameters. Thus, the synthetic multilocus cline of allele frequency at these SNPs showed a stronger association with either minimum annual SST or mean winter SST compared to latitude and longitude.

Gene ontology

A total of 432 candidate sequences contained the 505 unique SNPs significantly associated with temperature

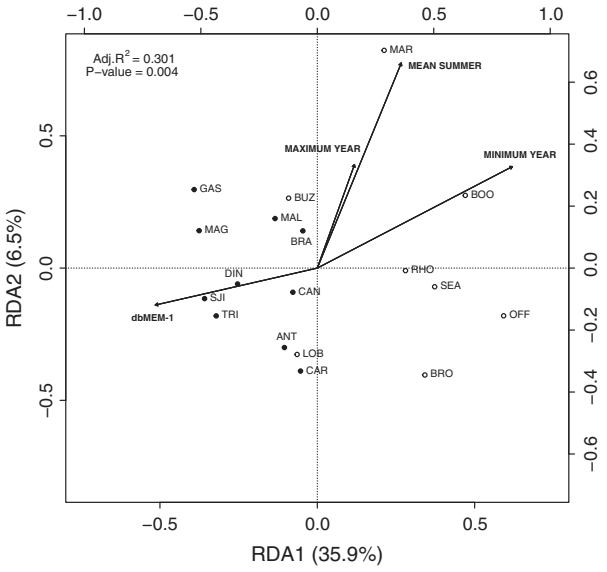

Fig. 3 Redundancy analysis (RDA) performed on the 28 SNPs putatively under divergent selection. RDA axes 1 (35.9%) and 2 (6.5%) show American lobster from 18 localities in relation to geographic vectors (dbMEM-1) and temperature descriptors (minimum annual, maximum annual and mean sea surface temperature), which are illustrated by black arrows. Lobsters from the 'south region' are in white and those from the 'north region' are in black. Positions of PC factors are according to scales on top and right axes. The RDA was globally significant and revealed an adjusted coefficient of determination of 0.301.

or defined as potentially under divergent selection by the genome scan analyses. The alignment of these candidate sequences to the complete transcriptome of the American lobster merged in a total of 122 contigs. The BLAST analysis on these 122 contigs against the SWISS-PROT database provided a total of 15 hits with an *E*-value smaller than 10^{-6}. From these 15 successfully annotated genes, five carried a nonsynonymous SNP (Table 3). Only two of these nonsynonymous SNPs—SNP *20131* and SNP *49442*—may have an impact at the protein-level as these substitutions lead to amino acid with different properties, which is not the case for the other three. The SNP *20131* is situated in the gene *GRID1*, which encodes glutamate receptor delta 1, a subunit of glutamate receptor channels that mediate most of the synaptic transmissions in the central nervous system (Guo *et al.* 2007). This mutation (Leu/Ile) is located in the extremity of the C-terminal protein that could interact with the N-ethylmaleimide-sensitive fusion (NSF) and soluble NSF attachment (SNAP) proteins, which are involved in glutamate activity. Similarly, the SNP *49442*, located in the *Vps16* gene, may interact with the SNP *20131* through the proteins NSF and SNAP (Osten *et al.* 1998). In the remaining nine synonymous polymorphisms, we also discovered the SNP *11147*, detected by the COR method ($r > 0.75$, $P < 0.001$ for mean year SST), which has a higher frequency of its alternate allele (T) in warmer populations than in colder populations (Fig. 6). Interestingly, this SNP (A/T) is located near the active site of the *β-galactosidase* gene, which produces a hydrolase enzyme well known to be involved in molecular cold adaptation processes in several organisms (Table 3; reviewed in D'Amico *et al.* 2002).

Discussion

Despite the socio-economic importance of the American lobster in the northwest Atlantic, we have very limited understanding of how the marine environment affects this species' genetic structure. In response to this knowledge gap, we conducted what may be the broadest seascape genomics study to date on a nonmodel invertebrate species. Using 13 688 RAD-sequencing markers, we applied traditional population genetics approaches (population differentiation (PD) and environmental association (EA) analyses) jointly with more general multivariate statistical frameworks (RDA and sPCA) in an attempt to gain new insights into the key determinants of genetic structure and local adaptation in this species. Our results revealed that both geographic distance but more importantly ocean currents were involved in explaining and shaping neutral genetic population structure, whereas minimum annual sea surface temperature (SST) was identified as a main potential selective agent driving local adaptation. From the combination of statistical analyses, we detected three candidate genes (*GRID 1*, *Vps16*, *β-galactosidase*), including one gene (*β-galactosidase*), with allele frequencies exhibiting a pronounced temperature-associated cline. This *β-galactosidase* gene has been identified as an important functional gene involved in cold adaptation in many microorganisms (Hoyoux *et al.* 2001; Karasova *et al.* 2002) and may play a similar role in American lobster.

Drivers of neutral and adaptive genetic structure

Marine species are typically characterized by high gene flow and weak genetic structure (Waples 1998). Nonetheless, there is a growing number of seascape studies highlighting the role of geographic distances and ocean currents in shaping patterns of marine species' population structure (White *et al.* 2010; Amaral *et al.* 2012; Iacchei *et al.* 2013; Jorde *et al.* 2015). White *et al.* (2010) highlighted the benefits of using oceanographic data to advance our ability to interpret population structure of species with pelagic larval stages and high gene flow. They demonstrated that ocean currents

Table 3 Characterization of high-quality BLAST matches obtained in comparison with American lobster RAD-sequencing SNP against American lobster transcriptome and then against SWISSPROT database

Uniprot ID	Program	Loci	Protein names	Gene names	Species	E-value	Hit length	Amino acid change	Uniprot GO
Q3LXA3	ARLEQUIN	11427	Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)	DAK	Homo sapiens	7.00E−85	573	ACC/GCC=Thr/Ala	ATP-binding, FAD-AMP lyase activity, glycerone kinase activity, metal ion binding, triokinase activity, carbohydrate and fructose metabolic processes, fructose catabolic process to hydroxyacetone phosphate and phosphorylation, glycerol metabolic process, innate immune response, regulation of innate immune response
Q62640	LFMM	**20131**	**Glutamate receptor ionotropic, delta-1**	**Grid1**	**Rattus norvegicus**	**5.00E−10**	**119**	**CTA/ATA=Leu/Ile**	**Extracellular-glutamate-gated ion channel activity, ionotropic glutamate receptor activity and ionotropic glutamate receptor signalling pathway, ion transmembrane transport, social behaviour and synaptic transmission, glutamatergic**
Q920Q4	OUTFLANK	**49442**	**Vacuolar protein**	**Vps16**	**Mus musculus**	**6.00E−120**	**454**	**ACT/CCT=Thr/Pro**	**Actin biding, endosomal transport, intracellular protein transport and regulation of SNARE complex assembly, regulation of vacuole fusion, nonatophagic, vacuole organization, viral entry into host cell**

