Received: 6 April 2020 | Revised: 15 January 2021 | Accepted: 1 February 2021

DOI: 10.1111/mec.15835

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Thermal adaptation rather than demographic history drives genetic structure inferred by copy number variants in a marine fish

Hugo Cayuela[1,2] | Yann Dorant[1] | Claire Mérot[1] | Martin Laporte[1] | Eric Normandeau[1] | Stéphane Gagnon-Harvey[3] | Marie Clément[4,5] | Pascal Sirois[3] | Louis Bernatchez[1]

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

[2]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

[3]Département des sciences fondamentales, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

[4]Center for Fisheries Ecosystems Research, Fisheries and Marine Institute of Memorial, University of Newfoundland, St. John's, NL, Canada

[5]Labrador Institute of Memorial University of Newfoundland, Happy Valley-Goose Bay, NL, Canada

Correspondence
Hugo Cayuela, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada.
Email: hugo.cayuela51@gmail.com

Funding information
Natural Sciences and Engineering Research Council of Canada; Ressources Aquatiques Québec; Department of Fisheries and Oceans Canada; Nunatsiavut Government, NunatuKavut Community Council; Labrador Fishermen's Union Shrimp Company; Department of Fisheries and Aquaculture – Government of Newfoundland and Labrador; World Wildlife Fund Canada; St. Lawrence Global Observatory; Parc Marin du Saguenay–Saint-Laurent; Greenland Institute of Natural Resources; Swiss National Science Foundation, Grant/Award Number: 31003A_182265

Abstract

Increasing evidence shows that structural variants represent an overlooked aspect of genetic variation with consequential evolutionary roles. Among those, copy number variants (CNVs), including duplicated genomic regions and transposable elements (TEs), may contribute to local adaptation and/or reproductive isolation among divergent populations. Those mechanisms suppose that CNVs could be used to infer neutral and/or adaptive population genetic structure, whose study has been restricted to microsatellites, mitochondrial DNA and Amplified fragment length polymorphism markers in the past and more recently the use of single nucleotide polymorphisms (SNPs). Taking advantage of recent developments allowing CNV analysis from RAD-seq data, we investigated how variation in fitness-related traits, local environmental conditions and demographic history are associated with CNVs, and how subsequent copy number variation drives population genetic structure in a marine fish, the capelin (*Mallotus villosus*). We collected 1538 DNA samples from 35 sampling sites in the north Atlantic Ocean and identified 6620 putative CNVs. We found associations between CNVs and the gonadosomatic index, suggesting that six duplicated regions could affect female fitness by modulating oocyte production. We also detected 105 CNV candidates associated with water temperature, among which 20% corresponded to genomic regions located within the sequence of protein-coding genes, suggesting local adaptation to cold water by means of gene sequence amplification. We also identified 175 CNVs associated with the divergence of three previously defined parapatric glacial lineages, of which 24% were located within protein-coding genes, making those loci potential candidates for reproductive isolation. Lastly, our analyses unveiled a hierarchical, complex CNV population structure determined by temperature and local geography, which was in stark contrast to that inferred based on SNPs in a previous study. Our findings underline the complementarity of those two types of genomic variation in population genomics studies.

KEYWORDS
copy number variants, fish, local adaptation, population genetic structure, reproductive isolation, structural variants, transposable elements

1 | INTRODUCTION

Genetic variation is an essential component of evolution, contributing to adaptation, reproductive isolation and genetic structure. A part of this genetic variation, called single-nucleotide polymorphism (SNP), has been well characterized and its evolutionary role well demonstrated over the last couple of decades (Helyar et al., 2011; Leaché & Oaks, 2017; Morin et al., 2004). However, there is increasing evidence that structural variants (SVs) represent an overlooked aspect of genetic variation with a possible important evolutionary role (Chain & Feulner, 2014; Spielmann et al., 2018; Wellenreuther & Bernatchez, 2018; Wellenreuther et al., 2019). A generalized conceptual framework is emerging (Mérot et al., 2020; Wellenreuther et al., 2019), which proposes to study those SVs and their diversity using integrative approaches toward deciphering their respective mechanisms and the consequences of their interactions for micro- and macro-evolutionary processes.

Among the different types of SVs, copy number variants (CNVs) have been proposed to play a crucial role in genome evolution, adaptation and speciation (Freeman et al., 2006; Zhang et al., 2009). CNVs are genomic SVs in which a segment of DNA can be absent (in comparison to the sequence of reference) or present in two or more copies due to either gene duplication, deletion or transposable elements (TEs) (Mérot et al., 2020). CNVs can arise from a variety of mechanisms, including nonallelic homologous recombination, nonhomologous end-joining and retrotransposition (Hastings et al., 2009). They may be maintained in a population via neutral evolutionary processes (i.e., migration, genetic drift) and either positive, purifying or balancing selection (Katju & Bergthorsson, 2013; Qian & Zhang, 2014; Zhang et al., 2009).

