MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 3299–3315 doi: 10.1111/mec.13245

RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*)

LAURA BENESTAN,* THIERRY GOSSELIN,* CHARLES PERRIER,* BERNARD SAINTE-MARIE,† RÉMY ROCHETTE‡ and LOUIS BERNATCHEZ*

*Institut de Biologie Intégrative des Systèmes (IBIS), Université Laval, Québec, Quebec, Canada G1V0A6, †Pêches et Océans Canada, Institut Maurice-Lamontagne, CP 1000, Mont-Joli, Quebec, Canada G5H 3Z4, ‡Department of Biology, University of New Brunswick, PO Box 5050, Saint John, New Brunswick, Canada E2L 4L5

Abstract

Deciphering genetic structure and inferring connectivity in marine species have been challenging due to weak genetic differentiation and limited resolution offered by traditional genotypic methods. The main goal of this study was to assess how a population genomics framework could help delineate the genetic structure of the American lobster (*Homarus americanus*) throughout much of the species' range and increase the assignment success of individuals to their location of origin. We genotyped 10 156 filtered SNPs using RAD sequencing to delineate genetic structure and perform population assignment for 586 American lobsters collected in 17 locations distributed across a large portion of the species' natural distribution range. Our results revealed the existence of a hierarchical genetic structure, first separating lobsters from the northern and southern part of the range (F_{CT} = 0.0011; *P*-value = 0.0002) and then revealing a total of 11 genetically distinguishable populations (mean F_{ST} = 0.00185; CI: 0.0007–0.0021, *P*-value < 0.0002), providing strong evidence for weak, albeit fine-scale population structuring within each region. A resampling procedure showed that assignment success was highest with a subset of 3000 SNPs having the highest F_{ST}. Applying Anderson's (*Molecular Ecology Resources*, 2010, 10, 701) method to avoid 'high-grading bias', 94.2% and 80.8% of individuals were correctly assigned to their region and location of origin, respectively. Lastly, we showed that assignment success was positively associated with sample size. These results demonstrate that using a large number of SNPs improves fine-scale population structure delineation and population assignment success in a context of weak genetic structure. We discuss the implications of these findings for the conservation and management of highly connected marine species, particularly regarding the geographic scale of demographic independence.

Keywords: American lobster, assignment test, fishery management, population genomics, single-nucleotide polymorphism marker

Received 6 August 2014; revision received 9 May 2015; accepted 12 May 2015

Introduction

Determining genetically distinct populations and establishing appropriate management units are primary goals of modern conservation biology and population management (Palsboll *et al.* 2007). Towards that end, assignment tests are very useful and versatile tools (Manel *et al.* 2005; Schwartz *et al.* 2007), encompassing a wide array of applications, ranging from population structure inferences to the 'real-time' detection of

Correspondence: Laura Benestan, E-mail: laura.benestan@icloud.com

migrants (reviewed by Manel *et al.* 2005). However, highly connected and/or recently diverged populations with large effective population sizes often show very weak genetic differentiation, thus decreasing the power of genetic tools for defining management units and assigning individuals to their origin (Allendorf *et al.* 2010). The advent of next-generation sequencing (NGS) genotyping methods (Davey *et al.* 2011) promises an increase in the usefulness of genomics markers to finely define weakly structured populations (Hess *et al.* 2013; Ogden *et al.* 2013; Willette *et al.* 2014) and more accurately assign individuals (Nielsen *et al.* 2012; Larson *et al.* 2014; Candy *et al.* 2015).

Elucidating the genetic structure of populations for conservation and management purposes is particularly challenging in marine species (Allendorf *et al.* 2010). Over the last several decades, numerous studies have attempted to interpret the very weak genetic differentiation (typically $F_{ST} < 0.01$) found in most marine species and determine how to link this genetic information to management plans (Palumbi 2003; Waples & Gagiotti 2006; Waples *et al.* 2008). Here, a major issue concerns the biological meaning of such weak genetic differentiation in terms of levels of demographic independence between populations (Waples & Gagiotti 2006; Waples *et al.* 2008). In many marine species characterized by large effective population size (N_e), weak genetic structure generally translates into pronounced genetic connectivity (N_em), but it is unclear how this relates to demographic connectivity (m), which matters most for short-term population management (Cano *et al.* 2008). Indeed, the transition from demographic dependence to independence in populations with large N_e occurs within the asymptotic region of the hyperbolic relationship between F_{ST} and N_em, where genetic data have typically been insufficiently precise to discriminate between migration rates that are meaningful or not to demographic independence.

Working on a larger genomic scale by very substantially increasing the number of markers could overcome previous methodological limitations by (i) improving the accuracy of population genetic estimates, (ii) allowing the use of assignment tests for inferring 'real-time' migration and (iii) providing new insights from previously unexplored genomic regions (Kohn *et al.* 2006). Recent studies on sturgeon (Ogden *et al.* 2013) and sea anemone (Reitzel *et al.* 2013) are examples in nonmodel marine species, where NGS highlighted previously undetected demographic and evolutionary patterns. Even though the number of NGS-based genotyping studies has increased exponentially over the last few years, there has been little investigation into the possible gains that such data offer for deciphering fine-scale population structure in nonmodel marine species (but see Lamichhaney *et al.* 2012; Nielsen *et al.* 2012; Hess *et al.* 2013; Pujolar *et al.* 2013) and particularly in invertebrate species (Reitzel *et al.* 2013).

Performance of assignment methods depends mainly on the degree of population differentiation among candidate source populations, sample sizes of individuals and the number of markers used (Cornuet *et al.* 1999; Bernatchez & Duchesne 2000; Banks *et al.* 2003). In principle, genotyping thousands of SNP markers in a large number of individuals should help circumvent these constraints. However, we are not aware of any study that specifically investigates the improvement of assignment methods through the use of large sets of NGS markers in situations of weak genetic differentiation (typically $F_{ST} < 0.01$).

The main goal of this study was to assess how NGS could help delineate the genetic structure of the American lobster (*Homarus americanus*) throughout much of the species' range and increase the assignment success of individuals to their location of origin. The American lobster (henceforth lobster) supports one of the most valuable fisheries in North America. Its distribution ranges from Cape Hatteras (North Carolina, USA) in the south to the Strait of Belle Isle (Labrador, Canada) in the north. Typically inhabiting coastal waters <50 m deep, lobster can be found offshore in some localities at depths reaching 700 m (Cooper & Uzmann 1971). The carapace length at which 50% of females are sexually mature decreases with increasing temperature and varies from about 70 to 108 mm depending on locality (Watson *et al.* 2013). Mating and spawning occur during summer, usually one or more years apart, and larvae are hatched after an incubation period of 11–12 months on the abdomen of the female (Templeman 1940; Waddy *et al.* 1995). The planktonic/pelagic larval phase lasts on average 3–6 weeks, and its duration is inversely related to temperature (Ennis 1986; Quinn *et al.* 2013).

Early studies based on allozymes and random amplified polymorphic DNA (RAPD) revealed virtually no genetic differentiation in lobster from geographically separate regions (Tracey *et al.* 1975; Harding *et al.* 1997). More recently, Kenchington *et al.* (2009) conducted a detailed study of lobster along the northeast coast of North America with 13 microsatellite markers. A north–south genetic discontinuity centred on southwest Nova Scotia was detected, and a weaker, smaller-scale substructure was revealed in the southern region but not in the northern region. Weak genetic structure in American lobster might reflect potential for extensive dispersal (Incze & Naimie 2000; Xue *et al.* 2008) via ocean currents during the long pelagic larval period (Ennis 1986). Adult lobsters have also been shown to

undertake extensive seasonal migrations over distances of up to 100 km in some regions (Campbell 1986), but also exhibit homing behaviour (Pezzack & Duggan 1986). Moreover, as mating and larval release may be separated in time by about 2 years (Waddy *et al.* 1995), these events may occur in different locations, and mating rather than larval release could determine genetic patterns. Therefore, the contribution of adult movements to gene flow and population structure remains unclear.

In this study, we genotyped 586 adult American lobsters collected in 17 locations using 10 156 SNPs discovered by RAD sequencing. We first document the regional and finer-scale population genetic structure among the sampled locations and then quantify the efficiency of assignment tests as a function of number of SNPs used and sample size per location. Finally, we discuss the benefits of genotyping a large number of SNP markers for the study, conservation and management of the American lobster as well as other marine species that experience high levels of gene flow.