Table 3 *Continued*

Uniprot ID	Program	Loci	Protein names	Gene names	Species	E-value	Hit length	Amino acid change	Uniprot GO
Q95SX7	LFMM	21341	Probable RNA-directed DNA polymerase from transposon BS	RTase	Drosophila melanogaster	4.00E−14	248	AAT/AGT=Asn/Ser	RNA-directed DNA polymerase activity
Q96 MW7	COR	12449	Tigger transposable element-derived protein 1	T1GD1	Homo sapiens	9.00E−08	108	GCT/GTT= Ala/Val	Tigger transposable element-derived protein 1
Q81W92	COR	**11147**	**Beta-galactosidase-like protein 2**	**GLB1L2**	**Homo sapiens**	**2.00–119**	**616**	**CTT/CTA=Leu/Leu**	**Carbohydrate metabolic processes**

The five SNPs that involve an amino acid change are listed as well as the one located in beta-galactosidase gene. SNPs in bold are located in genes with putative functions that are compatible with the hypothesis of thermal adaptive selection acting on encoded protein.

better explained genetic patterns of the whelk, *Kelletia kelletii*, than geographic distance. Similarly, another recent study by Jorde *et al.* (2015) revealed that both geographic distances and larval drift with currents help elucidate large-scale genetic differentiation patterns in northern shrimp, *Pandalus borealis.* In agreement with these studies, we found that ocean currents (21.0%) were more useful in explaining genetic structure in American lobster than geographic distances alone (7.6%).

In agreement with Benestan *et al.* (2015), the most significant Moran's eigenvector maps (dbMEM-1; Fig. 5A), which represent the influence of distances on neutral genetic structure, highlighted the North and South dichotomy resulting in two genetic groups of lobster. For both regions, the most significant asymmetrical eigenvector maps (AEM-4; Fig. 5B), representing larval dispersal within a single generation, indicated that the Gaspé and the Scotian Shelf Currents impact neutral genetic structure (Fig. 5C). Indeed, the Gaspé Current is likely to carry pelagic larvae along the Gaspé Peninsula towards the southern Gulf of St Lawrence and western coast of Cape Breton, connecting sampling sites in this area (GAS, MAL, MAG and DIN) that showed very low and nonsignificant F_{ST} values (Appendix S1, Fig. S3, Supporting information). Similarly, the Scotian Shelf current could contribute to 'homogenizing' lobsters in and near the eastern Gulf of Maine, potentially causing the lack of significant genetic divergence previously observed among offshore (OFF and BRO) and inshore (LOB; Fig. S3, Supporting information) sampling sites near the southwestern part of the Scotian Shelf. However, current-mediated drift of larvae from the Gulf of St. Lawrence to the Gulf of Maine almost never occurred within one generation (Fig. S2, Supporting information). Over multiple generations some connectivity likely occurs between these regions, following a 'stepping stone' model of gene flow, which would prevent complete isolation of lobsters in these two regions; however, this would not be enough to homogenize them, thus supporting the north–south genetic divide observed for this species (Benestan *et al.* 2015 and present study). On average, lobster larvae drift approximately 129 km between hatch and settlement, with the majority (90%) drifting ≤410 km (B. K. Quinn, J. Chassé & R. Rochette, in prep). Therefore, genetic dissimilarities observed between sites in the north and south regions, as well as between far-apart sites within these regions (e.g. TRI and GAS), are likely due, at least in part, to the limited amount of current-mediated larval exchange between them.

Importantly, these findings provide empirical support for modelled estimates of larval drift and connectivity for this species (Quinn 2014) and they demonstrate that

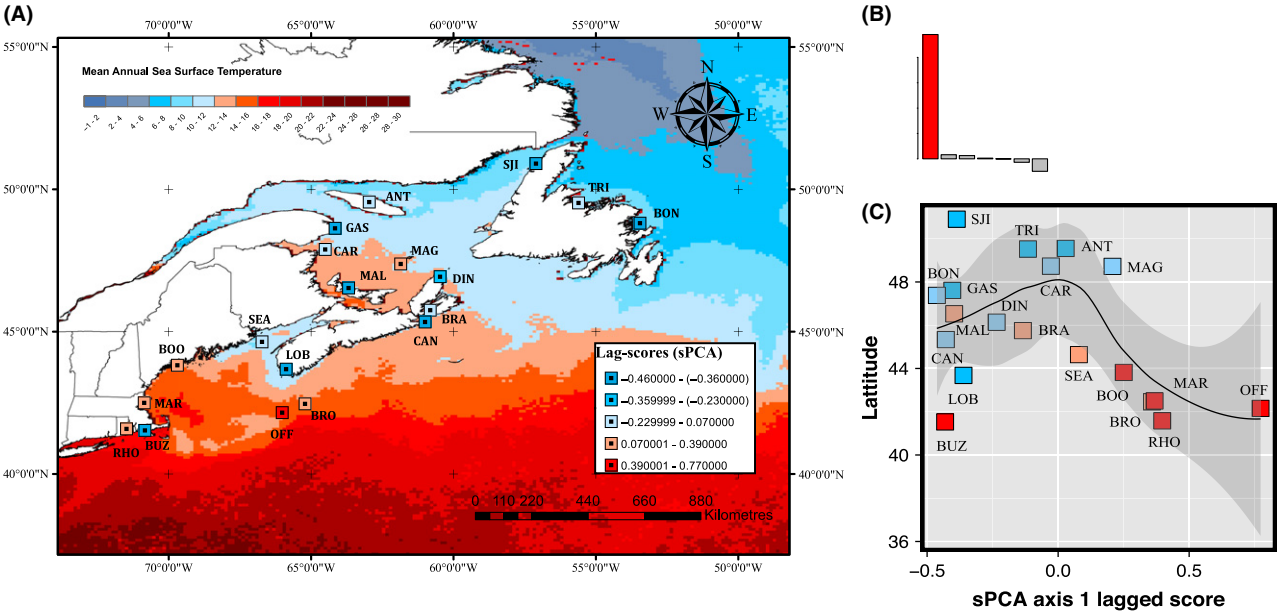

Fig. 4 (A) Synthetic multilocus putatively adaptive variation in American lobsters from 19 sampling sites. This spatial analysis was based on genetic variation at the seven SNPs that were significantly associated with explanatory variables and detected commonly using three environmental association analyses (see Fig. 2B, LFMM, BAYESCENV and COR). The 19 sampling sites are represented on the map by squares coloured according to each locality's lagged score on the first principal component. Mean annual sea surface temperatures, averaged from January to December 2012, are represented on the same colour scale. (B) Barplot of the positive and negative eigenvalues obtained when running the sPCA. Here, the first and positive eigenvalue retained is indicated in red. (C) Graphical representation of the synthetic multilocus cline considering the relationship between latitude and sPCA lagged score. The gradient of colours represents the mean annual sea surface temperature (see A).