As copy-number changes may reach high frequency and can be fixed in a short time (Farslow et al., 2015), CNVs have been hypothesized to promote local adaptation and enable the colonization of new habitats (Kondrashov, 2012; Qian & Zhang, 2014). By affecting gene dosage and expression, gene duplication may determine phenotypic performance and fitness under specific environmental conditions (Kondrashov, 2012; Qian & Zhang, 2014). For instance, studies have reported that gene sequence amplification may underlie adaptive responses to both abiotic (e.g., insecticide, Raymond et al., 1991; heavy metals, Hull et al., 2017; temperature, Tigano et al., 2018) and biotic factors (e.g., nutrient limitation, Kondrashov et al., 2002). Furthermore, the activity of TEs, a mechanism generating CNVs, can also promote local adaptation (van't Hof et al., 2016; Casacuberta & González, 2013; Schrader et al., 2014; Stapley et al., 2015). TEs generate a broad variety of mutations that may lead to phenotypic and fitness variation through the modification of gene expression, the inactivation of genes, and the alteration of gene sequence and reading frame (Chuong et al., 2017). The activation of TEs in response to stress experienced during an individual's lifetime induces structural variation that may help organisms to adapt to environmental conditions (McClintock, 1950) such as temperature and rainfall (e.g., González et al., 2010).

CNVs may also play a major role in the process of speciation (Lynch & Force, 2000; Serrato-Capuchina & Matute, 2018). As their evolutionary rate is higher than that of SNPs (Paudel et al., 2015; Sudmant et al., 2013), CNVs enhance the accumulation of genetic incompatibilities (Nosil et al., 2009), and thus accelerate postzygotic reproductive isolation and speciation rate (Böhne et al., 2008; Laporte et al., 2019; Ricci et al., 2018). TEs may be also involved in reproductive barriers by causing disruptions of gene expression (Dion-Côté et al., 2014), genomic expansions and generating new chromosomal inversions (Serrato-Capuchina & Matute, 2018).

Clearly, the central role of CNVs in adaptation and reproductive isolation demonstrated in previous studies suggests that they could play a major role in modulating patterns of population structure, whose study is increasingly focused on the use of SNPs (Helyar et al., 2011; Hendricks et al., 2018). Thus, short tandem repeats (i.e., microsatellites), a specific family of CNVs, have been used extensively to study population genetic structure over the last three decades (Li et al., 2002; Schlötterer & Pemberton, 1998). Given their faster evolution rate (Paudel et al., 2015; Redon et al., 2006; Sudmant et al., 2013) and the effect of the environment on their accumulation within the genome (for TEs, see Chuong et al., 2017), CNVs could reveal patterns of genetic structure that are different from those drawn by SNPs. Consequently, they would provide a valuable and complementary set of genetic markers to analyse neutral and adaptive population structure in both basic and applied studies. However, with the exception of targeted candidate genes studies, CNV genotyping remained expensive until recently as it required whole genome sequencing data with deep coverage (Makałowski et al., 2019; Pirooznia et al., 2015), precluding the use of these markers in population genomic studies that involve thousands of samples from dozens of populations. This situation has recently changed through the development of a novel methodological approach allowing the identification of CNVs in a cost-effective way using RAD-seq (restriction site associated DNA sequencing) data (Dorant et al., 2020; McKinney et al., 2017; Tigano, 2020). Using this method, Dorant et al. (2020) recently showed that CNVs can be more efficient than SNPs at revealing genotype–temperature associations in marine invertebrates, namely the American lobster. Yet, the usefulness of CNVs from RAD-seq data as valuable DNA markers for population genomic studies still needs to be more broadly exemplified (Tigano, 2020). In particular, it needs to be investigated whether these CNVs may be associated with variation in fitness-related phenotypic traits and can be informative for organisms with a small-size genome

containing far fewer repeated elements than that of the American lobster (~4.5 Gbp; Jimenez et al., 2010).