Methods

Sampling

We collaborated with commercial fishermen to sample lobsters from 17 locations throughout much of the species' range, 15 that were inshore and two that were offshore (Fig. 1). Sampling was done between May and August 2012. We only sampled adult females bearing late-stage eggs that would hatch in the coming weeks (n = 624 total), to standardize the sampling design and to estimate the genetic structure of individuals that had survived to reproduce. We reasoned that this sampling design would perhaps be most likely to reveal genetic

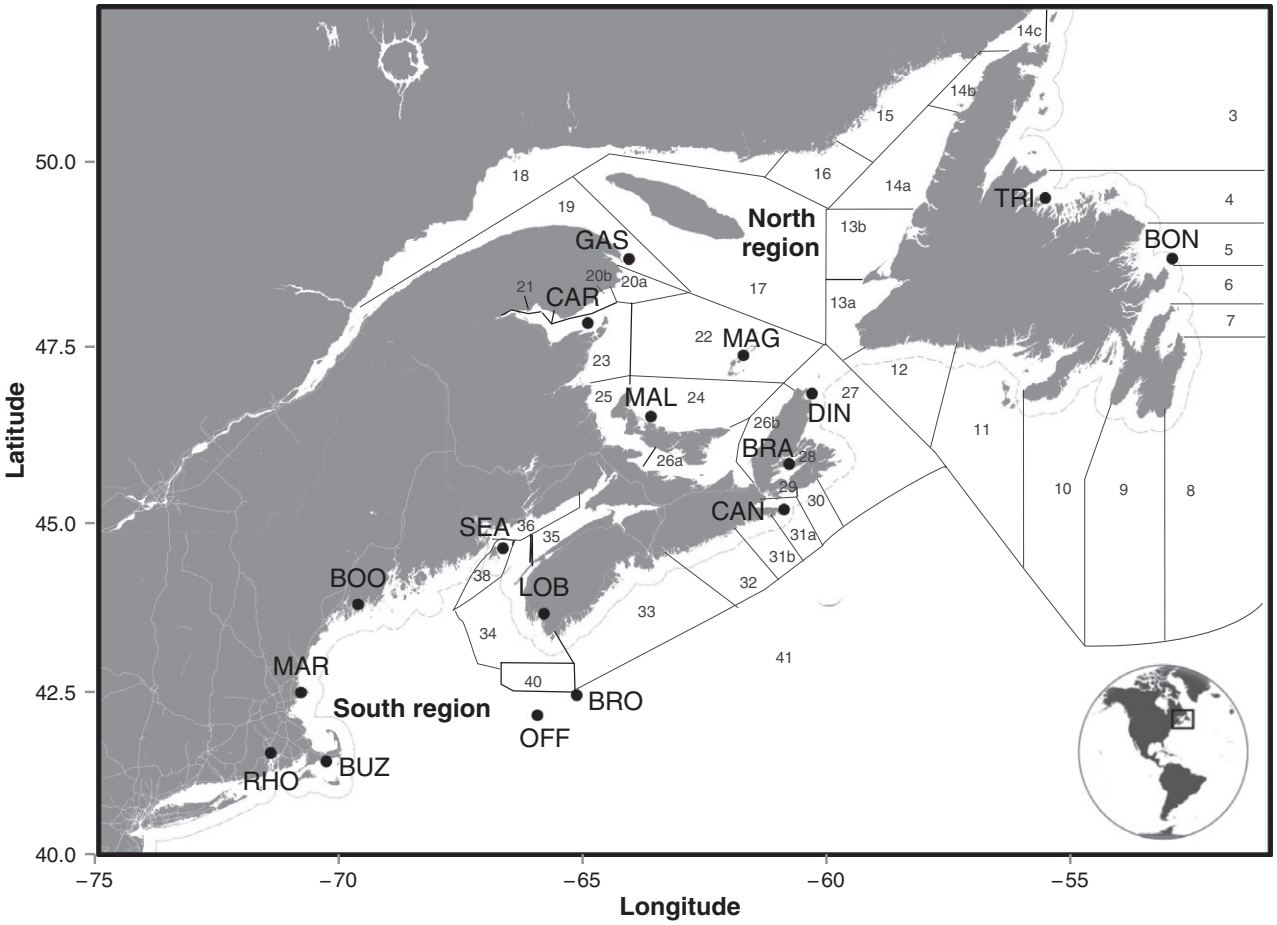

Fig. 1 Map of lobster sampling locations. South region including Gulf of Maine: BOO, Boothbay Harbour; BRO, Browns Bank; BUZ, Buzzard's Bay; LOB, Lobster Bay; MAR, Marblehead; OFF, Georges Basin; RHO, Rhode Island; SEA, Seal Cove. North region including Gulf of St. Lawrence: BON, Bonavista; BRA, Bras d'Or Lake; CAN, Canso; CAR, Caraquet; DIN, Dingwall; GAS, Gaspé; MAG, Magdalen Islands; MAL, Malpeque Bay; TRI, Triton. The figure also illustrates the limits of the 41 current management units in Canada (designated by a number), called Lobster Fisheries Areas (LFAs).

structure, particularly if females displayed homing behaviour related to spawning and hatching (Pezzack & Duggan 1986). The second walking leg of each individual was removed and preserved in 95% EtOH until DNA extraction. A total of 36 individuals were sampled in all but one study location (n = 48 for MAG) (Table 1).

Molecular techniques

Genomic DNA was extracted using a salt-extraction protocol (Aljanabi & Martinez 1997) with additional RNase A treatment (Qiagen) following the manufacturer's recommended protocols. DNA integrity (i.e. presence of degradation or smears) was inspected on a 1% agarose gel. Samples with degraded DNA were excluded. Extracted genomic DNA (gDNA) was quantified using Quantit Picogreen dsDNA assay kits (Invitrogen). RAD-sequencing libraries were prepared following a protocol modified from Miller *et al.* (2007) (see Appendix S1, Supporting information). Each library contained 48 individuals barcoded with a unique six-nucleotide sequence. Real-time PCR was used to quantify libraries. Single-read, 100-bp target length sequencing on Illumina HiSeq2000 platform was conducted at the Genome Quebec Innovation Centre (McGill University, Montreal, Canada).

Bioinformatics and genotyping

The libraries were demultiplexed using the *process_radtags* program in STACKS v.1.09 (Catchen *et al.* 2013). Polymorphic SNPs were identified on reads truncated to 90 bp and filtered for overall quality and presence of barcodes. The formation of RAD loci was allowed with a maximum of two nucleotide mismatches (M = 2)—identified as an optimum threshold according to the method developed by Ilut *et al.* (2014)—and a minimum stack depth of three (m = 3) among reads with potentially variable sequences (*ustacks* module in STACKS, with default parameters). Then, reads were aligned de novo with each other to create a catalogue of putative RAD tags (*cstacks* module in STACKS, with default parameters). In the *populations* module of STACKS and following consecutive filtering steps, we first retained RAD tags with a minimum stacks depth (m) of 10 to a maximum stacks depth of 100. This step removed SNPs genotyped with too low coverage (m < 10) to be accurately called as well as SNPs genotyped with too high coverage (m > 100), which could be located on highly overrepresented sites due to repeats in the lobster genome. Then, we retained SNPs genotyped in at least 70% of the individuals and 70% of the sampling locations. Potential homeologs were excluded by removing markers showing heterozygosity >0.50 within samples (Hohenlohe *et al.* 2011). We also removed markers out of Hardy–Weinberg equilibrium (P-value = 0.01) at more than 60% of the locations. Individuals and SNPs with more than 30% of missing data were also eliminated. To avoid bias in the estimation of the baseline differentiation and eliminate any sequencing and PCR error from the SNP data set, polymorphisms with a minor allele frequency (MAF) > 0.1 in at least one location (i.e. minor allele occurring at least 4 times in one location) and polymorphisms with MAF > 0.05 on average across sampling locations were kept. It has been shown that very low-frequency SNPs (MAF < 0.05) create biases in quantifying genetic connectivity and should therefore be removed when inferring demographic processes (Roesti *et al.* 2012). Details of the number of SNPs kept after each filtering step are provided in Table 2. The resulting filtered VCF file was converted into the file formats necessary for the following analyses using PGDSPIDER v.2.0.5.0 (Lischer & Excoffier 2012).

Detecting SNPs under selection

SNPs potentially under balancing and divergent selection should also be removed when assessing genetic connectivity between populations (Beaumont & Nichols 1996; Luikart *et al.* 2003). This was achieved using BAYESCAN v.2.1 (Foll & Gaggiotti 2008) as well as the Fdist approach (Beaumont & Nichols 1996) implemented in ARLEQUIN v.3.5 (Excoffier & Lischer 2010). BAYESCAN estimates population-specific F_{ST} coefficients by the Bayesian method described in Beaumont & Balding (2004) and uses a cut-off based on the mode of the posterior distribution to detect SNPs under selection (Foll & Gaggiotti 2008). SNPs with a posterior probability over 0.95 were considered as outliers, after running 100 000 iterations on all samples together (i.e. not pairwise, with remaining default parameters). We specified a 'prior' odd of 10 000, which set the neutral model being 10 000 times more likely than the model with selection in order to minimize false positives (Lotterhos & Whitlock 2014). ARLEQUIN was executed with 200 000 simulations and 100 demes simulated as recommended by the authors, and SNPs were considered as outliers based on their F_{ST} and P-value.