ocean currents play a meaningful role in shaping American lobster neutral population genetic structure. Nevertheless, larval connectivity via ocean currents 'only' explained approximately 21.0% of the neutral genetic variation observed among lobsters from our 18 study locations. This could be partly due to limitations of the dispersal modelling system we used, which at present lacks some aspects of lobster biology (*e.g.* larval behaviour, mortality, egg production) that could impact dispersal patterns, but for which information from across the species' range is currently unavailable (Quinn 2014). Processes occurring at other points in the lobster's life cycle (*e.g.* movement by adults on the sea floor, postlarval swimming and settlement behaviours) might also play a role in structuring lobster populations (Campbell & Stasko 1986; Chiasson *et al.* 2015) and would thus lead to different connectivity patterns than inferred by larval dispersal alone. Additionally, processes occurring over multiple generations could lead to different patterns than those observed in single-generation simulations and should thus be comprehensively investigated in the future.

We used redundancy analysis (RDA) instead of performing a linear regression between Euclidian or oceanographic distances and F_{ST}, which has been the most common approach used in seascape studies thus far (White *et al.* 2010; Godhe *et al.* 2013; Jorde *et al.* 2015). However, the assumption of independence is violated when performing linear regressions on F_{ST} values, which may make this approach statistically inappropriate (Boldina & Beninger 2016). The approach we used overcame this issue by synthesizing multivariate genetic data (SNPs) into vectors that were compared to Moran's eigenvector map (dbMEM) of geographic distances and asymmetrical eigenvector maps (AEMs) of larval dispersal mediated by ocean currents. Moreover, these methods depicted a greater influence of ocean currents and geographic distances on genetic variation than if we had used Euclidian distances or latitude and longitude data in a linear regression analysis. Indeed, performing RDA based on latitude and longitude alone would have resulted in $R_{adj}^2 = 0.030$ (details not shown), which is twice lower than the $R_{adj}^2 = 0.076$ obtained with dbMEM variables. Our study therefore provides evidence of the relevance of considering dbMEM for future landscape studies, especially when the spatial context is potentially nonlinear (see Garroway *et al.* 2013; Breyne *et al.* 2014).

The effects of demographic history and isolation by distance on genetic variation can confound effects of

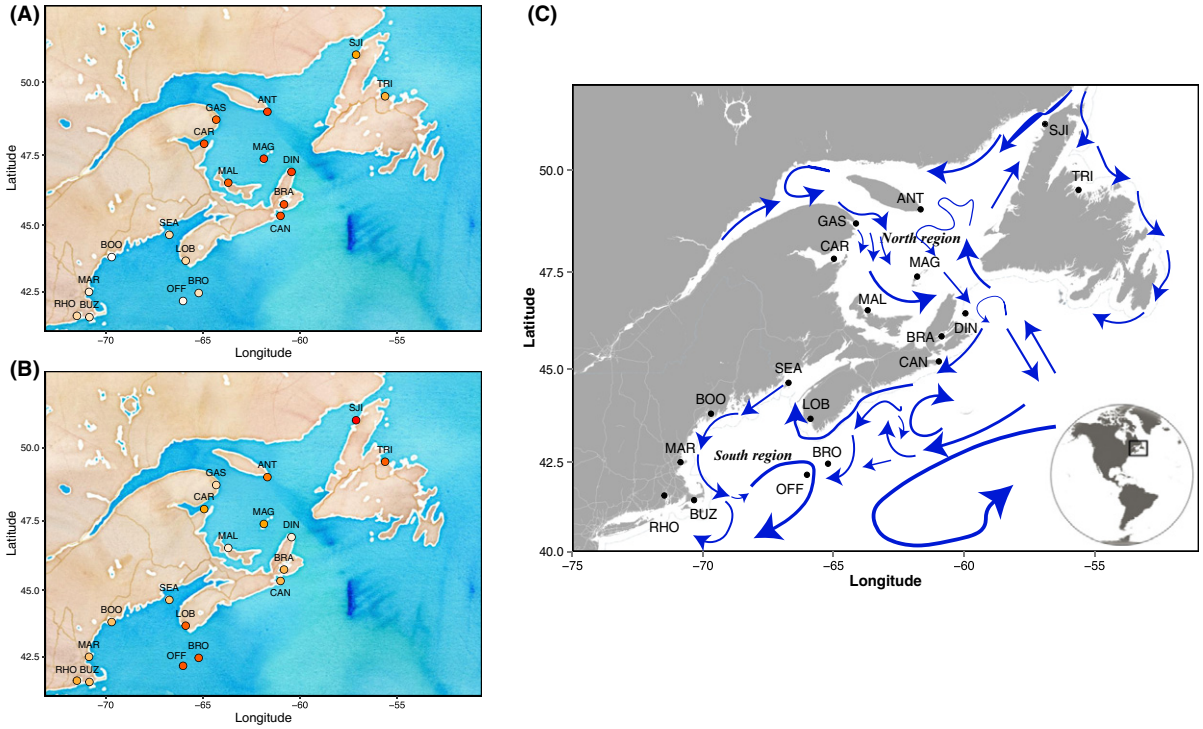

Fig. 5 Mapping of environmental data: (A) Map view of the values of the first distance-based Moran's eigenvector (dbMEM-1) attributed to each site. Color similarity represents similarity in dbMEM-1 values. (B) Map view of the values of the fourth asymmetric eigenvector maps (AEM-4), representing connectivity via larval dispersal attributed to each site. Color similarity represents similarity in AEM-4 values. (C) Map showing the 18 sampling sites of the present study related to ocean circulation (blue arrow) along the eastern seaboard of Canada with permission of Brickman & Drozdowski 2012a.

