Copy number variants outperform SNPs to reveal genotype-temperature association in a marine species

Yann Dorant[1], Hugo Cayuela[1], Kyle Wellband[1], Martin Laporte[1], Quentin Rougemont[1], Claire Mérot[1], Eric Normandeau[1], Rémy Rochette[2] and Louis Bernatchez[1]

1) Institut de Biologie Intégrative des Systèmes (IBIS), Université Laval, Québec, QC, Canada, G1V0A6

2) Department of Biology, University of New Brunswick, P.O. Box 5050, Saint John, NB, Canada, E2L 4L5

Correspondence: yann.dorant.1@ulaval.ca

Abstract

Copy number variants (CNVs) are a major component of genotypic and phenotypic variation in genomes. Yet, our knowledge on genotypic variation and evolution is often limited to single nucleotide polymorphism (SNPs) and the role of CNVs has been overlooked in non-model species, partly due to their challenging identification until recently. Here, we document the usefulness of reduced-representation sequencing data (RAD-seq) to detect and investigate copy number variants (CNVs) alongside SNPs in American lobster (*Homarus americanus*) populations. We conducted a comparative study to examine the potential role of SNPs and CNVs in local adaptation by sequencing 1141 lobsters from 21 sampling sites within the southern Gulf of St. Lawrence which experiences the highest yearly thermal variance of the Canadian marine coastal waters. Our results demonstrated that CNVs accounts for higher genetic differentiation than SNP markers. Contrary to SNPs for which no association was found, genetic-environment association revealed that 48 CNV candidates were significantly associated with the annual variance of sea surface temperature, leading to the genetic clustering of sampling locations despite their geographic separation. Altogether, we provide a strong empirical case that CNVs putatively contribute to local adaptation in marine species and unveil stronger spatial signal than SNPs. Our study provides the means to study CNVs in non-model species and underlines the importance to consider structural variants alongside SNPs to enhance our understanding of ecological and evolutionary processes shaping adaptive population structure.

Introduction

The paradigm of local adaptation states that contrasting habitats act as selective forces, generating and maintaining variation in morphology, physiology, behaviour or life history traits that provide higher fitness under heterogeneous environmental conditions (Williams 1966; Savolainen et al. 2013; Hoban et al 2016). Therefore, investigating how genotypic variation is influenced by environmental heterogeneity is critical for understanding the evolution of adaptive traits. Local adaptation has been extensively studied in terrestrial landscapes and freshwater environments (Salvolainen et al. 2013; Manel & Holderegger, 2013). Although many recent studies have reported that local adaptation plays a major role in shaping populations in marine habitats, much remains to be done toward deciphering the genomic bases underlying local adaptation in marine species (see reviews by Bernatchez 2016; Grummer et al. 2019; Palumbi et al., 2019). In fact, understanding how local adaptation unfolds has been particularly challenging in marine species due to their generally large effective population size combined with high dispersal ability resulting in high gene flow and weak population structure as well as the complex biophysical environments found in oceans (Carr et al., 2003, Selkoe et al., 2008). In this context, studying local adaptation at the genomic level requires extensive sampling designs based on large sample sizes and dense genotypic data, particularly so in marine species.

Over the last decade, high throughput single nucleotide polymorphisms (i.e. SNPs) data have facilitated population genomics studies and substantially improved our understanding about the mechanisms underlying local adaptation (Savolainen et al. 2013). In particular, the development of reduced-representation shotgun sequencing approaches (i.e. RRS; including RADseq, ddRADseq or GBS) has allowed genotyping of SNPs across large sample sizes to address the genetic basis of adaptation in non-model species, including marine organisms (Hess et al., 2013; Hemmer-Hansen et al., 2014; Guo et al., 2015, Bernatchez et al., 2019; Sandoval-Castillo 2018; Xuereb et al. 2018). Yet, to date, studies pertaining to link genotypic variation with adaptation in non-model species has largely focussed on SNPs, representing point mutation events, whereas the role of structural variants has been overlooked for a long time due to technical and financial constraints (Wellenreuther et al., 2019). Structural variants (SVs) are genomic rearrangements affecting the presence, position or direction of a nucleotide sequence, including deletions, duplications, insertions, inversion or translocations (Mahmoud et al., 2019). SVs are pervasive variants that shape genome architecture and they can cover a larger proportion of genetic variation than SNPs (Redon et al., 2006; Wellenreuther et al., 2019; Catanach et al., 2019). They have functional consequences, notably impacting the regulation of gene expression (Gamazon & Stranger 2015) and recombination rate by diminishing the frequency of meiotic crossing-over, which can preserve the integrity of an adaptive haplotype (Rowan et al., 2019).

SVs are increasingly recognized for their prevalent role in evolutionary processes such as environmental adaptation, mating systems, reproductive isolation and speciation (Serrato-Capuchina & Matute 2018; Wellenreuther & Bernatchez, 2018; Wellenreuther et al., 2019).

Copy number variants (CNVs) are a subtype of SVs that involve insertions, deletions and duplications of DNA sequences with a size ranging from di-nucleotides repeats to millions of bases (Clop, Vidal & Amills, 2012). Their presence brings additional genomic diversity which may complement SNPs to inform about the extent of genomic divergence and can contribute to inter-individual phenotypic variation (Nguyen et al., 2006, Levy et al., 2007). For example, while studies on human initially reported that the genomes of two unrelated individuals differ by ~0.1% when considering only SNPs; the introduction of CNVs as an additional source of variation revised this assertion up to ~0.5% (Levy et al., 2007), with most of this difference due to CNVs (Redon et al. 2006; Levy et al. 2007). Although the majority of CNVs occurs within intergenic regions of eukaryote genomes (Emerson et al., 2008; Conrad et al., 2010; Makino, MCLysaght & Kawata, 2013), large CNVs may encompass entire genes, which leads to modifying downstream expression of regulatory regions as well as multiple protein-coding genes (i.e., pleiotropic effects; Gamazon and Stranger 2015). Originally associated with human genome studies (i.e. diseases, pathogen susceptibility, drug response, ancestry; Feuk, Carson & Scherer 2006; Wong et al., 2007; Ionita-Laza et al., 2009), CNVs have since been studied in a few other pecies, notably to understand domestication or speciation processes (Clop et al., 2012). Nevertheless, the role of CNVs in local adaptation relative to SNPs in non-model species is poorly understood and remains mostly restricted to a very limited number of animal and plant species in terrestrial environments (Tigano et al., 2018; Rinker et al., 2019; Nelson et al., 2018, Prunier et al., 2019), making this issue still largely unexplored in marine species. CNVs are also rarely studied on a large subset of individuals due to financial and technical constraints because classical approaches to detect CNVs, such as paired-end mapping, split read, *de novo* assembly or depth of coverage, required a reference genome and whole-genome resequencing at deep coverage (Roca et al., 2019). However, recent technological and methodological improvements in sequences analyses hold promise to address such issues. In particular, an approach has recently been proposed to use cost-efficient RADseq data to distinguish multiple-copy (e.g. duplicated, CNV) from non-duplicated single-copy loci (e.g. McKinney et al., 2017). Originally developed to deal with specific issues on whole genome duplications (WGD) and mixed ploidy problems in salmonids, McKinney's et al. (2017) method has been used to filter SNPs and to avoid miscalled genotypes, but it can also inform on CNVs.

Here, we conducted a comparative study between SNPs and CNVs to quantify their respective association with thermal conditions and thus their potential role in local adaptation in marine non-

model species, the American lobster (*Homarus americanus)*. This species appears to be a relevant model for investigating this issue as recent insights from crustacean genome assemblies indicate that repeat regions represent a prominent proportion of the genome (Song et al., 2016; Zhang et al 2019), suggesting that CNVs could play an important role in evolutionary processes in invertebrates. In addition, this ectothermic species has been reported to be physiologically sensitive to thermal regimes (Quinn & Rochette 2015*)*. First, we aimed to discover CNVs and SNPs using McKinney et al. (2017) method. Second, we compared the extent of genetic differentiation among sampling sites using the two types of markers. Third, we examined the potential role of CNVs and SNPs in local adaptation by investigating their respective associations with temperature. Overall, this work aims to highlight the role of CNVs as an important source of genetic polymorphism present in the genome of the American lobster that could lead to a better understanding of its population biology and as such improve its conservation and management.

2. Material and Methods

2.1 Sample collection and DNA extraction

A total of 1,141 lobster samples were collected from 21 sites between May and July in 2016 in the southern area of the Gulf of St-Lawrence, hereafter referred as sGSL (figure 1a and Table S1). DNA was extracted from half of the walking leg of each lobster using salt extraction (Aljanabi & Martinez, 1997) with an additional RNAse treatment following the manufacturer's instructions. DNA quality was assessed using 1% agarose gel electrophoresis. Genomic DNA concentrations were normalized to 20ng/µl based on a fluorescence quantification method (AccuClear™ Ultra High Sensitivity dsDNA Quantitation Solution). Individual reduced-representation shotgun sequencing (i.e. RRS) libraries were prepared following the Rapture approach (Ali et al., 2016). Briefly, the Rapture approach is a form of RRS sequencing which combines double-digested libraries (i.e. GBS, ddRADseq) with a sequence capture step using DNA probes designed from known genomic sequences. Here, we used the same 9,818 targeted loci previously used for the American lobster and all the details about the wet protocol are described in Dorant et al. (2019). All Rapture libraries were sequenced on the Ion Torrent p1v3 chip at the Plateforme d'analyses génomiques of the Institute of Integrative and Systems Biology (IBIS, Université Laval, Québec, Canada http://www.ibis.ulaval.ca/en/home/). Two rounds of sequencing (i.e. two separated chips) were conducted for all Rapture libraries.