Individual and population clustering

We first inferred population structure using two Bayesian clustering methods that are implemented in the programs STRUCTURE v2.3.4 (Falush *et al.* 2003) and ADMIXTURE v1.23 (Alexander *et al.* 2009). Both programs provide a means of identifying the best value for K, the number of putative populations. With STRUCTURE, we used 10 000 burn-in iterations followed by another

10 000 Markov chain Monte Carlo (MCMC) steps assuming an admixture model based on individuals and including no prior information on sampling location. We ran ADMIXTURE using 20 000 bootstraps. For both programs, we varied the number of groups (K) from 1 to 17 with 5 iterations for each value and stabilization of parameters was checked for this length of burn-in and MCMC. We then performed a discriminant analysis of principal components (DAPC) in the R package *adegenet* (Jombart *et al.* 2010), without prior information on group individual populations, and we used the function *find.clusters* to assess the optimal number of groups with the Bayesian information criterion (BIC) method. The DAPC is a non-model-based method, which maximizes the differences between groups while minimizing variation within groups. Therefore, retaining too many discriminant functions with respect to the number of populations can lead to overfitting the discriminant functions, which results in spurious discrimination of any set of clusters. To avoid this bias, we evaluated the optimal number of discriminant functions ($n = 100$) to retain according to the optimal α-score obtained from our data (Jombart *et al.* 2010). In addition, a K-means clustering analysis was performed on sampling locations with the GENODIVE v.2.0b25 program (Meirmans & Van Tienderen 2004), using simulated annealing and testing for K clusters from 1 to 10, for 5000 permutations. This analysis provides the Calinski–Harabasz pseudo-F-statistic for determining the number of clusters (Caliński & Harabasz 1974).

Population differentiation

The extent of pairwise population differentiation was quantified using the unbiased F_{ST} estimator θ (Weir & Cockerham 1984), and 95% confidence intervals were calculated for each pairwise comparison based on 5000 permutations using GENODIVE. Significance of the observed F_{ST} values was determined by running 10 000 permutations and assessed against a FDR-adjusted P-value to account for multiple testing (Benjamini & Hochberg 1994; $\alpha = 0.05$, $n = 171$). We used the function *hclust* available in the R package *ggdendro* to create a UPGMA dendrogram based on the F_{ST} values. A heatmap was produced to illustrate the F_{ST} matrix considering four different F_{ST} groups delimited from the distribution of pairwise F_{ST} values (see Results). A hierarchical analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) based on north vs. south regional groupings (see Results) was performed. In addition, we conducted three standard Mantel tests to correlate genetic distances (F_{ST}) and natural logarithm of geographic distances. Geographic distances between each pairs of sampling locations were calculated using ARCGIS software. The first Mantel test included all pairwise comparison, whereas the two others were based only on pairwise comparison of samples belonging to the same region (either south or north) in order to take the spatial dependence in the data into account (Meirmans 2012). The Mantel tests were performed with the library *adegenet* (Jombart *et al.* 2010), and significances of the tests were assessed using 10 000 permutations.

Table 1 Regional groupings of lobster sampling locations and information on locations and samples: latitude and longitude, sampling date and number of individuals successfully genotyped (N_{GEN})

Region	Sampling location	Code	Latitude	Longitude	Sampling date	N_{GEN}
North	Malpeque Bay, PEI	MAL	46.529	−63.6874	May-12	31
	Caraquet, QC	CAR	48.7208	−64.2789	May-12	36
	Magdalene Islands, QC	MAG	47.379	−61.853	Jun-12	38
	Gaspé, QC	GAS	48.7313	−64.3065	May-12	32
	Triton, NF	TRI	49.5218	−55.6107	Jun-12	35
	Bonavista, NF	BON	47.6113	−57.5873	Jun-12	32
	Dingwall, NS	DIN	46.9139	−60.4285	Jun-12	35
	Bras d'Or Lake, NS	BRA	45.7516	−60.817	Jul-12	32
	Canso, NS	CAN	45.3362	−60.9944	Jul-12	35
South	Lobster Bay, NS	LOB	43.6792	−65.8784	Jul-12	36
	Seal Cove, NB	SEA	44.6403	−66.7199	Jul-12	33
	Boothbay Harbour, US	BOO	43.8165	−69.6897	Jul-12	35
	Marblehead, US	MAR	42.4999	−70.8578	Jul-12	34
	Buzzard's Bay, US	BUZ	41.5292	−70.8357	Jul-12	36
	Browns Bank	BRO	42.4588	−65.2083	Jul-12	35
	Georges Basin	OFF	42.1538	−66.0143	Jul-12	36
	Rhode Island, US	RHO	41.58	−71.4774	Aug-12	35

Population assignment

Pairwise genetic differentiation (F_{ST}) between the 17 sampling locations was calculated for each SNP using *hierfstat* library in R (Goudet 2005). All of the 10 156 SNPs were ranked according to their F_{ST}, from the highest to the lowest. As recommended by Anderson (2010), the calculation of F_{ST} and the ranking of the SNPs were based on a *training-set* of individuals (50% of the individuals for each sampling location), and the assignment success was assessed using the other, that is *holdout-set* of individuals. As such, pools of individuals to select markers (*training-set*) and used to assess assignment success (*holdout-set*) were totally independent, thus circumventing the problem of *high-grading bias* (Anderson 2010). To assess the impact of the number of SNPs on the assignment test results, we performed assignment tests on the subsets of SNPs (500, 1000, 2000, 3000, 4000, 5000, 6000, 7000 and 10 156 SNPs) selected according to their ranking using the *training-set* of individuals, and these subsets were tested for local assignment on the *holdout-set* of individuals. Linkage disequilibrium among markers could introduce bias when we estimated assignment success for the different subsets of markers (Manel *et al.* 2005). We therefore tested for linkage disequilibrium between each pair of loci for the 3000 most differentiated SNPs using VCFTOOLS to minimize bias of linkage disequilibrium on assignment success.

To assess the impact of the number of individuals per sampling location on assignment success, we created five random data sets of 10, 15, 20, 25 and 30 individuals, which were randomly chosen without varying the number of SNPs used (using the optimal number of 3000 SNPs, see Results), and this procedure was repeated three times. Then, we performed a standard leave-one-out assignment test on these five data sets (Paetkau *et al.* 2004). In order to further test the null hypothesis that assignment estimates obtained from our empirical data set were not due to some stochastic process, we performed assignment tests on a randomized data set with populations of identical size and randomly chosen individuals shuffled among populations. To obtain confidence intervals (CI) on estimates, we ran each assignment test on 10 generated bootstrapped data sets using repeated resampling of individuals with replacement.

Assignment tests were performed on the *holdout-set* of individuals for each population both at the regional (north/south) and at the local (i.e. putative population) scales using GENODIVE with the *frequentist* method of Paetkau *et al.* (1995). In a given genotype, when the observed frequency of any allele was zero (a missing allele), the frequency of this allele was replaced by a fixed value of 0.005 as recommended by Paetkau *et al.* (2004), to avoid the calculation of a multilocus likelihood of zero. A null distribution of likelihood values was generated using a Monte Carlo chain (Cornuet *et al.* 1999) for 5000 permutations. In an attempt to distinguish migrants from miss-assignments, we used Cornuet *et al.*'s (1999) algorithm with a statistical threshold calculated separately for every population based on an α value of 0.05 (Berry *et al.* 2004). Individuals with likelihood values of originating from their sampling location (L_H) inferior to this threshold are thus defined as putative migrants. As the GENECLASS2 program (Piry *et al.* 2004) has been more commonly used for population assignment in previous studies (e.g. Berry *et al.* 2004; Castric & Bernatchez 2004; Paetkau *et al.* 2004), we also compared the local assignment test results obtained from GENODIVE to those given by GENECLASS2, using the same parameters (0.005 for missing alleles, alpha value of 0.05 and L_H criterion).

Results

Genotyping results

The average number of sequence reads among the 16 libraries was 169 million (range: 112–189 million) and the average number of quality-filtered reads per library was 130 million (range: 87–156 million), providing an average depth of coverage per individual over all SNPs of 43× and a mean depth per nucleotide position ranging from 18× to 448×. Thirty-eight individuals (~6.0%) had an insufficient mean coverage (<10×) and were removed from subsequent analyses. After applying the

Table 2 Number of putative SNPs retained following each filtering step

From reads to SNPS	SNP count
Stacks catalogue	200 313
Population filters	
Genotyped	
>70% of the samples	74 229
>70% of the populations	
Minor allele frequency (MAF) filters	
Local MAF > 0.05	15 552
Local MAF > 0.1	
Coverage filter	
From 10 to 100×	15 505
HWE filters	
Hardy–Weinberg equilibrium (*P*-value 0.05)	10 324
$H_{OBS} < 0.5$	10 156
Genome scan filter	
Putatively neutral	8144
Putatively under divergent selection	32

different filtering steps, 10 156 SNPs were retained for subsequent analyses (Table 2).

Selecting candidate SNPs for demographic inference

From the 10 156 SNPs retained, a genome scan using ARLEQUIN detected 8645 SNPs seemingly not under selection (~85.1%), 406 SNPs (~4.0%) under divergent selection and 1105 SNPs (~10.9%) potentially under balancing selection. BAYESCAN identified 8324 SNPs (~82.0%) seemingly not affected by selection, 32 SNPs (~0.3%) potentially under divergent selection and 1800 SNPs (~17.7%) potentially under balancing selection (Fig. S1, Supporting information). Here, we used the most conservative neutral model available in BAYESCAN (pr_odds = 10 000) to minimize false positives detected as being under positive or balancing selection (Lotterhos & Whitlock 2014). The finding of a high number of SNPs potentially under balancing selection may also support several studies suggesting or showing that balancing selection is more prevalent in the genome than previously expected (Nielsen 2005; Charlesworth 2006; Shimada *et al.* 2011). Subsequent inferences of genetic structure were carried out using the 8144 SNPs' (~80.1%) candidate markers that were concluded not to be under selection by both BAYESCAN and ARLEQUIN.