environmental variables, potentially leading to incorrect interpretations regarding local adaptation (Excoffier & Ray 2008). It is therefore important to account for the spatial distribution of populations or sample locations when attempting to assess the effect of environmental factors on genetic variation. To that end, we used a partial RDA to investigate genetic variation in lobster and found that when accounting for effects of spatial distribution of sample locations SST was not a significant explanatory variable of neutral genetic variation, whereas adaptive genetic differences were significantly related to minimum annual SST. SST likely provides the best available index of spatial variation in selection imposed by temperature on all life stages of lobsters (see Methods), and our results suggest that spatially varying selection in American lobster populations are mainly driven by minimal temperatures encountered by larval or benthic stages. Spatially varying selection is a signature of local genetic differentiation caused by disparate in situ mortalities within a single generation (Endler 1986). Spatially varying selection has been evidenced in several marine species, for example American eel [*Anguilla rostrata*: (Gagnaire *et al.* 2012; Laporte *et al.* 2016)] and acorn barnacle [*Semibalanus balanoides:* (Schmidt & Rand 2001; Véliz *et al.* 2004)]. Following the method proposed by Gagnaire *et al.* (2012), we also revealed that the genetic cline based on the seven candidate SNPs identified commonly by EA approaches was better explained by minimum annual SST (R_{adj}^2 = 0.382) than by geography (R_{adj}^2 = 0.178 for latitude and 0.157 for longitude). This suggests again that the effect of temperature prevails over that of the spatial structure alone.

Here, we highlighted that minimum annual SST may be a potential selective agent driving local adaptation. Whereas SST estimates are correlated with bottom temperatures (Drinkwater & Gilbert 2004; Brickman & Drozdowski 2012a,b), which describe the environment occupied by sampled benthic stages (adults) of the lobster life cycle, SST is most likely to be experienced by pelagic larval phase where it could be a significant source of mortality. For instance, in situ observations showed that postlarvae tend to remain in waters above 12 °C (Annis 2005) and an increase in mortality below that temperature has been documented in experimental conditions (MacKenzie 1988). It is also noteworthy that larvae originating from a cold-water region have been found to exhibit a shorter development time in cold

water than larvae originating from a warm–water region (Quinn *et al.* 2013), which may also suggest that lobsters are adapted to the thermal regime they occupy. However, minimum annual SST occurs during winter months, which is a period when the larval phase is already over. Therefore, this outcome might suggest that cold-tolerance is more important for the benthic life stages than larvae, where some juveniles/adults may be better able than others to tolerate certain low temperature and will remain in the population, through the process of natural selection.

Combining population differentiation and environmental association approaches

Detecting local adaptation occurring in complex landscapes is not optimally achieved using a single approach (Rellstab *et al.* 2015). Combining population differentiation (PD) and environmental association (EA) approaches to detect candidate loci of thermal adaptation not only reduces false positive discoveries, but also maximize our chances of detecting potential signals of selection (Francois *et al.* 2016). Recently, Vatsiou *et al.* (2016) showed that combining seven analyses for the detection of selective sweeps could greatly increase the ability to pinpoint the most likely genomic regions under selection. In this study, we employed three different analyses for each approach (Fig. 2), which led to the identification of 505 candidate SNPs, a small fraction of which (six SNPs) were identified by all six analyses. Overall, we found that EA analyses identified more candidate markers (370 SNPs) than PD analyses (170 SNPs). These outcomes are in agreement with a simulation study demonstrating that EA approach have more power to detect loci under divergent selection than PD approach (Villemereuil *et al.* 2014), which is not surprising given that the former (but not the latter) utilize environmental information (here SST) to depict signals of selection.

We found 28 candidate genes that were identified by all three PD analyses, which represent only 16.5% of all outliers detected by at least one of these analyses. The number of outliers discovered by BAYESCAN and OUTFLANK tests (36 and 41 outliers, respectively) was about four times lower than the number found by ARLEQUIN (123 outliers). This outcome is in agreement with results of a simulation studies showing that ARLEQUIN consistently found more outliers and had highest type I and type II errors in their simulation scenarios in comparison with other methods such as BAYESCAN (Narum & Hess 2011). In contrast, BAYESCAN and OUTFLANK performed much more similarly by finding 80% of the same candidate SNPs. OUTFLANK identified slightly more candidate SNPs than BAYESCAN (41 against 36) although it is supposed to have a lower false discovery rate than the latter (Whitlock & Lotterhos 2015). However, the slightly higher identification rate of OUTFLANK does not necessarily result from more false positives (type I errors) but could also be due to fewer false negatives (type II errors). In species exhibiting isolation by distance (IBD), such as American lobster, a large number of false positives may be detected when testing for SNPs under selection. In the presence of IBD, Whitlock & Lotterhos (2015) recommended using other methods (*e.g.* OUTFLANK, Fdist2, FLK) than BAYESCAN because of its higher rate of false positive in such circumstances. Here, we followed this recommendation and underlined that BAYESCAN and OUTFLANK gave very similar results in an IBD system where F_{ST} is very low, which was never shown before. Indeed, the assumptions is that BAYESCAN may handle the differences between heterozygosity among loci better in a cases of less structured populations (*e.g.* $F_{ST} < 0.005$), which was different from the system tested ($F_{ST} > 0.05$) by Whitlock & Lotterhos (2015).

We identified only a small subset of seven overlapping SNPs (1.8%) that displayed temperature-associated clines in all three of the genotype-temperature association tests we conducted. Villemereuil *et al.* (2014) similarly found on average from 1 to 5% of overlap between loci considered as positives by all three analyses they used, which were very similar to ours; LFMM, BAYESCAN (we used BAYESCENV, but results were 90% similar to those obtained with BAYESCAN) and a simple linear regression analysis (similar to our COR method). This low number of overlapping SNPs reiterates the high degree to which outcomes differ between analytical approaches. As Villemereuil *et al.* (2014) revealed that these methods tend to agree more on true positives, consistency among methods can be used to account for the errors that each analysis makes and improve the identification of true positives. Nevertheless, none of the SNPs detected by all PD and EA approaches combined was among the most likely candidate to thermal adaptation detected by the BLAST. More broadly, we found that several candidate SNPs were only detected by one analysis, including the three strongest SNPs candidate (SNP *49442*, SNP *20131* and SNP *11147*; Table 2). As each approach has its advantages and disadvantages (Rellstab *et al.* 2015), our results reiterate the importance of utilizing several analyses and approaches in the field of landscape genomics.