2.2 Genotyping and data filtering

Single-end raw reads were trimmed to 80bp and shorter reads were removed using cutadapt (Martin,

2011). We only retained samples with at least 200,000 sequencing reads for downstream pipeline. This threshold was fixed to maximize the expected locus read depth among samples and remove poorly sequenced samples. BWA mem (Li 2013) was used to map the resulting individual-based sequences to the Rapture reference catalog previously developed by Dorant et al. (2019) that consisted of 9,818 targeted sequences in total. SNP discovery was done with stacks v.1.48 (Catchen, Hohenlohe and Bassham, 2013). A minimum stack depth of four (pstacks module : m=4) and a maximum of three nucleotides mismatches (cstacks module : n=3) were allowed. We ran the population module requiring a locus to be genotyped at a frequency greater than 60% in at least four populations. We then filtered genotypes data and characterized CNV loci using filtering procedure and custom scripts available in *stacks_workflow* (https://github.com/enormandeau/stacks_workflow). In an initial step, we filtered the raw VCF file keeping only genotypes showing a minimum depth of four (parameter "m" hereafter) and called in at least 70% of the samples in each site (parameter "p" hereafter) and an MAS value of two (parameter "S" hereafter), the *05_filter_vcf_fast.py* available in *stacks_workflow*. The MAS parameter, which is akin to the minor allele frequency (MAF) or minor allele count (MAC) filter, refers to the number of different samples possessing the rare allele. Here, we implemented the MAS filter for short-read data as it does not suffer from the same biases inherent to MAF and MAC, which are boosted by genotyping errors where one heterozygous sample is erroneously genotyped as a rare-allele homozygote. Then, we removed samples showing more than 15% of missing data. To prevent sample DNA contamination, relatedness between samples was estimated following the equation proposed by Yang et al. (2010) and implemented in vcftools. One of each of the two pairs of samples originating from the same sampling site and showing a relatedness level > 0.90 was excluded in the goal to avoid potential errors from libraries preparations. Finally, the inbreeding coefficient was estimated for each sample using a method of moments implemented in vcftools. From graphical observation of sample inbreeding, we defined a cutoff value (i.e. -0.25) to exclude abnormal samples. We then removed all bad samples identified from the raw VCF file and then re-ran the *05_filter_vcf_fast.py* from *stack_workflow*, keeping the same parameters previously used (i.e. m=4; p=70; S=2).

2.3 Identification of copy number variants

To explore locus duplication and then characterize putative CNVs, we based our analyses on the previous work by McKinney et al (2017). In particular McKinney et al. (2017) used population genetics simulations as well as empirical data and demonstrated that a set of simple summary statistics (e.g. the proportion of heterozygous and read ratio deviation of alleles) could be used to confidently discriminate SNPs exhibiting a duplication pattern without the need for a reference genome. Thus, we

investigated SNPs "anomalies" based on a suit of four parameters to discriminate high confidence SNPs (hereafter singleton SNPs) from duplicated SNPs (i.e. (1) median of allele ratio in heterozygotes (MedRatio), (2) proportion of heterozygotes (PropHet), (3) proportion of rare homozygotes (PropHomRare), (4) Fis. Each parameter was calculated from the filtered VCF file using the *08_extract_snp_duplication_info.py* available in *stacks_workflow.* The four parameters calculated for each locus were plotted against each other to visualize the distributions of these parameters across all loci. Based on the methodology of McKinney et al (2017) and by plotting different combinations of each parameter, we graphically fixed cut-off for each parameter. Full details of this step are available in table S2 and figure S2. Finally, two separated data sets were generated: the "SNP dataset", keeping only singletons and the "CNV dataset", keeping only duplicated SNPs. To construct the SNP dataset, we kept the genotype calls from the VCF file containing singleton SNPs, post-filtered by keeping all unlinked SNPs within each locus using the *11_extract_unlinked_snps.py* available in *stacks_workflow*. Briefly, the first SNP is kept and all remaining SNPs showing strong genotype correlation are pruned (i.e. Two SNPs show strong genotype correlation if samples with the rare allele in one of the SNPs have the same genotypes as samples with the rare allele in the other SNP more than 50%). The procedure was repeated until all SNPs were either kept or pruned. To construct the CNV dataset, we extracted the locus read depth of SNPs identified as duplicated using vcftools. Note that the CNV dataset contains duplicated loci that could be invariant in copy number among the 21 sampling sites. Hence, for more simplicity, we use the term "CNV" to represent all loci classified as duplicated by our approach. As libraries sequenced at a greater depth will result in higher overall read counts, CNV locus read counts were normalized to account for differences in sequencing coverage across all samples. Normalization was performed using the Trimmed mean of M-values method originally described for RNA-seq count normalization and implemented in the R package edgeR (Robinson & Oshlack, 2010). The correction accounts for the fact that for an individual with a higher copy number at a given locus, that locus will contribute proportionally more to the sequencing library than it will for an individual with lower copy number at that locus.

2.4 Genetic differentiation analyses

We estimated pairwise F_{ST} values for the SNP dataset (Weir & Cockerham, 1984) using *stamppFst* function from the R package StAMPP 1.5.1 (Pembleton et al., 2013). To estimate population genetic differentiation of loci identified as CNVs, we calculated the variant fixation index V_{ST} (Redon et al. 2006). V_{ST} is an analog of F_{ST} estimator of population differentiation (Wright, 1949; Weir & Cockerham, 1984) and is commonly used to identify differentiated CNV profiles between populations

(Redon et al., 2006; Dennis et al., 2017; Rinker et al 2019). For each pairwise population comparison, V_{ST} was estimated considering $(V_T-V_S)/V_T$, where V_T is the variance of normalized read depths among all individuals from the two populations and V_S is the average of the variance within each population, weighed for population size (Redon et al 2006). Then, we compared the magnitude of population differentiation estimated from SNPs and CNVs, respectively. Finally, we performed principal components analyses (PCAs) to visualize the pattern of genetic differentiation based on either all SNPs or all CNVs.

2.5 Genotype-Environment Association Analysis

Marine climatic data were extracted from geoTiff layers of the MARSPEC public database [http://marspec.weebly.com/modern-data.html] (Sbrocco & Barber, 2013). Five environmental layers (30 arcseconds resolution, i.e. ~1km) related to Sea Surface Temperature (hereafter SST) were considered. These include the annual mean, annual range, annual variance, the minimum value observed for the coldest month and the maximum value observed for the warmest month. We focused on Sea Surface Temperature, since this environmental parameter is one of the most critical for larval deposition and survival for the American lobster (Quinn 2017). Furthermore, lobsters were collected by Canadian fishers close to the shoreline within < 20m depth. At such shallow depths, we expected that sea surface parameters can be used as a reliable proxy of temperature for the entire water column. For all temperature layers, we defined a circular buffer of 15km radius around each of the 21 sampling sites and values within this buffer were averaged to minimize pixel anomalies due to known biases of correction algorithms used for remote sensing, especially in coastal areas (Smit et al 2013). Collinearity between explanatory environmental variables was checked using the Pearson correlation index as well as a Draftsman's plot. Among variables exhibiting high level of correlation (i.e. $r > 0.6$), we kept only one of them based on its biological relevance (see explanations in Results).

We performed a redundancy analysis (RDA) to investigate the association between environmental variables and genetic variation (either SNPs or CNVs data sets) using the vegan R library (Oksanen et al., 2018) and following Forester et al. (2018) (see https://popgen.nescent.org/2018-03-27_RDA_GEA.html for details). We applied a forward stepwise selection process to environmental predictors via the *OrdiR2step* function (1,000 permutations; limits of permutation $P < 0.05$) in order to identify environmental variables that significantly explained overall genetic variation (i.e. maximising adjusted R^2 at every step). Ultimately, global and marginal analyses of variance (ANOVA) with 1,000 permutations were performed to assess the significance of the models and evaluate the contribution of each environmental variable. Once genetic markers were loaded against the RDA axes, candidates for

positive response with environmental predictors were determined to be those that exhibit more than 2.5 standard deviations away from the mean ($P < 0.012$).

We also used the Latent Factor Mixed Models implemented in LFMM 2 (Caye et al,. 2019) for SNPs and linear mixed-effects models (LME) implemented in the lme4 R library (Bates et al., 2015) for CNV data. This second analysis aimed to document the form and strength of the relationship between genetic markers and environmental predictors. In the LME model, sampling sites were selected as a random effect and environment parameters as the explanatory variables. Then, the genetic information (i.e SNPs genotypes for singleton loci and normalized read depth for the putative CNV loci) were introduced in the model as the response variable. The effect of temperature predictors was assessed using ANOVA F-test applied on regression model output. To account for multiple testing issues, we used a false discovery rate (FDR) following the method of Benjamini & Hochberg (1995) for both LFMM and LME results. The p-value was adjusted to control threshold for statistical significance in multiple comparisons at $\alpha=0.01$. To minimize potential false positives among candidate markers, we defined the set of best candidate loci for local (thermal) adaptation as those that were found to be associated with at least one environmental factor in both RDA and statistical regression models (i.e. LFMM and LME).

The relationship between CNV read depth (normalized) and the environmental predictors displayed a non-linear growth-like model for most candidates CNVs (see results). Hence, we evaluated the fit of Gompertz (i.e. four parameters) and logistic functions (i.e. four to five parameters) for each locus. Model selection was done using the function *mselect* from the R library *drc* v3.0-1 (Ritz, 2015) and the model exhibiting the smallest log-likelihood coefficient was assumed to have the best fit to the data. Additionally, we assessed the clustering pattern among the 21 sampling sites using a Principal Component Analysis (PCA) on the candidate CNVs associated with the environmental predictors. The number of "meaningful" principal component (PCs) axes to retain for interpretation and downstream analyses was assessed based on the broken stick distribution (Legendre and Legendre, 2012).

2.6. Statistical discretization of copy groups.

To obtain reliable and robust reference points to interpret the copy number information data obtained from the distributions of normalized read depth, we performed a model-based unsupervised clustering approach using *Mclust*, V. 5.4.4, for R (Fraley et al., 2012). *Mclust* uses finite mixture estimation via iterative expectation maximization steps (EM) for a range of k components and the best model is selected using the Bayesian Information Criterion (BIC). For each locus, individuals were classified in discrete clusters read depth (that we coin copy number groups) that likely reflect variation in copy

number of any given locus. Based on the copy number group characterizing each individual for a given locus, we were then able to assess the geographic distribution of copy number groups across all 21 sampling sites.

3. Results

3.1 Data processing and classifying singleton and CNVs loci

Rapture sequencing yielded an average of 431,350 ± 172,538 reads per sample before any quality filtering. From the 1,141 samples sequenced, 60 were removed based on individual's quality filtering (i.e. 28, 25, 7 and 0 individuals for low sequencing depth threshold, up to 15% missing data, abnormal pattern of heterozygosity and abnormal relatedness, respectively). From the 1,081 remaining individuals, the median read depth per individual calculated among each sampling sites ranged from 364,437 ± 86,385 reads for T4P4 to 561,877 ± 219,624 reads for T3P1 (table S1). After the first SNPs calling process, 26,005 SNPs were successfully genotyped in at least 70% of the individuals forming a pre-cleaned VCF file prior to the identification of singleton vs. duplicated SNPs. The global genotyping rate was high with 3.5% of missing data on average.