F-statistics

Our results showed that the majority of sampling locations were genetically differentiated. Average F_{ST} was 0.00185 across all 8144 SNPs, and all pairwise comparisons of the 17 sampling sites ranged from 0.00002 (BRO vs. OFF) to 0.00374 (BON vs. BRO) (Table S1, Supporting information). Overall, 129 of the 136 pairwise comparisons of genetic differentiation between sampling locations were significant (P-value < 0.05), which resolved 11 genetically distinguishable populations among the 17 sampling sites. Eight of these 11 putative populations corresponded to unique sampling locations (BON, BOO, BRA, CAR, CAN, SEA, RHO and TRI), and three (hereafter South Gulf of Saint Lawrence = SGL, southwest Nova Scotia = SNS and Cape Cod = CCO) clustered together with neighbouring sampling locations (SGL: GAS, DIN, MAG and MAL; SNS: BRO, LOB and OFF; CCO: MAR and BUZ). Average F_{ST} was 0.00199 across all SNPs and the 11 putative populations and ranged from 0.00101 (SNS vs. SEA) to 0.00374 (BOO vs. SNS). Significant P-values for most of the pairwise comparisons of genetic differentiation were consistent with the very narrow 95% confidence intervals around F_{ST} estimates, which averaged ±0.0006, and never encompassed zero for all the significant comparisons (Table S1, Supporting information).

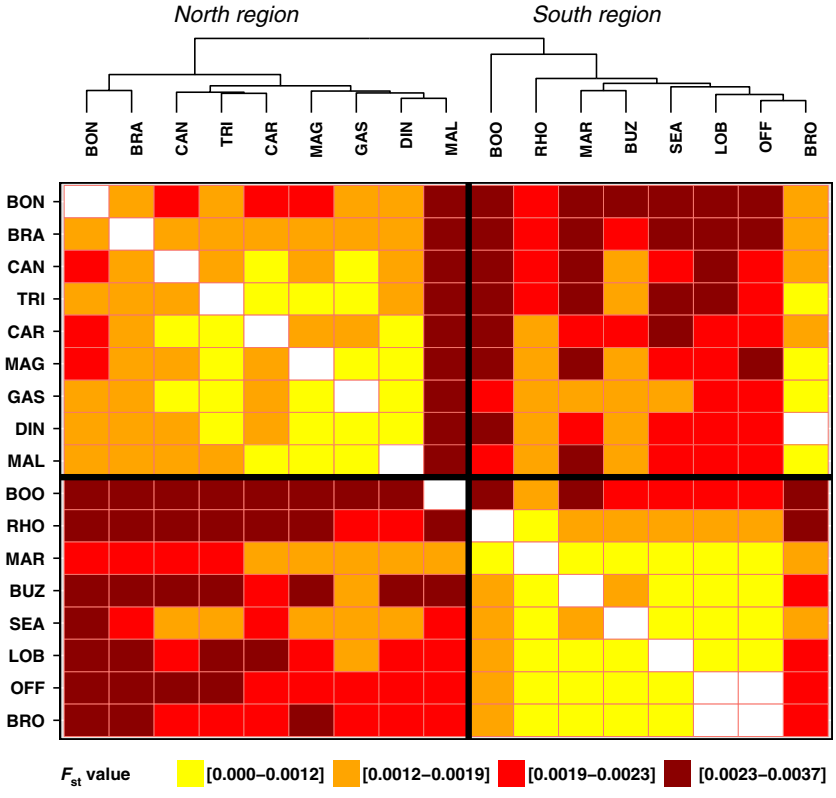

Fig. 2 F_{ST} population dendrogram and heatmap based on F_{ST} values among the 17 lobster sampling locations. The heatmap colour code illustrates the F_{ST} matrix considering four different F_{ST} groups delimited from the pairwise F_{ST} distribution: low F_{ST} (below 5th percentile, $F_{ST} < 0.0012$), low–intermediate F_{ST} (5th to 25th percentile, $0.0012 \leq F_{ST} < 0.0019$), intermediate (25th to 75th percentile, $0.0019 \leq F_{ST} < 0.0023$), high F_{ST} (above 75th percentile, $F_{ST} \geq 0.0023$).

Both the heatmap and the dendrogram based on F_{ST} values separated samples belonging to the north region from those belonging to south region of the sampled lobster distribution range (Fig. 2). The heatmap illustrated the dichotomic nature of the F_{ST} values, with lower F_{ST} values generally observed between sampling locations within each of the two large geographic regions (north or south) and higher F_{ST} values between locations belonging to the different geographic regions (Fig. 2). The AMOVA showed a modest yet highly significant net genetic differentiation between samples from the north and the south regions (F_{CT} = 0.0011, *P*-value = 0.0002; Table 3). The variation between sampling locations within each region was also significant (F_{ST} = 0.0010, *P*-value < 0.0002) and equal to than that found between regions (Table 3). We detected a strong and highly significant positive association between genetic and geographic distances (r^2 = 0.56, *P*-value < 9.999e-05) when considering all pairwise comparisons (Fig. 3). This association was still significant, albeit weaker, when considering samples only within the north region (r^2 = 0.41 and *P*-value = 0.046) or the south region (r^2 = 0.20 and *P*-value = 0.049).

Table 3 Analysis of molecular variance (AMOVA) among 17 sampling locations distributed in the north and south regions of the sampled distribution range of lobster

Source of variation	Percentage of variation	Variance	*P*-value
Between regions	0.11	0.001	0.0002
Among locations within regions	0.10	0.001	0.0002
Among individuals within locations	99.79	0.363	–

Clustering of individuals and populations

The genetic split between north and south regions was also discerned by both the DAPC and *K*-means analyses but not by STRUCTURE and ADMIXTURE. Thus, all lobsters analysed were grouped into a single cluster according to STRUCTURE and ADMIXTURE when using 8144 potentially neutral SNPs. The same result was obtained when we included all 10 156 SNPs (results not shown). In contrast, the DAPC revealed two clusters, according to the lowest BIC, separated along the first discriminant function (PC1), which explained 33.62% of the total genetic variation among individuals (Fig. 4). Discriminant functions 2 (PC2), 3 and 4 accounted for 6.27%, 3.84% and 2.28% of the variance, respectively, and did not reveal any particular clustering (results not shown). Although there was some overlap between the two groups, the first cluster resolved by discriminant function 1 corresponded mainly to individuals from the north region, whereas the second cluster contained mainly individuals from the south region (Fig. 4). Moreover, an optimal *K* of 2 clusters, corresponding to the north–south separation, was found when performing the analysis at the population level using the pseudo-*F*-statistics (Fig. 4).

Individual assignment analysis

The assignment success of individuals to their respective sampling locations was strongly affected by the number of SNPs used that were ranked based on their average F_{ST} value across all sampling locations (Fig. 5). Thus, the average assignment success to sampling location increased with the number of SNPs from 60.2% when using the top 500 most differentiated SNPs to a maximum of 80.8% using the top 3000 most differentiated SNPs and then decreased to only 8.9% using all 10 156 SNPs. Regarding the effect of individuals sampled per location, increasing the number of individuals from 10 to the maximum average of 34 increased the

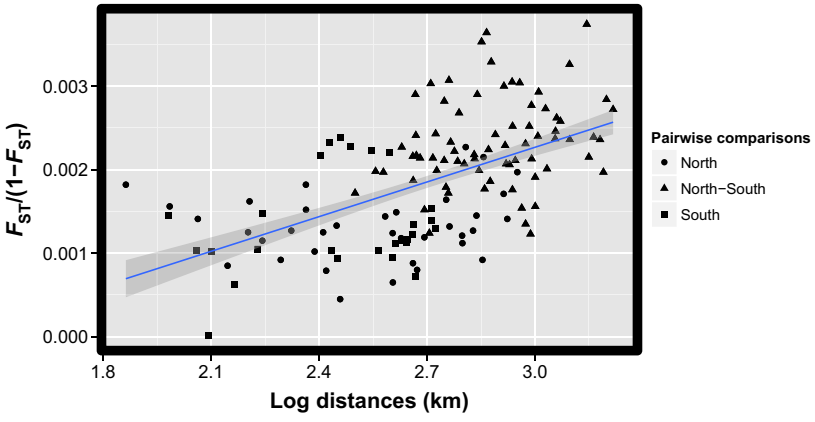

Fig. 3 Pairwise genetic distances (F_{ST}) in relation to geographic distances [Log (km)] between lobster sampling locations, with a linear regression line (in black) fitted with 95% confidence limits (in grey). Pairwise comparisons within and between the south and north regions are represented by circles (within north), squares (within south) or squares (north vs. south).