Finding a candidate gene for thermal adaptation

Numerous marine invertebrates have evolved biochemical adaptations to reduce the negative consequences of unfavourable changes in temperature (Hochachka & Somero 2014). By combining population PD and EA

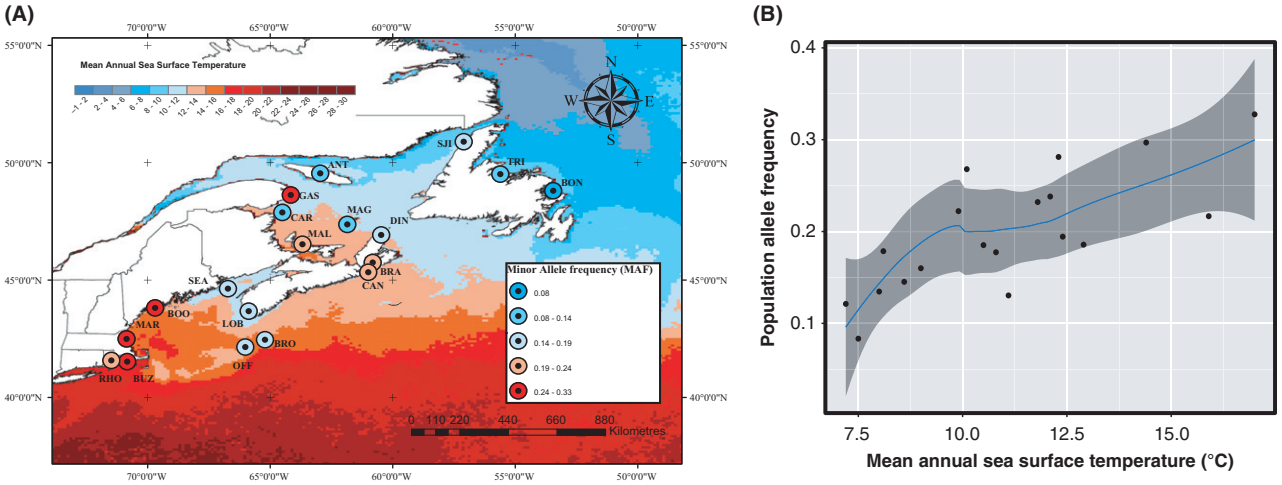

Fig. 6 Galactosidase gene characterization: (A) Map showing the minor allele frequency of the alternative allele (T) of the galactosidase gene (Haβ-GAL-1) at each of our 19 study sites in relation to the mean annual sea surface temperature (SST) (2012) over our study domain and (B) correlation between minor allele frequency of Haβ-GAL-1 and mean annual SST (2012), including loess smoothing function and confidence interval (grey area).

approaches, we identified a total of 505 SNPs as potential selection targets among the 19 sampling sites. We found only 15 SNPs in coding regions of known genes in the SWISSPROT database, which is not surprising given that the genome of the American lobster has not been sequenced and a large fraction of genes remains without any annotation (Pavey *et al.* 2012). Among these markers, we discovered two nonsynonymous polymorphisms (SNP *49442*, SNP *20131*) and one synonymous polymorphism (SNP *11147*) with putative functions that are compatible with the hypothesis of adaptive selection acting on encoded protein. The *20131* SNP is located in the *Grid1* gene, which may play key roles in synaptic plasticity (Guo *et al.* 2007) and was found to be potentially involved in high-altitude adaptation in Tibetan pigs (*Sus scrofa;* (Ai *et al.* 2014). The SNP *49442* belongs to the *Vsp16* gene, which is involved in vacuole protein sorting and organelle assembly in *Saccharomyces cerevisae* (Sato *et al.* 2000) and showed upregulated expression in sweet corn (*Zea mays*) under heat stress (Li *et al.* 2015). These findings suggest that these two SNPs may also be involved in thermal adaptation in American lobster, although more research will be needed to determine what their functions may be in this species as well as their protein structures.

The synonymous SNP *11147* is located near the active site of the β-galactosidase gene. β-galactosidases have a wide phylogenetic distribution, encompassing plants, animals and microorganisms (Wallenfels & Weil 1972). The β-galactosidase gene produces a cold-adapted enzyme, which hydrolyses lactose into galactose and glucose and has a stable enzymatic activity at temperatures below 8 °C. While the SNP *11147* is synonymous, there is a growing body of evidence demonstrating that synonymous polymorphism may face strong selection and could alter the phenotype by influencing several important cellular processes (*e.g.* transcription, splicing, mRNA transport or translation, enzyme activity and production; Plotkin & Kudla 2011). Here, for most of the sampling sites, we found that a greater proportion of individuals occupying warmer habitats had the alternate allele (T) compared to lobsters living in colder habitats (Fig. 6). Nevertheless, this is not true for four sites (CAR, GAS, OFF and BRO), where allele frequencies do not match well with the mean summer SST. We have no explanation for this gene-temperature mismatch at the CAR and GAS sites, but we can envision two reasons for the mismatch at the OFF and BRO sites. First, SST may not be the best predictor of allele frequency at these sites as they are located offshore, where lobsters occupy deeper (up to 200 m deep) and likely colder waters during summer months. However, during winter months temperature tends to actually be higher in deeper than in shallower water in this part of the species range, and adult lobsters make seasonal migrations from shallow to deep in the fall–winter to experience warmer conditions over the winter months (Robichaud & Campbell 1991). Second, it is plausible that there is a lot of mixing between animals sampled in LOB and those in OFF-BRO (see Fig. S3, Supporting information). Overall, our results suggest that functional *β-galactosidase* SNP may play a key role in the thermal adaptation of American lobster populations inhabiting varying

temperatures regimes in the northwest Atlantic. Nevertheless, the pattern we see needs to be investigated more in a future study.

Future directions

Considering processes that govern genetic structure with a broad perspective is crucial for understanding the forces that impact species' demography and evolution. Here, our best RDA model, which included spatial distribution, ocean currents and SST explained 31.7% of the neutral genetic structure. Consequently, much of this genetic structure remains unexplained. This result is comparable to that obtained by Selkoe *et al.* (2014), who found that variation in the genetic patterns of nine Hawaiian marine species cannot be explained by geography, dispersal ability and habitat factors alone (R_{adj}^2 ranging from 0.11 to 0.66). Indeed, part of the neutral genetic variation is the result of random processes (resulting from genetic drift and mutation) and it is unlikely that it would ever be possible to explain 100% of this structure. Still, other biophysical and geographic proprieties of habitats such as bathymetry, bottom temperature, productivity, salinity, colonization history, pollution and anthropogenic movements may also contribute to demographic isolation across American lobster populations. For example, previous marine genetic studies have shown that bottom temperature in northern shrimp (*Pandalus borealis*; (White *et al.* 2010; Godhe *et al.* 2013; Jorde *et al.* 2015) as well as bathymetry in cusk [*Brosme brosme*: (Knutsen *et al.* 2009)] and deep-sea sharks [*Centroscymnus coelolepis*: (Catarino *et al.* 2015)] may affect genetic structure. While these factors can also influence genetic structure in American lobster, we did not have the necessary data to test this possibility. This may help explaining the relatively low level of variance explained by our models and shows that the influence of these factors should be investigated in future studies. On the other hand, SST likely covaries with several of these other factors so the interpretation of temperature trends must be done cautiously, as is the case in any correlative study. Additionally, we only sampled females to investigate American lobster genetic structure, and thus, we have not observed genetic structure in males, which could potentially be different (*e.g.* if there is sex-biased dispersal).