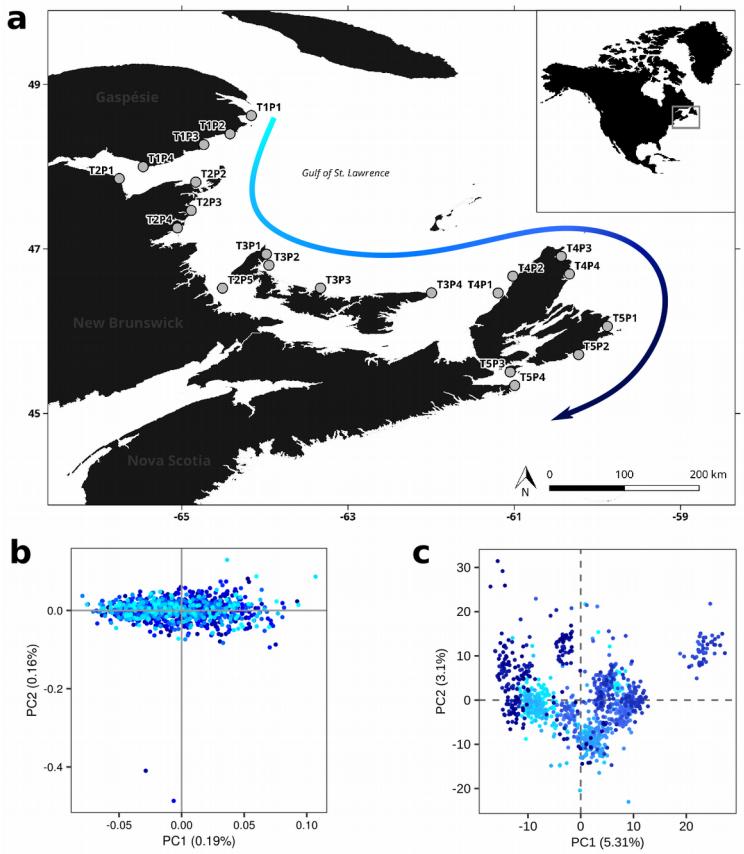


Figure 1: Geographic distribution and genetic divergence of American lobster in southern Gulf of St. Lawrence.

a. Map of American lobster sample collection. The blue-scale arrow represents the direction of the major current pattern within the region during the larval drift period between July and September(Galbraith et al., 2015). **b.** PCA plot inferred from the 13,854 SNPs genotyped. **c.** PCA plot inferred from normalized read-depth of 1,521 CNV loci identified. PCAs display the individuals spatially colored from the site T1P1 to the site T5P4 using the color-scale depicted by the arrow of ocean currents represented in the sampling map (1a).

The characterization of SNPs returned 14,534 SNPs as singletons and 9,659 SNPs as duplicated (figure 2; see table S3 and figure S2 for details). The final SNPs singleton data set contained 13,854 SNPs after keeping only unlinked SNPs within loci spread over 5,362 loci of 80bp. Finally, the CNV dataset consisted of read depth information extracted from the 1,521 loci which comprises the 9,659 SNPs classified as duplicated.

3.2 Measuring genetic differentiation

Pairwise population differentiation based on the 13,854 SNPs was extremely weak with F_{ST} values ranging from 0.0000 to 8.6^{e-4} (average F_{ST} = 0.0001 with 93% non-significant pairwise F_{ST} values). In contrast to SNPs, the genetic differentiation estimated using the V_{ST} index and across the 1,521 CNV loci exhibited a stronger and significant signal of genetic differentiation, with a V_{ST} value ranging from 0.01 to 0.10 (average V_{ST} = 0.04 and all pairwise values were significant).

SNPs PCA based on individual genotypes among all samples did not reveal any significant pattern of genetic differentiation among the 21 sampling sites (figure 1b). On the other hand, PCA based on individuals' CNV data revealed an unexpected pattern of spatial structure not matching geographic proximity only (figure 1c). Based on the first principal component, which explained 5.31% of the variance, we thus observed that the majority of lobsters sampled from sites at both edges of the sampling area (i.e. individuals illustrated by lightblue and darkblue colors and from sampling locations labelled as T1P., most of T2P. sites and T5P.) was more similar to each other than with the individuals collected from central sites (T3P. and T4P. sites). Furthermore, this spatial discontinuity did not reflect the shape and direction of the main current pattern observed in this region and illustrated in the figure 1a.

3.3 Genotype - environmental association

Environmental variables showed high level of correlation (figure S1). Hence, we only selected two SST variables (*i.e.*, SST annual minimum and annual variance) as predictors for the RDA analysis (SST values detailed in table S1). First, we selected the SST annual minimum because this predictor showed the lowest level of collinearity among the five thermal parameters considered. Moreover, this parameter has also been associated with thermal adaptation in the American lobster at a larger geographical scale than studied here (Benestan et al., 2016). Second, because most marine species are adapted to less thermal variability compared to terrestrial environments (Sunday et al., 2011), we considered that the SST annual variance was the most relevant factor which may influence biological processes (e.g. growth and survival) of the American lobster. In particular, the SST variance parameter

reflects the unpredictability of thermal conditions (e.g. extreme thermal events) through which pelagic larvae can be confronted. Furthermore, Larouche & Galbraith (2016) also observed that the southern part of the Gulf of St. Lawrence experiences the highest thermal amplitude of the Canadian marine coastal waters.

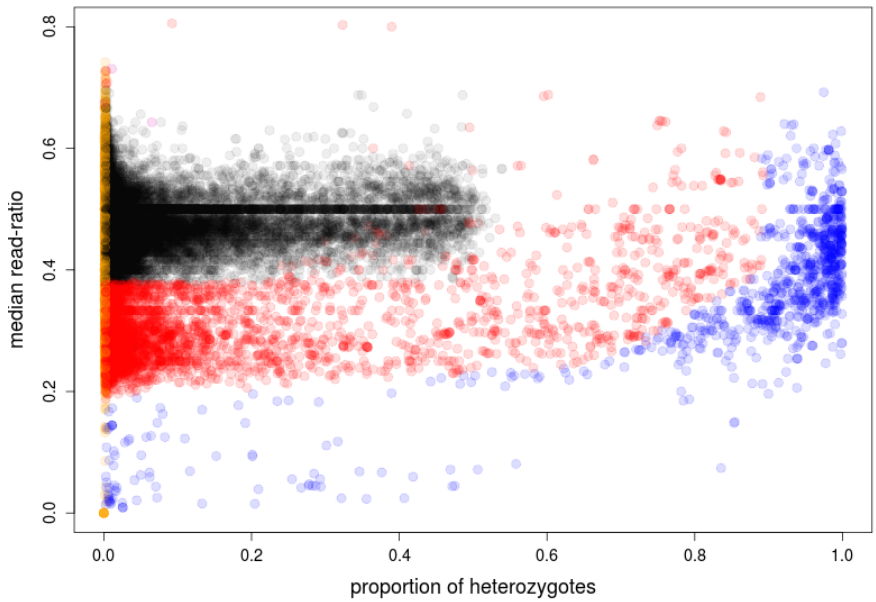
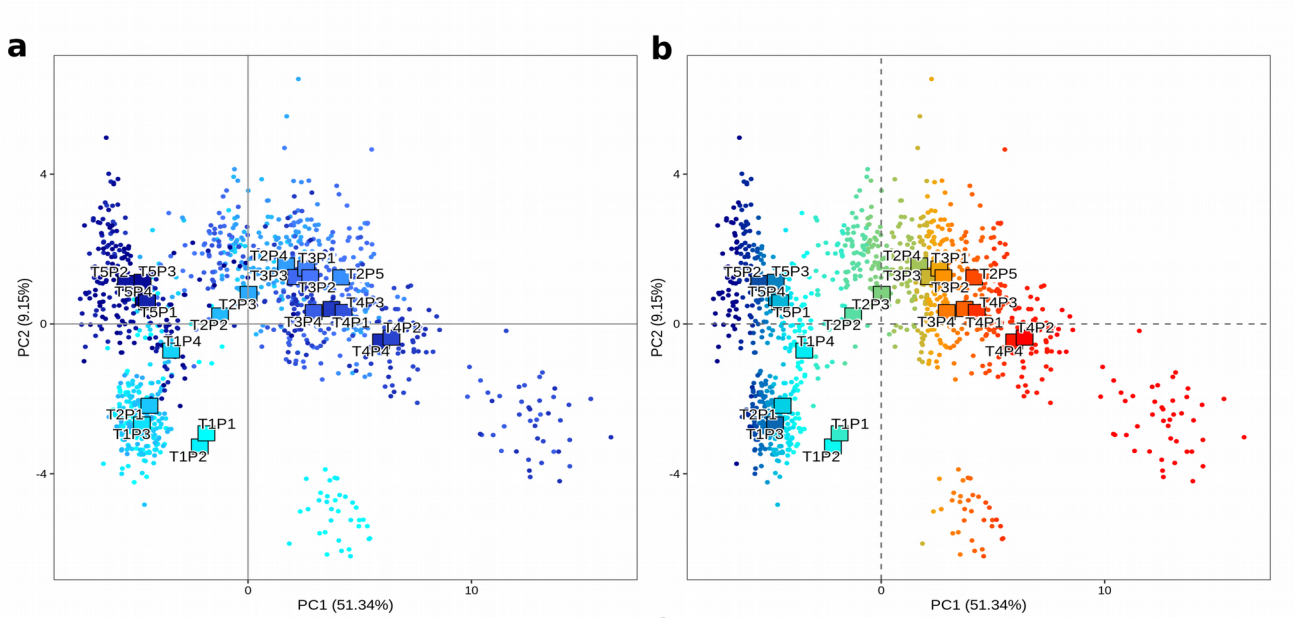


Figure 2: Characterization of duplication effect over the SNP dataset.
The bivariate scatter plot display the distribution of the 26,005 SNPs over the median read-ratio deviation of heterozygotes (y-axis) plotted against the proportion of heterozygotes (x-axis). The median read-ratio describes the deviation from equal alleles read-ratio (50/50) expected for heterozygotes. Black, red, blue and orange points represent singletons, duplicated, diverged and low confidence SNPs, respectively. Here, we only represent the most informative graphical representation of SNPs classification, which is derived from the graphical pattern of SNP categories (i.e. singleton, duplicated, diverged) demonstrated by McKinney et al. (2017) with data simulations as well as empirical analyses, Other graphical representations are presented in figure S2.