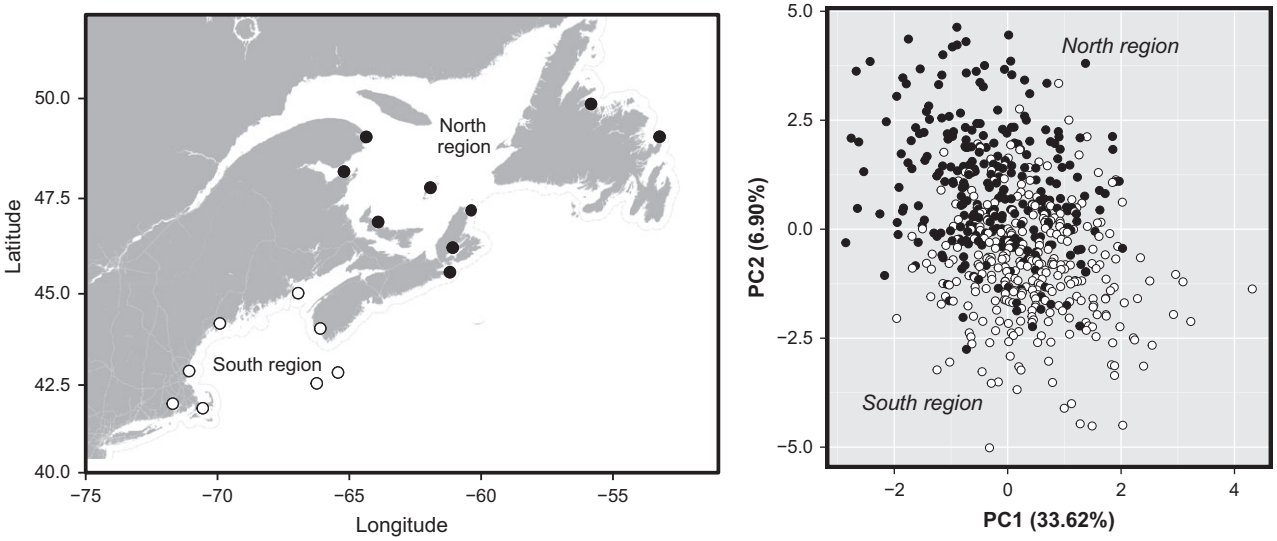

Fig. 4 Left panel: Discriminant analysis of principal components (DAPC) of genetic differentiation among the 586 genotyped lobsters based on 8144 single nucleotide polymorphism markers (each point represents one individual) with principal component 1 (PC1: 33.62% of variance) against principal component 2 (PC2: 6.90% of variance). Right panel: Pseudo-*F*-statistics analysis assigning each sampling location to either the south or the north region. The individuals (left panel) and sampling locations (right panel) from the south and north regions are represented by white and black symbols, respectively.

proportion of individuals (using the top 3000 SNPs) correctly assigned to their location of origin from 13.7% (range: 0–50.0%) to 80.8% (range: 56.6–95.6%) on average (Fig. 5). Visual inspection of this relationship indicates that sampling a greater number of individuals than were sampled in this study would have generated additional gains in assignment success.

At the regional scale, GENODIVE assigned lobsters to their region of origin with very high success. Lobsters sampled in the north and south regions were re-assigned correctly at 93.6% and 94.8%, respectively, using the top 3000 most differentiated SNPs. At the population level, that is considering the 11 putative genetically distinct populations as defined above, assignment success was lower than between the north and south regions but still high with an average of 80.8%. However, assignment success was highly variable depending on population, ranging between 55.5% (CAN) and 95.6% (SGL) (Fig. 6). Interestingly, the lowest assignment success is for a site (CAN) along the Scotia shelf where there might be a discontinuity in structure between the north and south regions. We also estimated assignment success for sampling sites that were pooled together as representing a same putative population based on F_{ST} values and alpha = 0.05. Assignment success was still high for these sites, averaging 77% for sites within SGL (GAS, MAL, MAG and DIN), 78% within SNS (LOB, OFF and BRO) and 83% within CCO (MAR and BUZ) (Fig. S2, Supporting information). As expected, miss-assigned individuals were generally assigned to other sites within each of these three putative populations. This indicates that despite the lack of statistically significant genetic differences between sites that were pooled as representing a same putative population, individuals from a given site were genetically more similar among themselves than they were to lobsters from other sites.

We found only 140 pairs of loci with an r^2 value > 0.5 in all sampling locations. Indeed, nonindependence of markers was expected to be low as the lobster genome is several times larger than that of many marine fish (~ n = 69 chromosomes: Coluccia *et al.* 2001; C = 4.75: Animal Genome Size Database). We randomly removed one of linked SNPs, and we assessed assignment success again using the remaining 2860 SNPs. Assignment success obtained in this case was very similar to assignment success using all 3000 SNPs, with on average 93.7% (instead of 94.2%) individuals correctly assigned to their region of origin and 79.6% (instead of 80.8%) individuals correctly assigned to their population of origin.

When using the randomized data set, <5% of individuals were correctly assigned to their location of origin, clearly indicating the rejection of the null hypothesis of random assignment based on the empirical data set. In contrast, the bootstrapped data set (using the top 3000 SNPs) gave a high assignment success of 80.3% on average, which is similar to the primary data set, further validating results of the assignment tests. However, assignment tests could not confidently tell apart

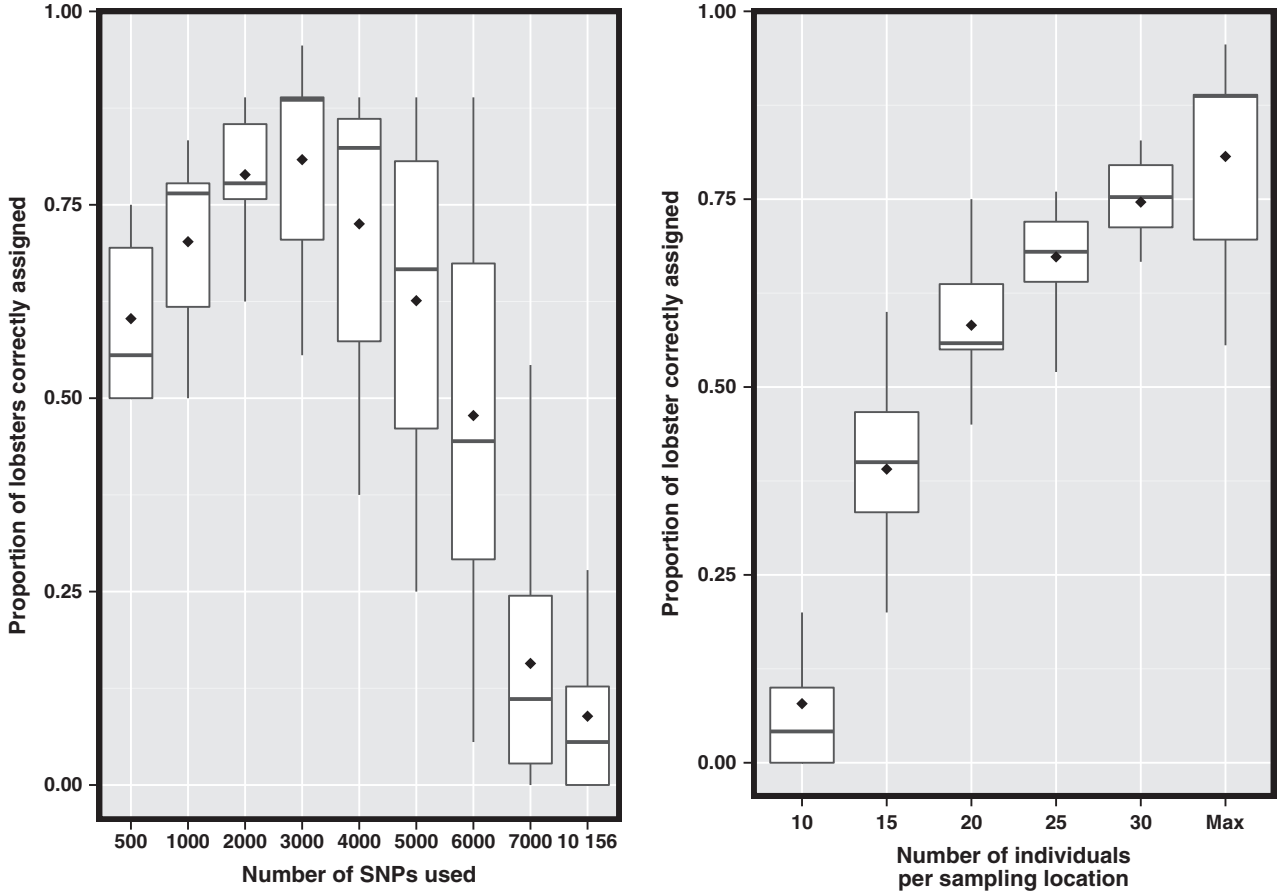

Fig. 5 Left panel: Boxplot of the proportion of lobsters correctly assigned to their sampling location (*y*-axis) as a function of number of SNPs ranked by decreasing order of average F_{ST} values (*x*-axis) following the Anderson (2010) method. Right panel: Boxplot of the proportion of lobsters correctly assigned to their sampling location (*y*-axis) as a function of number of individuals per sampling location and according to a standard leave-one-out procedure. The 'max' label refers to the maximum number of individuals per sampling location, which varies from 31 to 38 (average 34) individuals (see Table 1). In both panels, the vertical limits of the box represent one standard deviation around the mean (black diamond), the horizontal line within the box is the median, and the whiskers extend from the box to the 25th and 75th percentiles.

migrant individuals from incorrect assignments, as no individuals were outside the 95% likelihood limits of their respective population. When sampling locations were considered separately, GENECLASS2 and GENODIVE gave a similar assignment success (average 81.5% and 80.8%, respectively, Student's *t*-test, *P*-value = 0.82). Correlation between assignment successes obtained by both methods for a given population was also high (Rho = 0.84), indicating largely consistent conclusions between the two programs.