Without additional resources on the American lobster genome, we were only able to produce a list of loci that are potentially under selection or linked to alleles under selection and link the variation at these loci with relevant environmental variables that provided the selective pressures (here, SST). This list of loci represents only a very small portion of the genome potentially under divergent selection; other targets of selection may have become lost when generating the libraries and sampling DNA fragments. Nonetheless, we detected three candidate genes that may have potential effects on thermal adaptation of American lobster populations. However, polymorphisms identified as potential targets of selection are usually only statistically linked with close targets of adaptive significance, and performing a site-directed mutagenesis experiment on β-galactosidase is required to draw firmer conclusions about this gene's function (Barrett & Hoekstra 2011).

Acknowledgements

This research is part of the 'Lobster Node' of the NSERC Canadian Fisheries Research Network (CFRN). We are grateful to the fishermen of the Lobster Node without whom this project would have been impossible. Project design and work plan were developed in collaboration with scientists from the Department of Fisheries and Oceans (B. Sainte-Marie, J. Chassé, M. Comeau, J. Tremblay), representatives of fishermen associations and the facilitator of the Lobster Node, M. Allain. We would like to thank A. Boudreau, V. Brzeski, Y. Carignan, Clearwater, B. Comeau, M. Comeau, JP. Allard, M. Deraspe, N. Davis, S. Delorey, C. Denton, R. Doucelle, J. Grignon, M. Haarr, R. MacMillan, G. Paulin and M. Thériault who helped to collect the samples. We are very grateful to T. Gosselin and P. Legendre for their help in bioinformatic and statistical analyses. The manuscript was improved by comments from A.L. Ferchaud. We are also grateful to Associate Editor Paul Bentzen and four anonymous reviewers for their constructive input on an earlier version. This research was funded by the NSERC CFRN. L. Benestan was supported by a doctoral fellowship from NSERC CFRN and Réseau Aquaculture Québec (RAQ) and funds from LB's Canadian Research Chair in Genomics and Conservation of Aquatic Resources. Larval dispersal modelling was carried out with contributions from J. Chassé (DFO) and computing resources provided by ACENET, the regional advanced research-computing consortium for universities in Atlantic Canada. ACENET is funded by the Canada Foundation for Innovation, the Atlantic Canada Opportunities Agency and the provinces of Newfoundland & Labrador, Nova Scotia and New Brunswick. Access to the Biomina Galaxy server hosted by the Biomina research center at the University of Antwerp, Belgium. Funding for the lobster transcriptomics program was provided by the PEI Atlantic Shrimp Corporation Inc. (12-LSC-035) and the Atlantic Lobster Sustainability Measures Program. Sequencing of the lobster transcriptome was performed by Genome Quebec.

References

Ai H, Yang B, Li J, Xie X, Chen H, Ren J (2014) Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics*, **15**, 834.

Aiken DE, Waddy SL (1986) Environmental influence on recruitment of the American lobster *Homarus americanus*: a perspective. *Canadian Journal of Fisheries and Aquatic Sciences*, **43**, 2258–2270.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709.

Amaral AR, Beheregaray LB, Bilgmann K *et al.* (2012) Seascape genetics of a globally distributed, highly mobile marine mammal: the short-beaked common dolphin (Genus Delphinus). *PLoS ONE*, **7**, e31482.

Annis ER (2005) Temperature effects on the vertical distribution of lobster postlarvae (*Homarus americanus*). *Limnology and Oceanography*, **50**, 1972–1982.

Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, **28**, 45–48.

Banks SC, Piggott LMP, Williamson JE *et al.* (2007) Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology*, **88**, 3055–3064.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767–780.

Baums IB, Paris CB, Chérubin LM (2006) A bio-oceanographic filter to larval dispersal in a reef-building coral. *Limnology and Oceanography*, **51**, 1969–1981.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

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.

Blanchet FG, Legendre P, Maranger R, Monti D, Pepin P (2011) Modelling the effect of directional spatial ecological processes at different scales. *Oecologia*, **166**, 357–368.

Boldina I, Beninger PG (2016) Strengthening statistical usage in marine ecology: linear regression. *Journal of Experimental Marine Biology and Ecology*, **474**, 81–91.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Bourret V, Kent MP, Primmer CR *et al.* (2013) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **22**, 532–551.

Bourret V, Dionne M, Bernatchez L (2014) Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Molecular Ecology*, **23**, 4444–4457.

Breyne P, Mergeay J, Casaer J (2014) Roe deer population structure in a highly fragmented landscape. *European Journal of Wildlife Research*, **60**, 909–917.

Brickman D, Drozdowski A (2012a) *Atlas of Model Currents and Variability in Maritime Canadian Waters*. Fisheries and Oceans Canada, Maritimes Region.

Brickman D, Drozdowski A (2012b) *Development and Validation of a Regional Shelf Model for Maritime Canada Based on the NEMO-OPA Circulation Model* (pp. vii+–57). Fisheries and Oceans Canada, Maritimes Region.

Campbell A, Stasko AB (1986) Movements of lobsters (*Homarus americanus*) tagged in the Bay of Fundy, Canada. *Marine Biology*, **92**, 393–404.

Catarino D, Knutsen H, Veríssimo A *et al.* (2015) The Pillars of Hercules as a bathymetric barrier to gene flow promoting isolation in a global deep-sea shark (*Centroscymnus coelolepis*). *Molecular Ecology*, **24**, 6061–6079.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Chasse J, Miller RJ (2010) Lobster larval transport in the southern Gulf of St. Lawrence. *Fisheries Oceanography*, **19**, 319–338.

Chiasson M, Miron G, Daoud D, Mallet MD (2015) Effect of temperature on the behavior of stage IV American lobster (*Homarus americanus*) larvae. *Journal of Shellfish Research*, **34**, 545–554.

Cobb JS, Wang D, Campbell DB (1989) Timing of settlement by postlarval lobsters (*Homarus americanus*): field and laboratory evidence. *Journal of Crustacean Biology*, **9**, 60–66.