No association between environmental variables and SNP genetic variation was detected by the RDA (i.e. ordistep adj. $R^2 = 0$ for SST annual minimum and SST annual variance respectively) and the LFMM 2 analyses (data not shown). In contrast to SNPs, the step-wise model selection revealed that both SST annual variance and SST annual minimum were significantly associated with CNV variation across sampling sites with the strongest association observed for SST annual variance (ordistep adj. $R^2 = 0.105$; $P < 0.001$) relative to SST annual minimum (adj. $R^2 = 0.066$; $P < 0.001$). The RDA conducted with selected predictor (i.e. SST annual variance) and the CNV dataset was highly significant ($P <$

0.001; ANOVA 1,000 permutations) and explained 2.4% of the total CNV variation (adj.R^2 = 0.024). A total of 83 loci were detected as candidate outliers, where 67 and 16 were associated with SST annual variance and SST annual minimum, respectively. Linear mixed-effects models detected 288 CNV loci significantly associated with annual SST variance (adj. $P < 0.01$). Among these 288 CNVs, 48 were also identified from the RDA (i.e. 15.6% in common among all outliers detected with both methods combined). Although the LME models returned 19 CNV loci significantly associated with the SST annual minimum, none were in common with those 16 candidates identified by the RDA. V_{ST} estimation for all 1,521 CNV loci revealed nine outlier loci with substantially elevated levels of population differentiation (i.e. V_{ST} values above a threshold of six standard deviations) and all of them were identified in association with the SST annual variance.

The PCA performed with the set of 48 CNV candidates associated with the SST annual variance showed that sampling locations formed two distinct groups that were mainly discriminated by the first principal component (PC1), which explained 51.34% of the variance (figure 3a. 3b). Sampling sites located at both edges of the sampling area (i.e. T1P1, T1P2, T1P3, T1P4, T2P1, and T2P2 in the north; T5P1, T5P2, T5P3 and T5P4 in the south) were clearly separated from those sampling locations from the center part of the studied range (i.e. T2P4, T2P5, T3P1, T3P2, T3P3, T3P4, T4P1, T4P2, T4P3 and T4P4). The sampling site T2P3 displayed an intermediate position. Based on a Broken-stick model, only the first principal component was considered for interpretation. The geographical transposition of PC1 scores obtained from sampling sites centroids within the climate layer of the annual SST variance showed a strong association with CNV variation which was somewhat independent of geographic proximity (i.e. most northern and most southern samples being similar relative to centrally localized samples) (figure 3c). Although the Broken-stick model supported only the first PC, we also observed that PC2 separated sites located at both edges of the sampling area, albeit to a lesser extent that the separation observed on the first axis between north-south grouped samples vs. central locations (figure 3b).



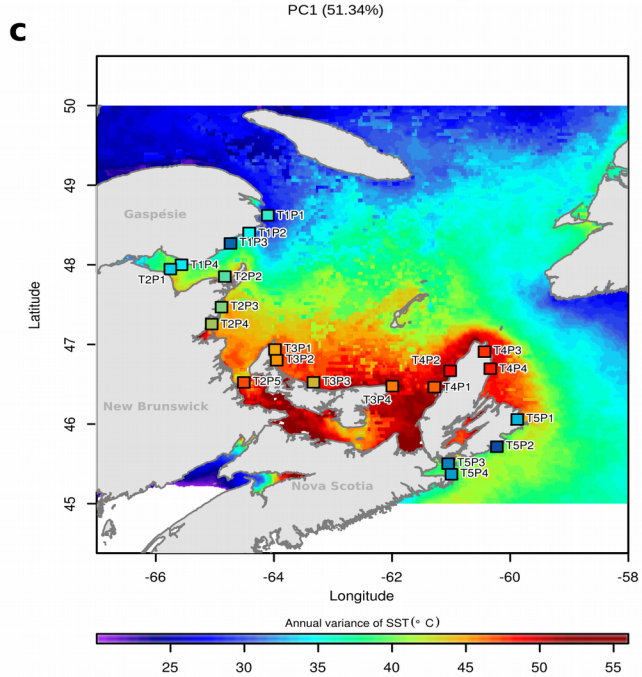


Figure 3 : PCA based analysis of the environmental associated CNVs.
PCA was applied to the normalized read depth matrix of 48 CNV loci associated with SST annual variance. For each PCA plot (a and b), points and squares represent individuals and sampling sites (i.e. centroid position) respectively. For (a) the color scale represents the main ocean current displayed in figure 1a. For (b) the colors from blue to red are scaled by the first PC1 axis. Note the clustering of sampling locations based on PC1 (namely those labelled as T1P. and T5P.) despite the geographic separation). (c) Climate map depicted the annual variance of sea surface temperature within the south of the Gulf of St. Lawrence. Sampling sites are represented by squares and colored according matching the color scale used in the PCA (i.e figure 3b).

The shape of the relationship between read depth (normalized) at all 48 CNV candidates and SST annual variance was quite similar (figure 4 and figure S4). Although these 48 CNV candidates were identified by linear mixed-effect models, our analyses revealed that this relationship was non-linear for 43 loci and was generally best described by a Gompterz function or logistic functions for 35 (73%) and 8 (16%) CNV loci respectively (figure S5). For most of the 43 CNV loci best described by a non-linear function, read depth remained stable until a SST annual variance value of ~45, and then drastically increased and stabilized around a value of ~50 (figures 4 and S4).

Based on the results above, we defined two main clusters: (i) the first one consisted of sampling sites experiencing low values of SST annual variance (i.e. $\sigma^2_{Temp} < 45$) and (ii) the second one consisted of

sampling sites experiencing high values of SST annual variance (i.e. $\sigma^2_{Temp} > 50$). The level of genetic differentiation estimated using the 1,521 CNV loci was higher between these two clusters than within each cluster. The average of pairwise V_{ST} values were 0.0548 (sd = 0.0131) and 0.0298 (sd = 0.0149) for *between clusters* versus *within cluster* respectively, and the Kolmogorov-Smirnov test showed that this difference was significant (KS: D = 0.616, P < 0.001).

3.4 Discrete groups of CNVs

Copy number information revealed by the variable distributions of normalized read depth among the 21 sampling sites was investigated further by the identification of discrete "copy groups" based on the "mclust" procedure (figure 4). This revealed that within each of the 48 CNV loci significantly associated with the SST annual variance, up to five discrete clusters putatively representing different levels of copy number were identified depending of the given locus (see figure S6, S7, S8 for details). While the first copy number group (the one best supporting a unimodal distribution of coverage) could represent individuals with a single copy for a given locus, we prefer to remain conservative in absence of a reference genome to firmly confirm this. We therefore prefer referring to copy number group 1 to 5, with 1 representing the group with the lowest copy number and 5, the highest. We observed spatial variation in the distribution of lobsters showing putatively different levels of copy numbers whereby the low-level copy number groups were preferentially distributed over sampling sites characterized by low SST annual variance (at the two edges of the sampling locations), while groups with higher level of copy numbers were distributed over sampling sites with high SST annual variance located in the central range of the studied sites (figure 4 and S7).

Discussion

Our study revealed that CNVs represent a non-negligible fraction of genetic variability in the American lobster genome which can be successfully detected within reduced representation sequencing data combined with McKinney's et al. (2017) framework. Our results showed that the level of genetic differentiation among the 21 sampling sites was significantly higher with CNVs than with SNPs. More importantly, we did not detect association between SNPs and any temperature variables whereas a strong and significant pattern of association was found between a set of CNVs and temperature variance, suggesting that variation in copy numbers at those loci could be involved in local (thermal) adaptation. Overall, our study supports the view that CNVs represent a significant portion of the genetic polymorphism and in turn, are relevant genomic markers for population genomic studies.