Discussion

The main goal of this study was to assess how using thousands of SNPs could help to better delineate fine-scale genetic structure and increase the assignment success of individuals to their site, putative population and region of origin in weakly genetically structured marine species using the American lobster as a case study. Results revealed the existence of a hierarchical genetic structure, first separating populations from the north and the south regions of the sampled range and then separating populations within each of these regions. Thus, 11 putative populations were resolved of the 17 sampling locations, revealing population genetic structuring at finer-spatial scale than previously revealed for this species. On the other hand, whereas F_{ST} values were often highly statistically significant, they were always small and comparable to values frequently reported for other species of marine vertebrates and invertebrates. These small F_{ST} values suggest pronounced genetic connectivity among sites and putative populations or recent separation and slow approach to equilibrium in very large popula-

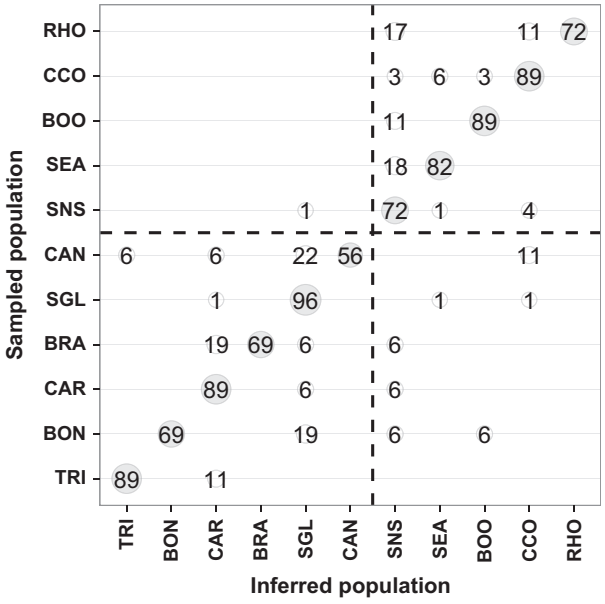

Fig. 6 Blind assignment success expressed as the percentage of lobsters sampled from one putative genetic population that are classified into their population of origin (grey-shaded numbers on diagonal) or inferred to belong to another putative population (nonshaded numbers). Eleven putative populations were identified (see text), of which 8 were single sampling locations (BON, BOO, BRA, CAR, CAN, SEA, RHO, TRI) and 3 were clusters of neighbouring sampling locations (South Gulf of St. Lawrence, SGL, grouping GAS, DIN, MAG and MAL; Southwest Nova Scotia, SNS, grouping BRO, LOB and OFF; Cape Cod, CCO, grouping MAR and BUZ). Dashed lines represent a higher-level genetic discontinuity separating putative populations in the south (above horizontal line on *y*-axis) from those in the north (below horizontal line on *y*-axis) of the sampled distribution range.

tions (Marko & Hart 2011). However, contrary to earlier studies on this species, confidence intervals around F_{ST} estimates were very narrow and excluded zero, as a consequence of the very large number of markers used. Results from the assignment tests provided further evidence for this general pattern of population structuring as 94.2% of individuals were correctly assigned to their region of origin and 80.8% were correctly assigned to their putative population of origin within each region. In addition, assignment success remained high when assigning individuals to sampling locations that were not significantly differentiated based on F_{ST}, indicating that lobsters from the same location are genetically more similar among themselves than they are with individuals from other locations. Overall, these results confirm the resolution gained by using a large number of SNP markers to delineate fine-scale population structuring and to perform assignment tests in highly genetically connected marine species (Waples & Gagiotti 2006). Below, we discuss the implications of these findings for the study, conservation and management of American lobster and other highly connected marine species.

Fine-scale population structuring

The small yet significant genetic differentiation found among 94.8% of the pairwise site comparisons, along with generally high site, population or regional assignment success, contributes to a growing literature finding that many marine organisms are subdivided into genetically separated units, sometimes at small spatial scales (e.g. Atlantic cod, *Gadus morhua*: Ruzzante *et al.* 1999; Knutsen *et al.* 2003; flathead mullet, *Mugil cephalus*: Krück *et al.* 2013; Atlantic herring, *Clupea harengus*; Pacific lamprey, *Entosphenus tridentatus*: Hess *et al.* 2013), which has changed the general perception that most marine species are panmictic across broad geographic scales (Swearer *et al.* 1999; Mora & Sale 2002; Banks *et al.* 2007; Iacchei *et al.* 2013). In the particular case of American lobster, earlier studies did indeed suggest that the species was panmictic over large geographic areas (Tracey *et al.* 1975; Harding *et al.* 1997). However, Kenchington *et al.* (2009) provided evidence of a north–south discontinuity in genetic structure that is corroborated by the genetic structure observed with SNPs reported here. Kenchington's study also showed a fine-scale genetic structure in the southern region, but not in the northern region where panmixia was proposed. In contrast, our results suggested the existence of six populations among the nine sampling sites from the northern region. Although genetic differences were small and variable depending on sampling sites comparison, they were accompanied by a relatively high assignment success. This outcome is most likely due to the increased accuracy and statistical power provided by screening thousands of SNPs across the lobster genome, as anticipated by Allendorf *et al.* (2010). Our results show that the use of thousands of SNPs returned very narrow (±0.0006) confidence intervals even around weak estimates of differentiation, therefore substantially increasing the accuracy of F_{ST} estimates. Willing *et al.* (2012) recently demonstrated via computer simulations that a large number of screened markers could be used to detect genetic differentiation as small as $F_{ST} = 0.001$, assuming there is a real genetic structure. This increased accuracy of genetic estimates may enhance our ability to relate indirect measures of gene flow and migration to demographic connectivity (that is *m*, the proportion of migrants among populations per generation), which matters more than genetic connectivity for short-term population management (Waples & Gagiotti 2006; Cano *et al.* 2008). Here, our results of population assignment suggest that at least some of the lob-

ster putative populations might be 'demographically independent', meaning that their dynamics is driven more by local birth and death than immigration and emigration (Hanski 1998). For instance, more than 89% of individual lobsters were correctly assigned for 6 of the 11 proposed populations, suggesting on average for these a maximum proportion of migrants (m) of about 0.11, that is considering that a proportion of that 0.11 most likely corresponds to spurious miss-assignment errors. Interestingly, and although this must be interpreted cautiously (Lowe & Allendorf 2010), Hastings (1993) proposed a value of $m = 0.1$ as the threshold below which populations may be considered demographically independent. Admittedly, however, interpretations regarding demographic independence must be done cautiously because our study was based on egg-carrying females, which is likely to have increased detectable genetic population differentiation. Whereas this strategy was used to standardize our sampling design, it may have biased the estimates of demographic independence, to which males and juveniles may also contribute. Therefore, future studies on this species should also compare patterns of connectivity in males and juveniles.

Our findings set the stage for future research into the demographic processes that are relevant to fine-scale genetic structuring in American lobster and other weakly differentiated marine species. For American lobster, bio-physical larval dispersal models have shown that lobster postlarvae may disperse up to 300–400 km from where they hatch (Incze & Naimie 2000; Xue *et al.* 2008; Chassé & Miller 2010), but it is not known what proportion of these individuals will successfully settle and survive to recruit into the 'local' reproductive adult population. Similarly, although some adults have a resident behaviour year-round, most undergo seasonal movements or long-range migrations to search for overwintering habitat that protects against harsh coastal winter conditions (e.g. ice scour or storms) and/or dampens seasonal thermal variability (Campbell 1986; Bowlby *et al.* 2007; Cowan *et al.* 2007). Despite the observation of long distance movements by some individuals, migrating adult lobsters tagged within the northern and southern regions defined here, including egg-bearing females, are generally recaptured within 5–10 km of their original tagging location, even after a number of years at liberty (Campbell 1986; Pezzack & Duggan 1986; Comeau & Savoie 2002). This would be congruent with the low migration rate suggested by our assignment tests. There is also evidence that adult American lobsters display homing behaviour (Comeau & Savoie 2002), as reported in palinurid lobsters (*Panulirus cygnus*: Chittleborough 1974; *Panulirus argus*: Herrnkind *et al.* 1975; *Jasus edwardsii*: Kelly & MacDiarmid 2003; *Panulirus versicolor*: Frisch 2007; *Palinurus elephas*: Follesa *et al.* 2009). Homing behaviour could result in large groups of adults belonging to a same population segregating to their coastal areas for reproduction, independent of other such groups, thereby potentially reducing genetic connectivity even if the adults undergo long-range migrations at certain times of the year (Lawton & Lavalli 1995).