Crossin G, Al-Ayoub S, Jury S, Howell W (1998) Behavioral thermoregulation in the American lobster *Homarus americanus*. *Journal of Experimental Biology*, **201**, 365–374.

D'Aloia CC, Bogdanowicz SM, Majoris JE, Harrison RG, Buston PM (2013) Self-recruitment in a Caribbean reef fish: a method for approximating dispersal kernels accounting for seascape. *Molecular Ecology*, **22**, 2563–2572.

D'Amico S, Claverie P, Collins T *et al.* (2002) Molecular basis of cold adaptation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **357**, 917–925.

Drinkwater K, Gilbert D (2004) Hydrographic variability in the waters of the Gulf of St. Lawrence, the Scotian Shelf and the eastern Gulf of Maine (NAFO Subarea 4) during 1991–2000. *Journal of Northwest Atlantic Fishery*, **34**, 85–101.

Endler JA (1986) Natural selection in the wild. *Monographs in Population Biology*, **21**, 1–336.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, **23**, 347–351.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Francois O, Martins H, Caye K, Schoville SD (2016) Controlling false discoveries in genome scans for selection. *Molecular Ecology*, **25**, 454–469.

Frichot E, Schoville S, Bouchard G, François O (2012) Landscape genomic tests for associations between loci and environmental gradients. *arXiv*, 1205.

Frichot E, Schoville SD, 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.

Gagnaire P-A, Normandeau E, Côté C, Møller Hansen M, Bernatchez L (2012) The genetic consequences of spatially varying selection in the panmictic American eel (*Anguilla rostrata*). *Genetics*, **190**, 725–736.

Gagnaire P-A, Broquet T, Aurelle D *et al.* (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, **8**, 769–786.

Galindo HM, Olson DB, Palumbi SR (2006) Seascape genetics: a coupled oceanographic-genetic model predicts population

structure of Caribbean corals. *Current Biology*, **16**, 1622–1626.

Garroway CJ, Radersma R, Sepil I *et al.* (2013) Fine-scale genetic structure in a wild bird population: the role of limited dispersal and environmentally based selection as causal factors. *Evolution*, **67**, 3488–3500.

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology*, **84**, 2989–2998.

Godhe A, Egardt J, Kleinhans D, Sundqvist L, Hordoir R, Jonsson PR (2013) Seascape analysis reveals regional gene flow patterns among populations of a marine planktonic diatom. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20131599.

Guo S-Z, Huang K, Shi Y-Y *et al.* (2007) A case–control association study between the GRID1 gene and schizophrenia in the Chinese Northern Han population. *Schizophrenia Research*, **93**, 385–390.

Guo B, DeFaveri J, Sotelo G, Nair A, Merilä J (2015) Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biology*, **13**, 19.

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR (2013) Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, **22**, 2898–2916.

Hochachka PW, Somero GN (2014) *Biochemical Adaptation*. Princeton University Press, Princeton.

Hoyoux A, Jennes I, Dubois P *et al.* (2001) Cold-adapted beta-galactosidase from the Antarctic psychrophile Pseudoalteromonas haloplanktis. *Applied and Environmental Microbiology*, **67**, 1529–1535.

Iacchei M, Ben Horin T, Selkoe KA *et al.* (2013) Combined analyses of kinship and FST suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. *Molecular Ecology*, **22**, 3476–3494.

Jensen JD, Foll M, Bernatchez L (2016) The past, present and future of genomic scans for selection. *Molecular Ecology*, **25**, 1–4.

Jombart T (2008) ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jorde PE, Søvik G, Westgaard JI *et al.* (2015) Genetically distinct populations of northern shrimp, *Pandalus borealis*, in the North Atlantic: adaptation to different temperatures as an isolation factor. *Molecular Ecology*, **24**, 1742–1757.

Karasova P, Spiwok V, Mala S, Kralova B (2002) Beta-galactosidase activity in psychrotrophic microorganisms and their potential use in food industry. *Czech Journal of Food Sciences*, **20**, 43–47.

Kenchington EL, Harding GC, Jones MW, Prodöhl PA (2009) Pleistocene glaciation events shape genetic structure across the range of the American lobster, *Homarus americanus*. *Molecular Ecology*, **18**, 1654–1667.

Kershaw F, Rosenbaum HC (2014) Ten years lost at sea: response to Manel and Holderegger. *Trends in Ecology and Evolution*, **29**, 69–70.

Knutsen H, Jorde PE, Sannæs H *et al.* (2009) Bathymetric barriers promoting genetic structure in the deepwater demersal fish tusk (*Brosme brosme*). *Molecular Ecology*, **18**, 3151–3162.

Laporte M, Pavey SA, Rougeux C *et al.* (2016) RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, **25**, 219–237.

Lawton P, Lavalli KL (1995) *Postlarval, Juvenile, Adolescent, and Adult Ecology*. Elsevier, Amsterdam.

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.

Li Y, Hu J, Liu J *et al.* (2015) Genome-wide analysis of gene expression profiles during early ear development of sweet corn under heat stress (R Tuberosa, Ed,). *Plant Breeding*, **134**, 17–27.

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24**, 1031–1046.

MacKenzie BR (1988) Assessment of temperature effects on interrelationships between stage durations, mortality, and growth in laboratory-reared *Homarus americanus* Milne Edwards larvae. *Journal of Experimental Marine Biology and Ecology*, **116**, 87–98.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution*, **28**, 614–621.

Manel S, Segelbacher G (2009) Perspectives and challenges in landscape genetics. *Molecular Ecology*, **18**, 1821–1822.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Oksanen J, Kindt R, Legendre P, O'Hara B (2007) The vegan package. *Community Ecology Package 10*.

Osten P, Srivastava S, Inman GJ *et al.* (1998) The AMPA receptor GluR2 C terminus can mediate a reversible, ATP-dependent interaction with NSF and α- and β-SNAPs. *Neuron*, **21**, 99–110.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547–572.

Pavey SA, Bernatchez L, Aubin-Horth N, Landry CR (2012) What is needed for next-generation ecological and evolutionary genomics? *Trends in Ecology & Evolution*, **27**, 673–678.

Peres Neto PR, Legendre P (2010) Estimating and controlling for spatial structure in the study of ecological communities. *Global Ecology and Biogeography*, **19**, 174–184.

Pezzack DS (1992) A review of lobster (*Homarus americanus*) landing trends in the northwest Atlantic, 1947–86. *Journal of Northwest Atlantic Fisheries Science*, **14**, 115–127.

Pezzack DS, Duggan DR (1986) Evidence of migration and homing of lobsters (*Homarus americanus*) on the Scotian Shelf. *Canadian Journal of Fisheries and Aquatic Sciences*, **43**, 2206–2211.

Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, **12**, 32–42.

Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**, R208–R215.

Pujolar JM, Jacobsen MW, Als TD *et al.* (2014) Genome-wide single-generation signatures of local selection in the panmictic European eel. *Molecular Ecology*, **23**, 2514–2528.

Qadri SA, Camacho J, Wang H *et al.* (2007) Temperature and acid-base balance in the American lobster *Homarus americanus*. *Journal of Experimental Biology*, **210**, 1245–1254.

Quinn BK(2014) *Assessing potential influence of larval development time and drift on large-scale spatial connectivity of American lobster (Homarus americanus)*. MSc thesis, University of New Brunswick, Saint John, NB.

Quinn BK, Rochette R (2015) Potential effect of variation in water temperature on development time of American lobster larvae. *ICES Journal of Marine Science: Journal du Conseil*, **72**, i79–i90.

Quinn BK, Sainte-Marie B, Rochette R, Ouellet P (2013) Effect of temperature on development rate of larvae from cold-water American lobster (*Homarus americanus*). *Journal of Crustacean Biology*, **33**, 527–536.

Reiss H, Hoarau G, Dickey Collas M, Wolff WJ (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, **10**, 361–395.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–4370.

Riginos C, Liggins L (2013) Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography Compass*, **7**, 197–216.

Robichaud DA, Campbell A (1991) Annual and seasonal size-frequency changes of trap-caught lobsters (*Homarus americanus*) in the Bay of Fundy. *Journal of Northwest Atlantic Fishery Science*, **11**, 29–37.

Sanford E, Kelly MW (2011) Local adaptation in marine invertebrates. *Annual Review of Marine Science*, **3**, 509–535.

Sato TK, Rehling P, Peterson MR, Emr SD (2000) Class C Vps protein complex regulates Vacuolar SNARE pairing and is required for vesicle docking/fusion. *Molecular Cell*, **6**, 661–671.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.

Schiavina M, Marino IAM, Zane L, Melià P (2014) Matching oceanography and genetics at the basin scale. Seascape connectivity of the Mediterranean shore crab in the Adriatic Sea. *Molecular Ecology*, **23**, 5496–5507.

Schmidt PS, Rand DM (2001) Adaptive maintenance of genetic polymorphism in an intertidal barnacle: habitat- and life-stage-specific survivorship of Mpi genotypes. *Evolution*, **55**, 1336–1344.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, **9**, 363–377.

Selkoe KA, Gaggiotti OE, Bowen BW, Toonen RJ (2014) Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology*, **23**, 3064–3079.

Selkoe KA, D'Aloia CC, Crandall ED *et al.* (2016) A decade of seascape genetics: contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, **554**, 1–19.

Storfer A, Murphy MA, Evans JS *et al.* (2006) Putting the "landscape" in landscape genetics. *Heredity*, **98**, 128–142.

Tepolt CK, Palumbi SR (2015) Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Molecular Ecology*, **24**, 4145–4158.

Thomas L, Bell JJ (2013) Testing the consistency of connectivity patterns for a widely dispersing marine species. *Heredity*, **111**, 345–354.

Vatsiou AI, Bazin E, Gaggiotti OE (2016) Detection of selective sweeps in structured populations: a comparison of recent methods. *Molecular Ecology*, **25**, 89–103.

Véliz D, Bourget E, Bernatchez L (2004) Regional variation in the spatial scale of selection at MPI* and GPI* in the acorn barnacle *Semibalanus balanoides* (Crustacea). *Journal of Evolutionary Biology*, **17**, 953–966.

Villemereuil P, Gaggiotti OE (2015) A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, **6**, 1248–1258.

Villemereuil P, Frichot E, Bazin E, Francois O, Gaggiotti OE (2014) Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, **23**, 2006–2019.

Wallenfels K, Weil R (1972) 20 β-Galactosidase. *The Enzymes*, pp. 617–663. Elsevier, Amsterdam.

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438–450.

White C, Selkoe KA, Watson J *et al.* (2010) Ocean currents help explain population genetic structure. *Proceedings Biological sciences/The Royal Society*, **277**, 1685–1694.

Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of FST. *The American Naturalist*, **186**(Suppl 1), S24–36.

Willette DA, Allendorf FW, Barber PH *et al.* (2014) So, you want to use next-generation sequencing in marine systems? Insight from the Pan-Pacific Advanced Studies Institute. *Bulletin of Marine Science*, **90**, 79–122.

Yeomans KA, Golder PA (1982) The Guttman–Kaiser criterion as a predictor of the number of common factors. *The Statistician*, **31**, 221.

L.B., R.R. and L.B. conceived and designed the study; L.B. performed sampling, libraries, bioinformatics and statistical analyses and wrote the manuscript. B.K.Q. and R.R. conceived the large-scale physical model of the American lobster. F. Clark and S. J Greewood provided the transcriptome dataset. All authors significantly contributed to improvement of statistical analyses and of the manuscript. L.B. is a PhD student in L. Bernatchez Lab and is interested in Genomics and Marine Ecology of marine organisms. B.K.Q. is a PhD student in R. Rochette Lab and is interested in Bio-physical modelling and fishery connectivity. M.L. is a postdoc in L. Bernatchez Lab and is interested in Evolutionary Biology of aquatic organisms. H.M. is a research scientist working on proteomics in the Institute of Integrative Biology. F.K.C and S.J.G. are two research scientists working on lobster in University of Prince Edward Island. R.R. is a faculty member at the University of New Brunswick, and lead PI of the CFRN Lobster Node. L.B. leads the Canadian Research Chair in Conservation Genomics and Conservation of Aquatic Resources at University Laval.

Data accessibility

DNA sequences demultiplexed with barcodes: NCBI SRA

- Bioproject Acc#: PRJNA281764
- BioSample Acc#s: SAMN03492800

The following files from this study are available from the Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.5vb8v

- *Homarus americanus*, RAD sequences for putative 13688 SNPs
- All inputs and outputs used for the PD and EA analyses
- Environmental data (spatial distribution, larval dispersal, sea surface temperature)
- All the home-made scripts used to perform the analyses

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Larval dispersal model and calculation of connectivity among areas

Figure S1 Plot of 5400 km^2 drift model "source-sink-areas" containing sampling sites used in this study.

Figure S2 Connectivity matrices among the 18/19 sample sites that fell within the drift model's domain (site BON not included).

Figure S3 Map representing the 11 differentiated genetic populations out of the 17 sampling sites analyzed by Benestan *et al.* (2015)