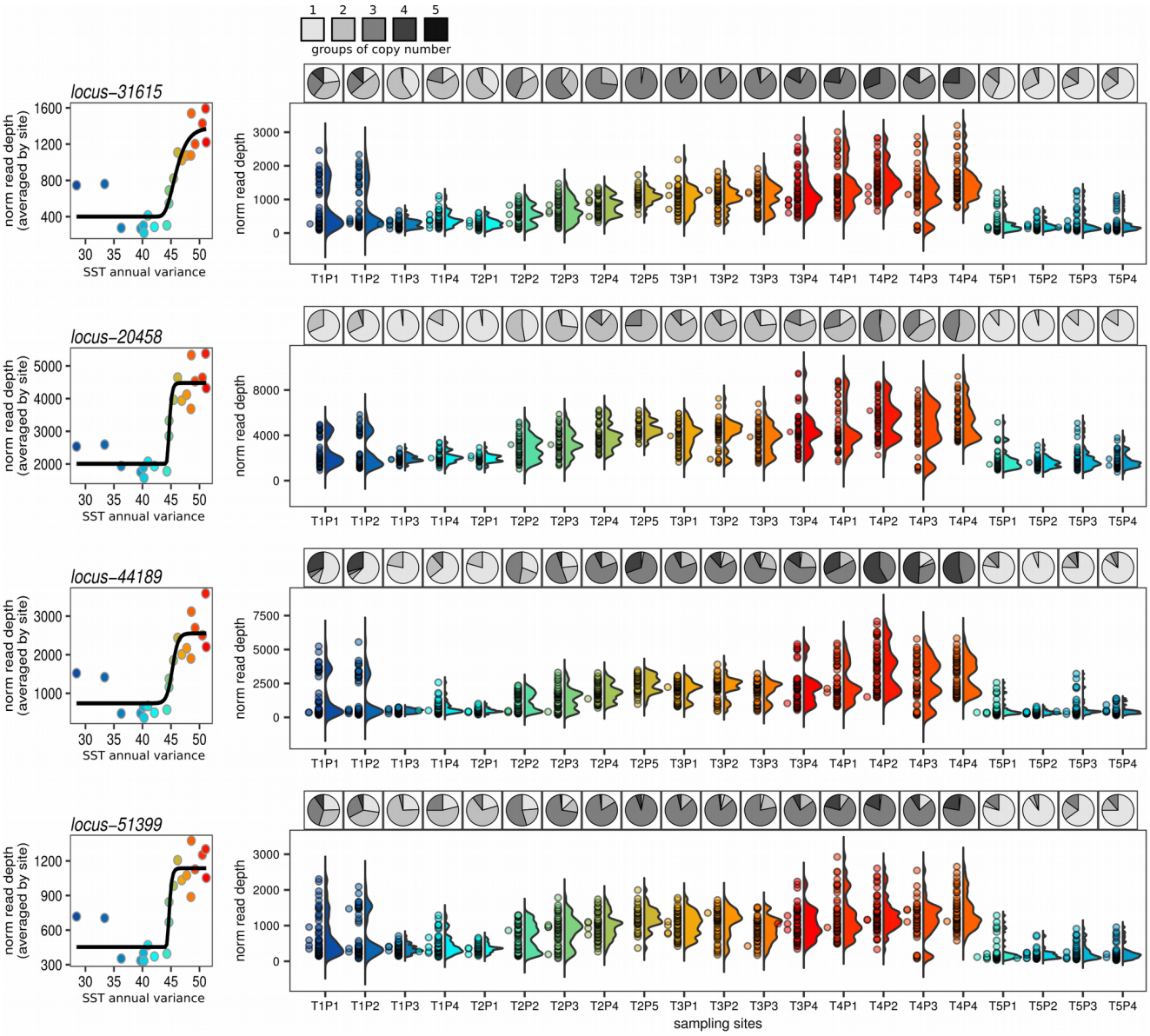


Figure 4: Overview of CNV profiles for the four most highly differentiated loci (VST) associated with the SST annual variance. (Left column panel) Correlation plot between SST annual variance and the average read depth observed at each sampling site. Black lines represent the Gompertz fit model for each locus. (Right column panel) For each sampling site (x-axis), half-violin plots represent the density of normalized read depth distribution (y-axis) while points represents individual lobsters. The colour for the different sites is according to their respective position depicted in the PCA (Figure 3b). Header pie charts represent the proportion of individuals classified within each putative 'copy group' (from 1 to 5) among each sampling site. For each CNV locus the putative number of "copy groups" was estimated based on independent EM algorithm (see M&M for details). White to black box-scale represents the different groups of copy number.

On the importance of detecting CNVs

Reduced-representation shotgun sequencing (RRS,, which include RADseq, ddRAD, GBS and Rapture) is a powerful tool for non-model systems and large-scale studies since RRS allow sequencing a broad number DNA sequences for thousands of samples simultaneously at a minimal cost (Davey et al., 2011). However, variation in copy-number at some loci is often confounding for DNA sequence alignment methods and can affect polymorphism inferences, leading for instance to errors in SNPs calling at those duplicated loci and bias in allele frequency estimation (McKinney et al., 2017). In this study, we improved confidence in SNP calling from RRS sequencing data by classifying and considering separately loci with copy-number variability, the CNVs, based on the methodology developed by McKinney et al (2017) (i.e. HDplot). Simulations and empirical data by McKinney et al. (2017) support the robustness of this approach, based on a set of simple summary statistics of SNP dataset, by showing that it correctly identified duplicated sequences with >95% concordance with loci of known copy status. The main limit identified by McKinney et al. (2017), is that the main statistic, *i. e.* , the read ratio deviation can be affected by the stochastic nature of RRS data, in particular at low to moderate sequencing depth (McKinney et al., 2017). Yet, this should be minimal in our study since the Rapture approach enable us to sequence 384 RRS captured libraries within the same sequencing chip, reducing unbalanced representation of individuals by a high rate of multiplexing. Moreover, genotype calling quality was assured by a high average read-depth per locus (average across all markers was ~31X reads/locus), leading to very few missing data over individual genotypes. Beyond providing a more polished SNP dataset, our classification allowed analyzing a new dimension of genetic variability, a CNVs dataset with the same original RRS sequencing data. The interpretation of CNV profiles among populations from the distribution of normalized read depths was also improved through the delineation of different discrete groups of copy number using a model-based clustering approach. This approach, which includes the Mclust procedure used here (Fraley & Raftery, 2012) or other methods such as Bayesian mixture model based clustering (Medvedic et al., 2004), offers advantage over heuristic methods (e.g. k-means clustering), such as the ability to measure uncertainty about the resulting clusters and to formally estimate the number of clusters (Fraley & Raftery 2007). Furthermore, model-based clustering procedures have been broadly used in genetics and ecology, for instance in studies using microarray to document patterns of gene expression or cancer cell type identification, as well as life-history traits variation (Freyhult et al., 2010; Côté et al. 2015; Barajas-Olmos et al., 2019). Altogether, the set of methods presented in this study thus provides a way to maximize the genetic information drawn from RRS sequencing by improving the accuracy of the SNP dataset and by providing a CNV dataset. Although the approach faces the same limits as other CNV

detection methods based on SNPs arrays or whole-genome sequencing (e.g. mappability issues, GC content, PCR duplicates, DNA libraries quality; Teo et al., 2012), it also provides the means to consider CNVs in non-model species even without extensive genomic or financial resources. Here, for instance, with a genome size of about 4.5Gb (Jimenez et al., 2010), and in absence of a reference genome, it would have essentially been impossible to screen for CNV variation across the entire genome for hundreds of individuals.

Copy number variants represent a substantial fraction of genetic diversity in the lobster genome

By applying the above-mentioned classifying methods, we identified that ~20% of the loci successfully sequenced represented copy number variation in the American lobster. It is well established that the amount of CNVs involving various structural variants such as whole genome duplication, transposable elements or duplication, deletion and rearrangements of various repeated DNA, are implicated in the change of genome size among eukaryotes (Biémont, 2008; Cioffi & Bertollo., 2012; López-Flores & Garrido-Ramos, 2012). For example, Prokopowich, Gregory and Crease (2003) showed a strong relationship between rDNA copy number and the genome size of 162 species of plants and animals. Thus, we expect that the 4.5 Gb genome of American lobster is made up of an important volume of repeated regions such as the CNVs detected in this study. This is supported by recent work on other crustacean genomes, which revealed prominent proportions of repeat-rich regions (Song et al., 2016, Verbruggen 2016; Yuan et al., 2018, Fang et al., 2019). For instance, up to 23.93% of the genome (~1.66 Gb) have been reported as repeated DNA sequences for the Pacific white shrimp, *Litopenaeus vannamei* (Zhang et al. 2018). Overall, this supports the hypothesis that CNVs may represent a significant proportion of the variation present in the American lobster' genome and could play functional roles.

CNVs reveal a stronger genetic structure than SNPs

Within the southern part of the Gulf of St. Lawrence, we observed an extremely weak level of genetic structure at SNP markers (average F_{ST} = 0.0001 and 93% of non-significant pairwise F_{ST}, i.e., p-value > 0.05), a pattern congruent with the findings of Benestan et al. (2015) within the same geographic region. Our results are also in accordance with many previous studies highlighting a weak genetic structure in other marine organisms (fishes, DiBattista et al., 2017; Junge et al., 2019; molluscs, Sandoval-Castillo et al., 2018; crustaceans, Al-Breiki et al., 2018) using SNPs. Similarly to these studies, the small F_{ST} values observed in our study most likely reflect a pattern of high gene flow and/or large effective population size among lobsters from the 21 sampling sites. Bio-physical larval dispersal

models showed that such a high connectivity could be due to high dispersal capacity of lobsters during early life stages (dispersal distance ranging from 5-400 km; Chassé & Miller 2010; Quinn et al., 2017). In addition, a recent mark-recapture study revealed that benthic movements at the adult stage also likely contribute to high gene flow and high connectivity in the American lobster (Morse et al., 2018).

In contrast to SNPs, CNVs showed higher level of differentiation in the study area (*i.e.* average V_{ST} = 0.043 and all pairwise V_{ST} were significant). This is consistent with previous studies in humans showing that CNVs contribute to higher genetic divergence than SNPs (Levy et al., 2007; Pang et al., 2010; Sudmant et al., 2015). Furthermore, recent studies in marine fishes also showed that reduced genomic regions involving structural variants account for higher genetic divergence than widely distributed SNPs (Berg et al., 2016; Barth et al 2019; Kess et al., 2020). Several hypotheses may explain this pattern. For instance, SVs such as inversions or some categories of CNVs locally limit recombination (Rowan et al., 2019), thus reducing genome homogenization despite gene flow. Some CNVs like transposable elements exhibit high evolutionary dynamics and have important implications for rapid genome evolution and adaptation (Kofler, Nolte and Schlötterer, 2015; Rey et al., 2016). Finally, by affecting a larger fraction of the genome or by having stronger effects on the phenotype, CNVs may be particularly sensitive to positive or negative natural selection. Altogether, although the mechanisms remain poorly known, the increasing evidence that CNVs underlies more differentiation than SNPs lead to consider CNVs as powerful genetic markers and to take advantage of present technology to include them in future population genomic studies aiming at characterize spatial genetic structure.

CNV-environment associations are stronger than SNP-environment associations

Although redundancy analysis (RDA) and LFMM represent suitable tools to identify candidate SNPs associated with environmental factors (Forester et al., 2018; Caye et al., 2019), forward model selection revealed no significant relationship between any of the 13,854 SNPs and the environmental variables tested in the analyses. This result differs from the findings of Benestan et al (2016), reporting a strong association between 505 SNPs and thermal conditions (i.e. minimum annual SST) in the American lobster. However, Benestan et al. (2016) sampling zone covered most of the species' habitat in North Western Atlantic, and the genotype-environment association was largely driven by a north vs south dichotomy, with the southern part of Benestan et al. (2016) study range not investigated here. This north-south genotype-environment association is also a prominent macroecological phenomenon, having been documented in four other species (i.e. Atlantic cod - *Gadus morhua*, European green crab -*Carcinus maenas*, Northern shrimp - *Pandalus borealis*, and the Sea scallop - *Placopecten*

magellanicus) (Stanley et al., 2018). In marine studies, broad-scale sampling regimes are often used and these are limited to examining large-scale environmental gradients which tend to overlook habitat heterogeneity at local scales.