Our results indicated that isolation by distance does play a role in the observed pattern of genetic structure and that this was not only driven by the hierarchical separation between the south and north regions, as significant isolation by distance existed within and between regions. Clearly, this underlines the need for a more comprehensive study investigating the impact of factors other than geography in determining the genetic structure of American lobster. This, however, was beyond the scope of this study and will be treated elsewhere (L. Benestan, B. Quinn, R. Rochette, L. Bernatchez, in preparation). Namely integrating larval dispersal and consideration of additional environmental factors (e.g. ocean temperature, salinity, bottom topography, coastline) into a seascape genetics framework could help better understand the ecological determinants underlying the observed pattern of genetic structure in lobster, similar to previous works on highly connected marine species (Banks *et al.* 2007; Selkoe *et al.* 2010; White *et al.* 2010).

Hierarchical structure between south and north regions

The genetic distinctiveness of the north and south regional groups of populations was previously interpreted as the result of a range expansion from south to north following the end of the last glacial period, approximately 10 000 years BP (Kenchington *et al.* 2009). An additional explanation could lie in oceanographic features that promote larval exchange and retention within each of these two regions (Urrego-Blanco & Sheng 2014). Moreover, the direction of larval dispersal between the two regions is probably constrained by the dominant southwesterly current outflow from the Gulf of St. Lawrence to the Gulf of Maine via the Atlantic coast of Nova Scotia, and not the other way around. At the mid-Scotian Shelf, off Mahone Bay, the surface currents disperse larvae away from the coast (Hannah *et al.* 2001), and this could act as a barrier to gene flow, assuming the larvae do not survive. This hypothesis is in agreement with previous studies showing difference in productivity between southern and northern populations along the Nova Scotia (reviewed by Miller 1997). At the same geographic area than our study, a recent genetic study also revealed the existence of a north/south dichotomy in northern shrimp (*Pandalus borealis*) that could be explained by oceanic circulation and temperature varia-

tion (Jorde *et al.* 2015). That being said, the net genetic differentiation between the north and south regions identified here was weak, which is also consistent with physical oceanographic studies suggesting that a proportion of larvae may drift through the strong Scotian Shelf current every generation and translate into long-term and pronounced genetic connectivity (Hannah *et al.* 2001). As discussed above, however, we cannot exclude the possibility that the weak differentiation between lobsters from the two regions may also reflect their very recent divergence along with presumably large effective population sizes. On the other hand, the assignment tests indicated again that the proportion of migrants between the two regions is very low. Thus, the 94.2% assignment success within each region suggests a short-term demographic independence between the two regional groups. This is also consistent with results of all the tagging studies involving adult lobsters, which report no long distance movements between the Gulf of Maine and Gulf of St. Lawrence lobsters (Lawton & Lavalli 1995).

The use of clustering software (DAPC, pseudo-*F*-statistics, AMOVA, STRUCTURE and ADMIXTURE) with different sensitivities to uncover subtle population structure resulted in contrasting findings. STRUCTURE and ADMIXTURE did not reveal any genetic structure (either regional or local), whereas DAPC, pseudo-*F*-statistics and AMOVA showed a significant division between the south and north regions. This is congruent with simulation studies (Waples & Gagiotti 2006; Kalinowski 2010) showing that Bayesian clustering methods fail to detect any genetic structure when genetic divergence is very low ($F_{ST} < 0.01$). Apparently, this still holds true even when using thousands of markers, as suggested in this study. Also, Kanno *et al.* (2011) and Jombart *et al.* (2010) showed the efficiency of DAPC to discern significant genetic clusters where STRUCTURE failed to detect any signs of clustering in the system. Thus, DAPC appears more efficient than STRUCTURE at detecting population clustering in systems of weakly ($F_{ST} < 0.01$) differentiated populations.

Assignment success as a function of number of markers and sample size

Several simulation-based studies and analytical models previously demonstrated that correct assignment varies as a function of the number of markers and individuals used (e.g. Cornuet *et al.* 1999; Bernatchez & Duchesne 2000; Paetkau *et al.* 2004). Here, our results empirically illustrate how the potential of using a large number of SNP markers may enhance the resolution of assignment methods for weakly differentiated populations. However, while we showed how increasing the number of markers genotyped up to a maximum of 3000 top-ranked markers improved assignment success, beyond that number, the assignment success decreased gradually, indicating that more weakly differentiated markers added noise and contributed to blurring rather than improving assignment. We believe that this is most likely due to a sampling error (arising from too few individuals being analysed), which is stronger on weakly differentiated markers with only modest allele frequency differences between populations relative to more differentiated markers (Roques *et al.* 1999). It would be important in future studies to assess whether this pattern of decreasing assignment success beyond a given number of top rank markers will be generalized in other marine species with similarly weak population structure. As for the effect of sample size, our results showed that our maximum number of individuals per sampling location in total ($n = 34$ on average) was not sufficient to reach the highest assignment success possibly attainable in this system with the top 3000 markers. Clearly, further improvement in assignment success could have potentially been gained by substantially increasing the number of individuals genotyped per sampling location.

Management implications

Our sampling design was largely based on obtaining samples belonging to different spatial units currently used for lobster management in the Northwest Atlantic [e.g. lobster fishing areas [LFAs] in Canada; Fig. 1]. Interestingly, the pattern of structuring we observed generally fitted these LFAs in the sense that most sampling sites representing different LFAs were genetically differentiated and lobsters belonging to different LFAs were often reassigned with high success. In some cases, however, such as the South Gulf of St. Lawrence, samples from different LFAs were not different based on F_{ST} values and assignment success was reduced, albeit remaining markedly more important than expected by chance alone. This suggests that there is a geographic distance below which demographic dependence may occur. Therefore, future studies should aim to refine the geographic scale of structuring by applying a sampling design including different geographic scales, many samples from the same LFA, different lobster life stages from larvae to adults, and both genders. Moreover, the temporal and seasonal stability of population structure should be addressed to properly document the match between population structure and management units. Finally, and although sample sizes should be increased, the promising results of individual assignment to their population of origin indicate that a lobster SNP database covering most, if not all populations, could also provide new informative tools in the context of com-

mercialization and marketing of American lobster. For instance, in the context of eco-certification and increased consumer awareness, such a database could provide a means for local managers and fishermen to define territorial branding. Moreover, the application of population assignment based on such a database could improve the traceability from fishers to consumers (e.g. FishPopTrace Consortium, Nielsen *et al.* 2012). We envision a bright future for the use of high-density genotyping facilitated by NGS-based genotyping protocols, both for improving our basic knowledge of population genetic structure of highly connected marine species and for using that knowledge to improve management and conservation practices of exploited species.

Acknowledgements

This research is part of the 'Lobster Node' of the NSERC Canadian Fisheries Research Network (CFRN). We first and foremost wish to thank the fishermen of the Lobster Node without whom this project would have been impossible. Project design and work plan were done in collaboration with scientists from the Department of Fisheries and Oceans (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 J. Gaudin and E. Normandeau for their help in bioinformatic analyses. This manuscript was improved by comments from B. Sutherland, J. S. Moore, A. Dalziel, G. Parent, A. M. Dion-Côté and Q. Rougemont. The NSERC CFRN funded this research. We also acknowledge Paul Hohenlohe, Lorenz Hauser and the two anonymous reviewers for their comments and advices, which greatly improved the quality of the manuscript. L. Benestan was supported by a doctoral fellowship from NSERC CFRN and Réseau Aquaculture Québec (RAQ).

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692–4693.

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

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, **10**, 701–710.

Banks MA, Eichert W, Olsen JB (2003) Which genetic loci have greater population assignment power? *Bioinformatics*, **19**, 1436–1438.

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.

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

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **263**, 1619–1626.

Benjamini Y, Hochberg Y (1994) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **1**, 289–300.

Bernatchez L, Duchesne P (2000) Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 1–12.

Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? *Molecular Ecology*, **13**, 551–561.

Bowlby HD, Hanson JM, Hutchings JA (2007) Resident and dispersal behavior among individuals within a population of American lobster *Homarus americanus*. *Marine Ecology Progress Series*, **331**, 207–218.

Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Communications in Statistics—Theory and Methods*, **3**, 1–27.

Campbell A (1986) Migratory movements of ovigerous lobsters, *Homarus americanus*, tagged off Grand Manan, Eastern Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **43**, 2197–2205.

Candy JR, Campbell NR, Grinnell MH, Beacham TD, Larson WA, Narum SR (2015) Population differentiation determined from putatively neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.12400. [Epub ahead of print].

Cano JM, Shikano T, Kuparinen A, Merilä J (2008) Genetic differentiation, effective population size and gene flow in marine fishes: implications for stock management. *Journal of Integrative Field Biology*, **5**, 1–10.

Castric V, Bernatchez L (2004) Individual assignment test reveals differential restriction to dispersal between two salmonids despite no increase of genetic differences with distance. *Molecular Ecology*, **13**, 1299–1312.

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

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PloS Genetics*, **2**, e64.

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

Chittleborough RG (1974) Home range, homing and dominance in juvenile western rock lobsters. *Marine and Freshwater Research*, **25**, 227.

Coluccia E, Cau A, Cannas R, Milia A, Salvadori S, Deiana AM (2001) Mitotic and meiotic chromosomes of the American lobster *Homarus americanus* (Nephropidae, Decapoda). *Hydrobiologia*, **449**, 149–152.