The goal of this study was to investigate how fitness, local thermal conditions and demographic history are associated with CNVs genotyped in multiple populations, and subsequently how copy number variation drives population genetic structure in a marine fish, the capelin (*Mallotus villosus*). This species is an excellent biological model to address this issue: first, capelin has a relatively small genome size (~700 Mbp, as usual in Osmeridae; Hardie & Hebert, 2003) that probably contains far fewer repeated elements than the genome of the American lobster studied by Dorant et al. (2020). Second, capelin has a complex demographic history; in the Northwest Atlantic, it comprises three ancient lineages (Arctic lineage ARC, Greenland lineage GRE and Northwest Atlantic lineage NWA) which diverged approximately from 1.8 (NWA-GRE) to 3.5 million years ago (ARC-NWA), display limited historical introgression, and exhibit no apparent contemporary admixture (Cayuela et al., 2020). Very low migration rates and the absence of admixture between lineages despite no obvious physical barriers suggest the existence of strong reproductive barriers and an ongoing speciation process (Cayuela et al., 2020), allowing the consideration of copy-number variation within and between nascent species. Third, within-lineage populations may reproduce in either demersal or beach-spawning sites (Christiansen et al., 2008). In the latter case, populations span large environmental gradients in which temperature seems to be an important selective driver affecting fitness components (i.e., survival and growth rate) at early stages (i.e., embryo and larvae; Frank & Leggett, 1981; Leggett et al., 1984) and probably later in life (Colbeck et al., 2011; Kenchington et al., 2015). In the NWA lineage, genotype–temperature associations based on SNPs suggest that beach-spawning individuals are locally adapted to thermal conditions prevailing in the intertidal zone (Cayuela et al., 2020).

We first used the method developed by McKinney et al. (2017) and refined by Dorant et al. (2020) to detect reliable CNVs. Second, we investigated how CNV-normalized read depth, a robust proxy of putative copy number (Dorant et al., 2020), correlates with the gonadosomatic index, a commonly used fitness proxy in fishes (Brewer et al., 2008; Ressel et al., 2020). Third, we examined the putative role of CNVs in thermal adaptation by assessing correlations between normalized read depth and water temperature in beach spawning sites. In particular, based on previous findings in Antarctic notothenioid fishes (Chen et al., 2008), we expected a higher number of gene copies in individuals from cold waters than in their counterparts from warmer waters. As recent studies suggested that long-term temperature changes determine TE abundance in teleost fish genomes (Carducci et al., 2019; Shao et al., 2019; Yuan et al., 2018), we also hypothesized that thermal stress could have led to temperature-dependent accumulation of TEs in the capelin genome. Fourth, we investigated the potential role of CNVs on the ongoing speciation process among the three capelin lineages by quantifying how the normalized read depth differs for each pair of lineages. Fifth, we examined how variation in copy number associated with thermal conditions and demographic history drive genetic structure in the study area. Lastly, we underlined the differences in the patterns of genetic structure drawn by CNVs and SNPs, discussed the evolutionary mechanisms driving those discrepancies, and emphasized the complementarity of both markers in the context of population genomic studies.

2 | MATERIALS AND METHODS

2.1 | Sampling area, phenotypic analyses

We sampled 1538 capelins from 35 spawning sites in the Northwest Atlantic, in Canadian and Greenland waters (Figure 1; and Table S1). Three sites were sampled in the ARC lineage, four sites in the GRE lineage and 28 sites in the NWA lineage. Within the NWA lineage, the sampling sites included 18 beach spawning sites, six demersal spawning sites and four offshore sites. In the whole data set, the median sample size was 46.5 (range: 19–50) individuals per site. The fish were collected and immediately frozen, and a piece of fin was preserved in RNAlater. The gonadosomatic index was measured in the laboratory on 843 individuals (302 females and 541 males) from the 18 beach spawning sites of the NWA lineage. Because male gonads were very small, resulting in large measurement error of the gonadosomatic index (data not shown), we focused our analyses on females only.

2.2 | DNA sequencing, genotyping and discovery of putative CNVs

Both DNA extractions and library preparations were performed following protocols fully described in Cayuela et al. (2020). Libraries were size-selected using a BluePippin prep (Sage Science), amplified by polymerase chain reaction (PCR) and sequenced on the Ion Proton P1v2 chip (single-end sequencing). Eighty-two individuals were sequenced per chip.

Barcodes were removed using CUTADAPT (Martin, 2011) and trimmed to 80 bp, allowing for an error rate of 0.2. They were then demultiplexed using the "process_radtags" module of STACKS version 1.48 (Catchen et al., 2013) and aligned to the capelin draft genome (Cayuela et al., 2020) assembly using BWA-MEM (Li, 2013) with default parameters. Next, aligned reads were processed with STACKS version 1.48 for SNP calling and genotyping. The "pstacks" module was used with a minimum depth of 3 and up to 3 mismatches were allowed in the catalogue creation. We then ran the "populations" module to produce a vcf file that was further filtered using python (https://github.com/enormandeau/stacks_workflow) and bash scripts. SNPs were kept if they displayed a read depth >4 and <70 (Dorant et al., 2020). Then, we kept SNPs present in at least 70% of individuals in each sampling location.