Contrary to SNPs, our analyses detected 48 CNV candidates out of 1,521 tested (normalized for sequencing effort over all samples), that were significantly associated with the SST annual variance, supporting the hypothesis that variation in copy numbers at those markers could be associated with local adaptation to temperature variance in the American Lobster. Several studies have shown distinct roles of CNVs in local adaptation in terrestrial plants and animals (e.g. Tigano et al., 2018; Weetman et al., 2018; Bazzicalupo et al., 2019; Nelson et al., 2018; Prunier et al., 2019; Ruytinx et al., 2019) but empirical studies on CNVs in marine systems remain limited. Previous studies in marine fishes supported a role for CNVs in freeze resistance (Hew et al., 1988; Hayes et al, 1991, Chen et al., 2008; Desjardins et al., 2012). For instance, Hayes et al. (1991) used DNA electrophoresis and southern blots to highlight differences in antifreeze proteins (AFPs) gene copy number and arrangement among some winter flounder (*Pseudopleuronectes americanus*) populations along the Northeastern Atlantic coast. Similarly, Desjardins et al. (2012) found that multiple copies of AFPs genes improved gene-dosage effect and also transcript level between two close-related wolf fish species (*Anarhichas lupus* and *Anarhichas minor*), facilitating the colonization of a shallow-water habitat where the risk of freezing is elevated. By showing empirical evidence for CNV-environment association, our results thus bring additional evidence for the importance of structural variation for local adaptation in marine species.

Processes underlying local adaptation by CNVs can imply various complex mechanisms such as remodeling chromatin architecture (Dunaway et al., 2016) or the modulation of gene expression (i.e. gene dosage-effect; Desjardin et al., 2012, Zhang et al. 2019). Although the alteration of gene-dosage balance is often prejudicial to fitness and have been mostly documented with diseases (Gamazon & Stranger 2015; Rice & McLysaght., 2017), in some cases, gene dosage-effect induce by CNVs can be positive and facilitate the adaptation of organisms to environmental changes. For example, CNVs have been highlighted as drivers of insecticide resistance in populations of *Aedes* mosquitoes by increasing the effectiveness of detoxification enzymes (Faucon et al., 2017, Weetman et al., 2018). Furthermore, multiple gene copy number have been highlighted as substantial drivers of gene expression differences between populations of three-spined stickleback (*Gasterosteus aculeatus*), which facilitate phenotypic variation and habitat-specific selection to contrasted environments (Huang et al., 2019).

Interestingly, we found that it was the variance of environmental conditions (i.e. SST annual variance) that was best associated with CNVs instead of commonly reported average, minimum or maximum values. Organisms often use phenotypic plasticity as a strategy for maximizing their fitness

in variable environments (Schlichting & Pigliucci 1998; Aubin-Horth & Renn, 2009; Laporte et al., 2016). It is possible that annual temperature variation could be compensated for by a gradation of gene copy silencing throughout the year. Indeed, epigenetic mechanisms are extremely labile and affect gene expression (McCue et al., 2012; Rougeux et al., 2019). For instance, DNA methylation can change in minutes following an environmental change and return in their initial configuration in days in previous environmental conditions (Huang et al., 2017). Furthermore, environmental stress can also induce a reactivation of transposable elements via DNA demethylation, which could relatively quickly result in an increase of CNVs through the genome that will be submitted to selective pressures in the new environment (Rey et al., 2019). Here, the particular repeated observation of a similar non-linear relationship among 43 CNV loci associated with the annual variance of SST raises questions about the nature of this mechanism within the genome (e.g. large DNA region successively repeated, transposable elements, etc.). However, the inherent features of our sequencing approach, which represents only a reduced sample of the entire genome, coupled with the absence of a reference genome for American lobster, does not allow us to address this issue. Future works should be able to address this by taking advantage of technologies advances in long-read sequencing (i.e. Nanopore or PacBio technologies) as well as the decreasing costs to develop references genomes in non-model species.

Admittedly, the possible mechanisms by which variation in CNVs in association with thermal variance could indeed be involved local adaptation remains hypothetical here. Nevertheless, our observations make a strong point regarding the relevance of investigating such mechanisms in future studies while the possible adaptive role of CNVs have been largely overlooked in most studies thus far, and particularly so in marine species.

Conclusion

Our study highlights the importance of structural variants such as CNVs in evolutionary genomics by showing their implication in genetic differentiation and local adaptation in a marine species. Moreover,, our contribution provides an approach to characterize and analyze CNVs with reduced-representation sequencing, an approach which is applicable with limited financial resources and without reference genome. We believe that despite some limitations as identified above such development will enable to take advantage of CNV markers to better characterize population structure of non-model species for conservation or management. Moreover, extending CNVs characterization and analysis to a wider range of species and ecological contexts will allow better understanding the ecological and evolutionary significance of those structural variants.

Data accessibility

(1) The following datasets generated for this study and available from the Dryad Digital Repository: http://dx.doi.org/xxxxx

- VCF file of no filtered and filtered SNPs dataset
- Matrix of normalized read depth for CNV loci
- Estimated statistics for SNPs duplicated effect detection
- Series of R scripts (environment data reduction, CNVs identification, read-depth normalization, RDA...)

Authors' contribution:

L.B and R.R designed and supervised the project. Y.D conducted literature mining and laboratory work. Bioinformatics and data analyses were conducted by Y.D with input from E.N., M.L. H.C. K.W. and Q.R. Original draft preparation was lead by Y.D. and all authors contributed to the writing-revew and editing of the final version of the manuscript.

Competing interest:

None declared

Acknowledgements

This research was financially supported a Strategic Partnership Grants for Projects from the Natural Sciences and Engineering Research Council of Canada to LB and RR. We thank scientists from the Department of Fisheries and Oceans and Canadian fishers who helped collecting the samples. We are grateful to Alison Devault and the Arbor Biosciences team for DNA probes synthesis and methodological advices. We also thank the personal of the IBIS sequencing platform for their assistance in developing the Rapture assay for highly multiplexed configuration.

References:

Al-Breiki, R. D., Kjeldsen, S. R., Afzal, H., Al Hinai, M. S., Zenger, K. R., Jerry, D. R., Al-Abri M., Delghandi, M. (2018). Genome-wide SNP analyses reveal high gene flow and signatures of local adaptation among the scalloped spiny lobster (Panulirus homarus) along the Omani coastline. *BMC Genomics*, *19*(1). doi: 10.1186/s12864-018-5044-8

Ali, O. A., ORourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics*, *202*(2), 389–400. doi: 10.1534/genetics.115.183665

Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, *25*(22), 4692–4693. doi: 10.1093/nar/25.22.4692

Aubin-Horth, N., & Renn, S. C. P. (2009). Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. *Molecular Ecology*, *18*(18), 3763–3780. doi: 10.1111/j.1365-294X.2009.04313.x

Barajas-Olmos, F. M., Ortiz-Sánchez, E., Imaz-Rosshandler, I., Córdova-Alarcón, E. J., Martínez-Tovar, A., Villanueva-Toledo, J., … Centeno, F. (2019). Analysis of the dynamic aberrant landscape of DNA methylation and gene expression during arsenic-induced cell transformation. *Gene*, *711*, 143941. doi: 10.1016/j.gene.2019.143941

Barth, J. M. I., Villegas-Ríos, D., Freitas, C., Moland, E., Star, B., André, C., … Jentoft, S. (2019). Disentangling structural genomic and behavioural barriers in a sea of connectivity. *Molecular Ecology*, *28*(6), 1394–1411. doi: 10.1111/mec.15010

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bazzicalupo, A. L., Ruytinx, J., Ke, Y.-H., Coninx, L., Colpaert, J. V., Nguyen, N. H., … Branco, S. (2019). Incipient local adaptation in a fungus: evolution of heavy metal tolerance through allelic and copy-number variation. *BioRxiv*, 832089. doi: 10.1101/832089

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., … Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). *Molecular Ecology*, *25*(20), 5073–5092. doi: 10.1111/mec.13811

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*(13), 3299–3315. doi: 10.1111/mec.13245

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., … Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports*, *6*(1), 1–12. doi: 10.1038/srep23246

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, *89*(6), 2519–2556. doi: 10.1111/jfb.13145

Bernatchez, S., Xuereb, A., Laporte, M., Benestan, L., Steeves, R., Laflamme, M., … Mallet, M. A. (2019). Seascape genomics of eastern oyster (Crassostrea virginica) along the Atlantic coast of Canada. *Evolutionary Applications*, *12*(3), 587–609. doi: 10.1111/eva.12741

Biémont, C. (2008). Genome size evolution: within-species variation in genome size. *Heredity*, *101*(4), 297–298. doi: 10.1038/hdy.2008.80

Carr, M. H., Neigel, J. E., Estes, J. A., Andelman, S., Warner, R. R., & Largier, J. L. (2003). Comparing Marine and Terrestrial Ecosystems: Implications for the Design of Coastal Marine Reserves. *Ecological Applications*, *13*(sp1), 90–107. doi:10.1890/1051-0761(2003)013[0090:CMATEI]2.0.CO;2

Catanach, A., Crowhurst, R., Deng, C., David, C., Bernatchez, L., & Wellenreuther, M. (2019). The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by

threefold in the marine teleost Chrysophrys auratus. *Molecular Ecology*, *28*(6), 1210–1223. doi: 10.1111/mec.15051

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. doi: 10.1111/mec.12354

Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msz008

Chassé, J., & Miller, R. J. (2010). Lobster larval transport in the southern Gulf of St. Lawrence. *Fisheries Oceanography*, *19*(5), 319–338. doi: 10.1111/j.1365-2419.2010.00548.x

Chen, Z., Cheng, C.-H. C., Zhang, J., Cao, L., Chen, L., Zhou, L., … Chen, L. (2008). Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences*, *105*(35), 12944–12949. doi: 10.1073/pnas.0802432105

Cioffi, M. B., & Bertollo, L. A. C. (2012). Chromosomal Distribution and Evolution of Repetitive DNAs in Fish. In M. A. Garrido-Ramos (Ed.), *Genome Dynamics* (Vol. 7, pp. 197–221). doi: 10.1159/000337950

Clop, A., Vidal, O., & Amills, M. (2012). Copy number variation in the genomes of domestic animals. *Animal Genetics*, *43*(5), 503–517. doi: 10.1111/j.1365-2052.2012.02317.x

Cockerham, C. C., & Weir, B. S. (1984). Covariances of Relatives Stemming from a Population Undergoing Mixed Self and Random Mating. *Biometrics*, *40*(1), 157–164. doi: 10.2307/2530754

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., … Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, *464*(7289), 704–712. doi: 10.1038/nature08516