Comeau M, Savoie F (2002) Movement of American lobster (*Homarus americanus*) in the southwestern Gulf of St. Lawrence. *Fishery Bulletin*, **100**, 181–192.

Cooper RA, Uzmann JR (1971) Migrations and growth of deep-sea lobsters, *Homarus americanus*. *Science*, **171**, 288–290.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**, 1989–2000.

Cowan DF, Watson WH, Solow AR, Mountcastle AM (2007) Thermal histories of brooding lobsters, *Homarus americanus*, in the Gulf of Maine. *Marine Biology*, **150**, 463–470.

Davey JW, Hohenlohe PA, Etter PD, *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Ennis GP (1986) Stock definition, recruitment variability, and larval recruitment processes in the American lobster, *Homarus americanus*: a review. *Canadian Journal of Fisheries and Aquatic Sciences*, **43**, 2072–2084.

Excoffier L, Lischer HE (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, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

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.

Follesa MC, Cuccu D, Cannas R, Sabatini A, Deiana AM, Cau A (2009) Movement patterns of the spiny lobster *Palinurus elephas* (Fabricius, 1787) from a central western Mediterranean protected area. *Scientia Marina*, **73**, 499–506.

Frisch AJ (2007) Short-and long-term movements of painted lobster (*Panulirus versicolor*) on a coral reef at Northwest Island, Australia. *Coral Reefs*, **26**, 311–317.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Hannah CG, Shore JA, Loder JW (2001) Seasonal circulation on the western and central Scotian Shelf. *Journal of Physical Oceanography*, **31**, 591–615.

Hanski I (1998) Metapopulation dynamics. *Nature*, **396**, 41–49.

Harding GC, Kenchington EL, Bird CJ, Pezzack DS, Landry DC (1997) Genetic relationships among subpopulations of the American lobster (*Homarus americanus*) as revealed by random amplified polymorphic DNA. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 1762–1771.

Hastings A (1993) Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. *Ecology*, **74**, 1362.

Herrnkind WF, Vanderwalker JA, Barr L (1975) Population dynamics, ecology and behavior of spiny lobsters (*Panulirus argus*) of St. John, US Virgin Islands: habitation, patterns of movements. Results of the Tektite program, vol. 2. *Science Bulletin of Natural History Museum Los Angeles County*, **20**, 31–45.

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.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

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

Ilut DC, Nydam ML, Hare MP (2014) Defining loci in restriction-based reduced representation genomic data from nonmodel species: sources of bias and diagnostics for optimal clustering. *BioMed Research International*, **2014**, 675158.

Incze LS, Naimie CE (2000) Modelling the transport of lobster (*Homarus americanus*) larvae and postlarvae in the Gulf of Maine. *Fisheries Oceanography*, **9**, 99–113.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

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.

Kalinowski ST (2010) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, **106**, 625–632.

Kanno Y, Vokoun JC, Letcher BH (2011) Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. *Molecular Ecology*, **20**, 3711–3729.

Kelly S, MacDiarmid AB (2003) Movement patterns of mature spiny lobsters, *Jasus edwardsii*, from a marine reserve. *New Zealand Journal of Marine Freshwater Research*, **37**, 149–158.

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.

Knutsen H, Jorde PE, Anré C, Stenseth NC (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology*, **12**, 385–394.

Kohn MH, Murphy WJ, Ostrander EA (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution*, **21**, 629–637.

Krück NC, Innes DI, Ovenden JR (2013) New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). *Molecular Ecology Resources*, **13**, 715–725.

Lamichhaney S, Barrio AM, Rafati N *et al.* (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences, USA*, **109**, 19345–19350.

Larson WA, Seeb LW, Everett MV *et al.* (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, **7**, 355–369.

Lawton P, Lavalli KL (1995) Postlarval, juvenile, adolescent, and adult ecology. In: *Biology of the Lobster Homarus americanus* (ed. Factor JR), pp. 47–88. Academic Press, San Diego, California.

Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038–3051.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, **20**, 136–142.

Marko PB, Hart HW (2011) The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, **26**, 448–456.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839–2846.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Miller RJ (1997) Spatial differences in the productivity of American lobster in Nova Scotia. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 1613–1618.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**, 240–248.

Mora C, Sale PF (2002) Are populations of coral reef fish open or closed? *Trends in Ecology & Evolution*, **17**, 422–428.

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

Nielsen EE, Cariani A, MacAoidh E *et al.* (2012) Gene-associated markers provide tools for tackling IUU fishing and false eco-certification. *Nature Communications*, **3**, 851.

Ogden R, Gharbi K, Mugue N *et al.* (2013) Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology*, **22**, 3112–3123.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, **4**, 347–354.

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Palsboll P, Berube M, Allendorf FW (2007) Identification of management units using population genetic data. *Trends in Ecology & Evolution*, **22**, 11–16.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, **13**, 146–158.

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.

Piry S, Alapetite A, Cornuet JM *et al.* (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *The Journal of Heredity*, **95**, 536–539.

Pujolar JM, Jacobsen MW, Frydenberg J *et al.* (2013) A resource of genome-wide single-nucleotide polymorphisms generated by RAD tag sequencing in the critically endangered European eel. *Molecular Ecology Resources*, **13**, 706–714.

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

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, **22**, 2953–2970.

Roesti M, Salzburger W, Berner D (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, **12**, 94.

Roques S, Duchesne P, Bernatchez L (1999) Potential of microsatellites for individual assignment: the North Atlantic redfish (genus *Sebastes*) species complex as a case study. *Molecular Ecology*, **8**, 1703–1717.

Ruzzante DE, Taggart CT, Cook D (1999) A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St. Lawrence. *Fisheries Research*, **43**, 79–97.

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, **22**, 25–33.

Selkoe KA, Watson JR, White C *et al.* (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, **19**, 3708–3726.

Shimada Y, Shikano T, Merilä J (2011) A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution*, **28**, 181–193.

Swearer SE, Caselle JE, Lea DW, Warner RR (1999) Larval retention and recruitment in an island population of a coral-reef fish. *Nature*, **402**, 799.

Templeman W (1940) Embryonic developmental rates and egg-laying of Canadian lobsters. *Journal of the Fisheries Research Board of Canada*, **5**, 71–83.

Tracey ML, Nelson K, Hedgecock D, Shleser RA, Pressick ML (1975) Biochemical genetics of lobsters: genetic variation and the structure of American lobster (*Homarus americanus*) populations. *Journal of the Fisheries Research Board of Canada*, **32**, 2091–2101.

Urrego-Blanco J, Sheng J (2014) Study on subtidal circulation and variability in the Gulf of St. Lawrence, Scotian Shelf, and Gulf of Maine using a nested-grid shelf circulation model. *Ocean Dynamics*, **64**, 385–412.

Waddy SL, Aiken DE, De Klejin DPV (1995) Control of growth and reproduction. In: *Biology of the American Lobster Homarus americanus* (ed. Factor JR), pp. 217–266. Academic Press, London.

Waples RS, Gagiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries*, **9**, 423–449.

Watson FL, Miller RJ, Stewart SA (2013) Spatial and temporal variation in size at maturity for female American lobster in

Nova Scotia. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**, 1240–1251.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358.
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.
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.
Willing EM, Dreyer C, Van Oosterhout C (2012) Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, **7**, e42649.
Xue H, Incze L, Xu D, Wolff N, Pettigrew N (2008) Connectivity of lobster populations in the coastal Gulf of Maine. *Ecological Modelling*, **210**, 193–211.

L.Ben., R.R. and L.Ber. conceived and designed the study; L.Ben. performed sampling, libraries, bioinformatics and statistical analyses and wrote the manuscript; all authors significantly contributed to the improvement of statistical analyses and of the manuscript. L.Ben. is a PhD student in L. Bernatchez Lab and is interested in genomics and marine ecology of marine organisms. T.G. is a postdoc in L. Bernatchez Lab and is interested in bio-informatics and genomics of marine and freshwater organisms. C.P. is a postdoc in L. Bernatchez Lab and is interested in Evolutionary Biology of aquatic organisms. B.S.-M. is a research scientist with the Department of Fisheries and Oceans. R.R. is a faculty member at the University of New Brunswick and lead PI of the CFRN Lobster Node. L.Ber. 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 Accession No.: PRJNA281764.
- BioSample Accession No.: SAMN03492800.

The following files from this study are available from the Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.q771r.

- *Homarus americanus*, RAD sequences for putative 8144 neutral SNPs
- Geographic distances and F_{ST} values between each pair of sampling location

Supporting information

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

Fig. S1 Bayescan test for selection on individual SNPs among 17 lobster sampling locations, implemented in BAYESCAN program. Red symbols represent SNPs potentially under.

Fig. S2 Blind assignment success expressed as the percentage of lobsters sampled from one sampling location that are classified into their sampling location of origin (greyshaded numbers on diagonal) or inferred to belong to another sampling location (nonshaded numbers).

Table S1 Bayescan test for selection on individual SNPs among 17 lobster sampling locations, implemented in BAYESCAN program. Red symbols represent SNPs potentially under divergent selection whereas black symbols represent SNPs potentially neutral and green symbols represent SNPs potentially under balancing selection.

Appendix S1 Protocol for RAD Library preparation.