Côté, C. L., Pavey, S. A., Stacey, J. A., Pratt, T. C., Castonguay, M., Audet, C., & Bernatchez, L. (2015). Growth, Female Size, and Sex Ratio Variability in American Eel of Different Origins in Both Controlled Conditions and the Wild: Implications for Stocking Programs. *Transactions of the American Fisheries Society*, *144*(2), 246–257. doi: 10.1080/00028487.2014.975841

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499–510. doi: 10.1038/nrg3012

Dennis, M. Y., Harshman, L., Nelson, B. J., Penn, O., Cantsilieris, S., Huddleston, J., … Eichler, E. E. (2017). The evolution and population diversity of human-specific segmental duplications. *Nature Ecology & Evolution*, *1*(3), 1–10. doi: 10.1038/s41559-016-0069

Desjardins, M., Graham, L. A., Davies, P. L., & Fletcher, G. L. (2012). Antifreeze protein gene amplification facilitated niche exploitation and speciation in wolffish. *The FEBS Journal*, *279*(12), 2215–2230. doi: 10.1111/j.1742-4658.2012.08605.x

DiBattista, J. D., Travers, M. J., Moore, G. I., Evans, R. D., Newman, S. J., Feng, M., … Berry, O. (2017). Seascape genomics reveals fine-scale patterns of dispersal for a reef fish along the ecologically divergent coast of Northwestern Australia. *Molecular Ecology*, *26*(22), 6206–6223. doi: 10.1111/mec.14352

Dorant, Y., Benestan, L., Rougemont, Q., Normandeau, E., Boyle, B., Rochette, R., & Bernatchez, L. (2019). Comparing Pool-seq, Rapture, and GBS genotyping for inferring weak population structure: The American lobster (*Homarus americanus*) as a case study. *Ecology and Evolution*. doi: 10.1002/ece3.5240

Dunaway, K. W., Islam, M. S., Coulson, R. L., Lopez, S. J., Vogel Ciernia, A., Chu, R. G., … LaSalle, J. M. (2016). Cumulative Impact of Polychlorinated Biphenyl and Large Chromosomal Duplications on DNA Methylation, Chromatin, and Expression of Autism Candidate Genes. *Cell Reports*, *17*(11), 3035–3048. doi: 10.1016/j.celrep.2016.11.058

Emerson, J. J., Cardoso-Moreira, M., Borevitz, J. O., & Long, M. (2008). Natural Selection Shapes Genome-Wide Patterns of Copy-Number Polymorphism in Drosophila melanogaster. *Science*, *320*(5883), 1629–1631. doi: 10.1126/science.1158078

Fang, S., Wu, R., Shi, X., Zhang, Y., Ikhwanuddin, M., Lu, J., … Ma, H. (2019). Genome survey and identification of polymorphic microsatellites provide genomic information and molecular markers for the red crab Charybdis feriatus (Linnaeus, 1758) (Decapoda: Brachyura: Portunidae). *Journal of Crustacean Biology*. doi: 10.1093/jcbiol/ruz074

Faucon, F., Gaude, T., Dusfour, I., Navratil, V., Corbel, V., Juntarajumnong, W., … David, J.-P. (2017). In the hunt for genomic markers of metabolic resistance to pyrethroids in the mosquito Aedes aegypti: An integrated next-generation sequencing approach. *PLOS Neglected Tropical Diseases*, *11*(4), e0005526. doi: 10.1371/journal.pntd.0005526

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics*, *7*(2), 85–97. doi: 10.1038/nrg1767

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*(9), 2215–2233. doi: 10.1111/mec.14584

Fraley, C., & Raftery, A. (2007). Model-based Methods of Classification: Using the mclust Software in Chemometrics. *Journal of Statistical Software*, *18*(6). doi: 10.18637/jss.v018.i06

Fraley, C., & Raftery, A. E. (2012). *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*. 57.

Freyhult, E., Landfors, M., Önskog, J., Hvidsten, T. R., & Rydén, P. (2010). Challenges in microarray class discovery: a comprehensive examination of normalization, gene selection and clustering. *BMC Bioinformatics*, *11*(1), 503. doi: 10.1186/1471-2105-11-503

Galbraith, P. S., Chassé, J., Nicot, P., Caverhill, C., Gilbert, D., Pettigrew, B., … Lafleur, C. (2015). *Physical Oceanographic Conditions in the Gulf of St. Lawrence in 2014*. Canadian Science Advisory Secretariat.

Gamazon, E. R., & Stranger, B. E. (2015). The impact of human copy number variation on gene expression. *Briefings in Functional Genomics*, *14*(5), 352–357. doi: 10.1093/bfgp/elv017

Grummer, J. A., Beheregaray, L. B., Bernatchez, L., Hand, B. K., Luikart, G., Narum, S. R., & Taylor, E. B. (2019). Aquatic Landscape Genomics and Environmental Effects on Genetic Variation. *Trends in Ecology & Evolution*, *34*(7), 641–654. doi: 10.1016/j.tree.2019.02.013

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. doi: 10.1186/s12915-015-0130-8

Hayes, P. H., Davies, P. L., & Fletcher, G. L. (1991). Population differences in antifreeze protein gene copy number and arrangement in winter flounder. *Genome*, *34*(1), 174–177. doi: 10.1139/g91-027

Hemmer-Hansen, J., Therkildsen, N. O., Meldrup, D., & Nielsen, E. E. (2014). Conserving marine biodiversity: insights from life-history trait candidate genes in Atlantic cod (Gadus morhua). *Conservation Genetics*, *15*(1), 213–228. doi: 10.1007/s10592-013-0532-5

Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F., & Narum, S. R. (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, *22*(11), 2898–2916. doi: 10.1111/mec.12150

Hew, C. L., Wang, N. C., Joshi, S., Fletcher, G. L., Scott, G. K., Hayes, P. H., … Davies, P. L. (1988). Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout, Macrozoarces americanus. *Journal of Biological Chemistry*, *263*(24), 12049–12055.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, *188*(4), 379–397. doi: 10.1086/688018

Huang, X., Li, S., Ni, P., Gao, Y., Jiang, B., Zhou, Z., & Zhan, A. (2017). Rapid response to changing environments during biological invasions: DNA methylation perspectives. *Molecular Ecology*, *26*(23), 6621–6633. doi: 10.1111/mec.14382

Huang, Y., Feulner, P. G. D., Eizaguirre, C., Lenz, T. L., Bornberg-Bauer, E., Milinski, M., … Chain, F. J. J. (2019). Genome-Wide Genotype-Expression Relationships Reveal Both Copy Number and Single Nucleotide Differentiation Contribute to Differential Gene Expression between Stickleback Ecotypes. *Genome Biology and Evolution*, *11*(8), 2344–2359. doi: 10.1093/gbe/evz148

Ionita-Laza, I., Rogers, A. J., Lange, C., Raby, B. A., & Lee, C. (2009). Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics*, *93*(1), 22–26. doi: 10.1016/j.ygeno.2008.08.012

Jimenez, A. G., Kinsey, S. T., Dillaman, R. M., & Kapraun, D. F. (2010). Nuclear DNA content variation associated with muscle fiber hypertrophic growth in decapod crustaceans. *Genome*, *53*(3), 161–171. doi: 10.1139/G09-095

Junge, C., Donnellan, S. C., Huveneers, C., Bradshaw, C. J. A., Simon, A., Drew, M., … Gillanders, B. M. (2019). Comparative population genomics confirms little population structure in two commercially targeted carcharhinid sharks. *Marine Biology*, *166*(2). doi: 10.1007/s00227-018-3454-4

Kess, T., Bentzen, P., Lehnert, S. J., Sylvester, E. V. A., Lien, S., Kent, M. P., … Bradbury, I. R. (2020). Modular chromosome rearrangements reveal parallel and nonparallel adaptation in a marine fish. *Ecology and Evolution*, *n/a*(n/a). doi: 10.1002/ece3.5828

Kofler, R., Nolte, V., & Schlötterer, C. (2015). Tempo and Mode of Transposable Element Activity in Drosophila. *PLoS Genetics*, *11*(7). doi: 10.1371/journal.pgen.1005406

Laporte, M., Claude, J., Berrebi, P., Perret, P., & Magnan, P. (2016). Shape plasticity in response to water velocity in the freshwater blenny Salaria fluviatilis. *Journal of Fish Biology*, *88*(3), 1191–1203. doi: 10.1111/jfb.12902

Larouche, P., & Galbraith, P. S. (2016). Canadian Coastal Seas and Great Lakes Sea Surface Temperature Climatology and Recent Trends. *Canadian Journal of Remote Sensing*, *42*(3), 243–258. doi: 10.1080/07038992.2016.1166041

Legendre, P., & Legendre, L. F. J. (2012). *Numerical Ecology* (3 edition). Amsterdam: Elsevier.

Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., … Venter, J. C. (2007). The Diploid Genome Sequence of an Individual Human. *PLOS Biology*, *5*(10), e254. doi: 10.1371/journal.pbio.0050254

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. Retrieved from http://arxiv.org/abs/1303.3997

López-Flores, I., & Garrido-Ramos, M. A. (2012). The Repetitive DNA Content of Eukaryotic Genomes. *Repetitive DNA*, *7*, 1–28. doi: 10.1159/000337118

Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: the long and the short of it. *Genome Biology*, *20*(1). doi: 10.1186/s13059-019-1828-7

Makino, T., McLysaght, A., & Kawata, M. (2013). Genome-wide deserts for copy number variation in vertebrates. *Nature Communications*, *4*(1), 1–10. doi: 10.1038/ncomms3283

Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution*, *28*(10), 614–621. doi: 10.1016/j.tree.2013.05.012

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. doi: 10.14806/ej.17.1.200

McCue, A. D., Nuthikattu, S., Reeder, S. H., & Slotkin, R. K. (2012). Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. *PLOS Genetics*, *8*(2), e1002474. doi: 10.1371/journal.pgen.1002474

McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*(4), 656–669. doi: 10.1111/1755-0998.12613

Medvedovic, M., Yeung, K. Y., & Bumgarner, R. E. (2004). Bayesian mixture model based clustering of replicated microarray data. *Bioinformatics (Oxford, England)*, *20*(8), 1222–1232. doi: 10.1093/bioinformatics/bth068

Morse, B. L., Quinn, B. K., Comeau, M., & Rochette, R. (2018). Stock structure and connectivity of the American lobster Homarus americanus in the southern Gulf of St. Lawrence: Do benthic movements matter? *Canadian Journal of Fisheries and Aquatic Sciences*, *75*(11), 2096–2108. doi: 10.1139/cjfas-2017-0346

Nelson, T. C., Monnahan, P. J., McIntosh, M. K., Anderson, K., MacArthur-Waltz, E., Finseth, F. R., … Fishman, L. (2018). Extreme copy number variation at a tRNA ligase gene affecting phenology and fitness in yellow monkeyflowers. *Molecular Ecology*. doi: 10.1111/mec.14904

Nguyen, D.-Q., Webber, C., & Ponting, C. P. (2006). Bias of Selection on Human Copy-Number Variants. *PLOS Genetics*, *2*(2), e20. doi: 10.1371/journal.pgen.0020020

Oksanen, J., Blanchet F.G., Friendly M., Kindt R., Legendre P., McGlinn D., …, Wagner H. (2018). Vegan: Community ecology package. R package version 2.5.5. Retried from http://r-forge.r-project.org/projects/vegan

Palumbi, S., Tyler, E., Pespeni, M., & Somero, G. (2019). Present and Future Adaptation of Marine Species Assemblages: DNA-Based Insights into Climate Change from Studies of Physiology, Genomics, and Evolution. *Oceanography*, *32*(3), 82–93. doi: 10.5670/oceanog.2019.314

Pang, A. W., MacDonald, J. R., Pinto, D., Wei, J., Rafiq, M. A., Conrad, D. F., … Scherer, S. W. (2010). Towards a comprehensive structural variation map of an individual human genome. *Genome Biology*, *11*(5), R52. doi: 10.1186/gb-2010-11-5-r52

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952. doi: 10.1111/1755-0998.12129

Prokopowich, C. D., Gregory, T. R., & Crease, T. J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, *46*(1), 48–50. doi: 10.1139/g02-103

Prunier, J., Giguère, I., Ryan, N., Guy, R., Soolanayakanahally, R., Isabel, N., … Porth, I. (2019). Gene copy number variations involved in balsam poplar (Populus balsamifera L.) adaptive variations. *Molecular Ecology*, *28*(6), 1476–1490. doi: 10.1111/mec.14836

Quinn, B. K. (2017). Threshold temperatures for performance and survival of American lobster larvae: A review of current knowledge and implications to modeling impacts of climate change. *Fisheries Research*, *186*, 383–396. doi: 10.1016/j.fishres.2016.09.022

Quinn, B. K., Chassé, J., & Rochette, R. (2017). Potential connectivity among American lobster fisheries as a result of larval drift across the species' range in eastern North America. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(10), 1549–1563. doi: 10.1139/cjfas-2016-0416

Quinn, B. K., & 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*(suppl 1), i79–i90. doi: 10.1093/icesjms/fsv010

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., … Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444–454. doi: 10.1038/nature05329

Rey, O., Danchin, E., Mirouze, M., Loot, C., & Blanchet, S. (2016). Adaptation to Global Change: A Transposable Element–Epigenetics Perspective. *Trends in Ecology & Evolution*, *31*(7), 514–526. doi: 10.1016/j.tree.2016.03.013

Rice, A. M., & McLysaght, A. (2017). Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nature Communications*, *8*(1), 1–11. doi: 10.1038/ncomms14366

Rinker, D. C., Specian, N. K., Zhao, S., & Gibbons, J. G. (2019). Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. *Proceedings of the National Academy of Sciences*, *116*(27), 13446–13451. doi: 10.1073/pnas.1901093116

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-Response Analysis Using R. *PLOS ONE*, *10*(12), e0146021. doi: 10.1371/journal.pone.0146021

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. doi: 10.1186/gb-2010-11-3-r25

Roca, I., González-Castro, L., Fernández, H., Couce, M. L., & Fernández-Marmiesse, A. (2019). Free-access copy-number variant detection tools for targeted next-generation sequencing data. *Mutation Research/Reviews in Mutation Research*, *779*, 114–125. doi: 10.1016/j.mrrev.2019.02.005

Rougeux, C., Gagnaire, P.-A., Praebel, K., Seehausen, O., & Bernatchez, L. (2019). Polygenic selection drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic sister species complex. *Molecular Ecology*, *28*(19), 4388–4403. doi: 10.1111/mec.15226

Rowan, B. A., Heavens, D., Feuerborn, T. R., Tock, A. J., Henderson, I. R., & Weigel, D. (2019). An Ultra High-Density Arabidopsis thaliana Crossover Map That Refines the Influences of Structural Variation and Epigenetic Features. *Genetics*, *213*(3), 771–787. doi: 10.1534/genetics.119.302406

Ruytinx, J., Coninx, L., Op De Beeck, M., Arnauts, N., Rineau, F., & Colpaert, J. V. (2019). Adaptive zinc tolerance is supported by extensive gene multiplication and differences in cis-regulation of a CDF transporter in an ectomycorrhizal fungus. *BioRxiv*. doi: 10.1101/817676

Sandoval-Castillo, J., Robinson, N. A., Hart, A. M., Strain, L. W. S., & Beheregaray, L. B. (2018). Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (Haliotis laevigata), along a longitudinal environmental gradient. *Molecular Ecology*, *27*(7), 1603–1620. doi: 10.1111/mec.14526

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820. doi: 10.1038/nrg3522

Sbrocco, E. J., & Barber, P. H. (2013). MARSPEC: ocean climate layers for marine spatial ecology. *Ecology*, *94*(4), 979–979. doi: 10.1890/12-1358.1

Schlichting, C. D., & Pigliucci, M. (1998). Phenotypic evolution: a reaction norm perspective. *Phenotypic Evolution: A Reaction Norm Perspective.* Retrieved from https://www.cabdirect.org/cabdirect/abstract/19980108896

Selkoe, K. A., Henzler, C. M., & Gaines, S. D. (2008). Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, *9*(4), 363–377. doi: 10.1111/j.1467-2979.2008.00300.x

Serrato-Capuchina, A., & Matute, D. R. (2018). The Role of Transposable Elements in Speciation. *Genes*, *9*(5), 254. doi: 10.3390/genes9050254

Smit, A. J., Roberts, M., Anderson, R. J., Dufois, F., Dudley, S. F. J., Bornman, T. G., … Bolton, J. J. (2013). A Coastal Seawater Temperature Dataset for Biogeographical Studies: Large Biases between In Situ and Remotely-Sensed Data Sets around the Coast of South Africa. *PLOS ONE*, *8*(12), e81944. doi: 10.1371/journal.pone.0081944

Song, L., Bian, C., Luo, Y., Wang, L., You, X., Li, J., … Xu, P. (2016). Draft genome of the Chinese mitten crab, Eriocheir sinensis. *GigaScience*, *5*(1), 5. doi: 10.1186/s13742-016-0112-y

Stanley, R. R. E., DiBacco, C., Lowen, B., Beiko, R. G., Jeffery, N. W., Wyngaarden, M. V., … Bradbury, I. R. (2018). A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Advances*, *4*(3), eaaq0929. doi: 10.1126/sciadv.aaq0929

Sudmant, P. H., Mallick, S., Nelson, B. J., Hormozdiari, F., Krumm, N., Huddleston, J., … Eichler, E. E. (2015). Global diversity, population stratification, and selection of human copy-number variation. *Science*, *349*(6253), aab3761. doi: 10.1126/science.aab3761

Sunday, J. M., Bates, A. E., & Dulvy, N. K. (2011). Global analysis of thermal tolerance and latitude in ectotherms. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1713), 1823–1830. doi: 10.1098/rspb.2010.1295

Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S., & Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics*, *28*(21), 2711–2718. doi: 10.1093/bioinformatics/bts535

Tigano, A., Reiertsen, T. K., Walters, J. R., & Friesen, V. L. (2018). A complex copy number variant underlies differences in both colour plumage and cold adaptation in a dimorphic seabird. *BioRxiv*. Doi: 10.1101/507384

Verbruggen, B. (2016). *Generating genomic resources for two crustacean species and their application to the study of White Spot Disease*. https://ore.exeter.ac.uk/repository/handle/10871/25535

Weetman, D., Djogbenou, L. S., & Lucas, E. (2018). Copy number variation (CNV) and insecticide resistance in mosquitoes: evolving knowledge or an evolving problem? *Current Opinion in Insect Science*, *27*, 82–88. doi: 10.1016/j.cois.2018.04.005

Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends in Ecology & Evolution*, *33*(6), 427–440. doi: 10.1016/j.tree.2018.04.002

Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular Ecology*, *28*(6), 1203–1209. doi: 10.1111/mec.15066

Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.

Wong, K. K., deLeeuw, R. J., Dosanjh, N. S., Kimm, L. R., Cheng, Z., Horsman, D. E., … Lam, W. L. (2007). A Comprehensive Analysis of Common Copy-Number Variations in the Human Genome. *The American Journal of Human Genetics*, *80*(1), 91–104. doi: 10.1086/510560

Wright, S. (1949). The Genetical Structure of Populations. *Annals of Eugenics*, *15*(1), 323–354. doi: 10.1111/j.1469-1809.1949.tb02451.x

Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M.-J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (Parastichopus californicus). *Molecular Ecology*, *27*(10), 2347–2364. doi: 10.1111/mec.14589

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., … Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569. doi: 10.1038/ng.608

Yuan, J., Zhang, X., Liu, C., Yu, Y., Wei, J., Li, F., & Xiang, J. (2018). Genomic resources and comparative analyses of two economical penaeid shrimp species, Marsupenaeus japonicus and Penaeus monodon. *Marine Genomics*, *39*, 22–25. doi: 10.1016/j.margen.2017.12.006

Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y., Yu, Y., … Xiang, J. (2019). Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications*, *10*(1), 356. doi: 10.1038/s41467-018-08197-4

Supporting information

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

Table S1: Geographic coordinates, sample size (i.e. number of samples successfully genotyped), sequencing effort, environmental parameters across the 21 sampling locations.

Table S2: Filtering sequencing data - setup pipeline

Table S3: SNPs classification results

Figure S1: Correlation plots of temperature parameters

Figure S2: Characterization of duplication effect over SNPs dataset

Figure S3: Impact of read depth normalization for CNVs data

Figure S4: Relationship between SST annual variance and CNVs read depth

Figure S5: Model comparison

Figure S6: Discretization of copy groups for 48 CNVs associated with SST annual variance

Figure S7:Mclust model performance among the 48 CNVs locus associated with SST annual variance.

Figure S8: Mclust model features for the 48 CNV loci associated with SST annual variance.