We identified putative CNVs using the *HDplot* approach proposed by McKinney et al. (2017). The duplicated loci detected by this method can be variant or invariant in copy number among the

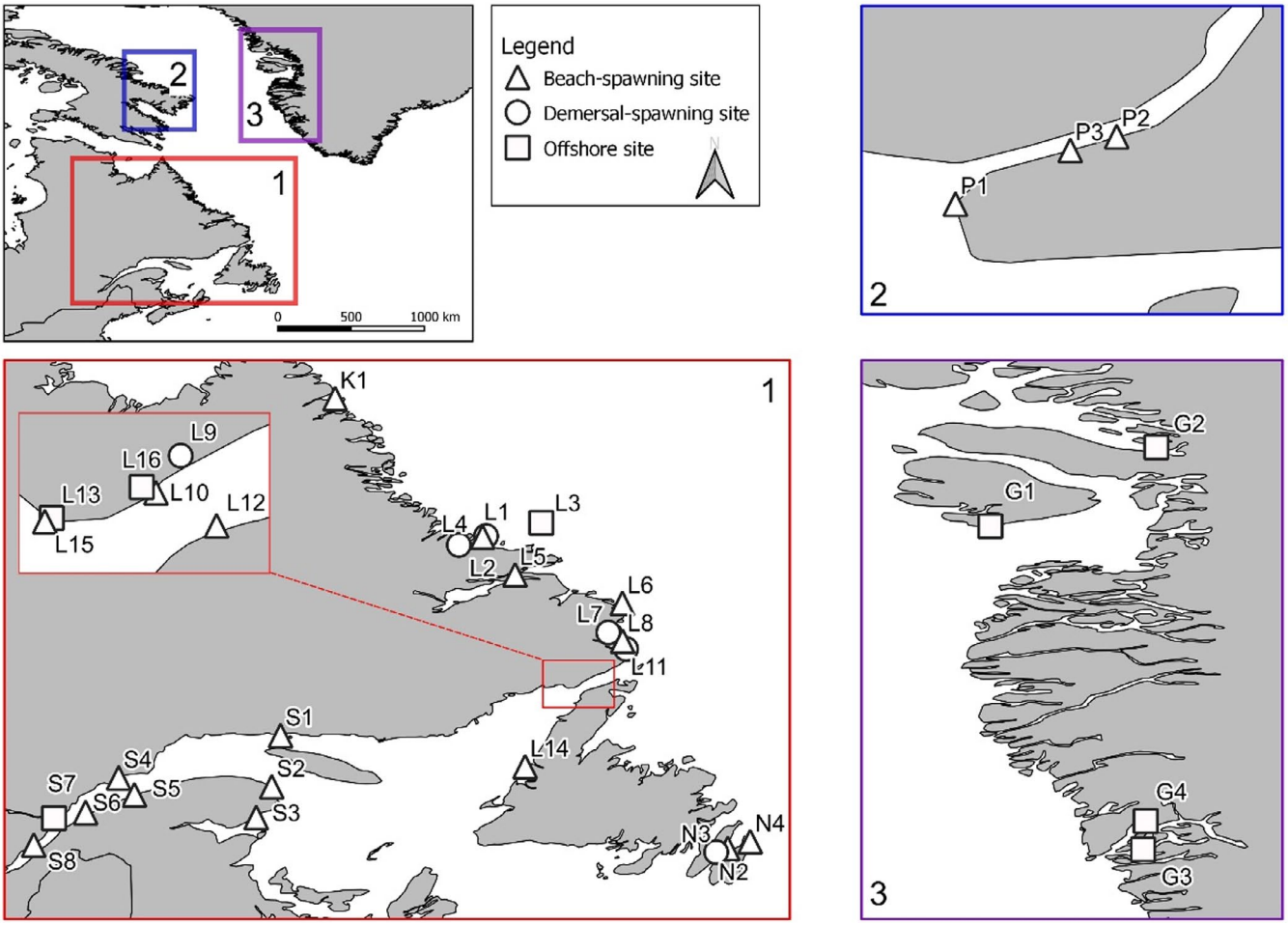

FIGURE 1 Map of the study area showing the sites sampled in the three capelin lineages. (1) Sampling sites in the NWA lineage. (2) Sampling sites in the ARC lineage. (3) Sampling sites in the GRE lineage. Three types of sites were sampled: beach-spawning sites (triangle), demersal-spawning sites (circle) and offshore sites (square) [Colour figure can be viewed at wileyonlinelibrary.com]

35 spawning sites of our study area. A duplicated locus is considered a CNV only if a correlation between the normalized read depth of a marker and a phenotypic trait or an environmental variable was detected. For this reason, we systematically use the term "putative CNVs" when we refer to the markers identified using the *HDplot* method. We used a refined version of this approach (Dorant et al., 2020) that discriminates "singleton" SNPs (i.e., nonduplicated) from "duplicated" SNPs (i.e., combining duplicated SNPs and SNPs with a high coverage) using five parameters: (i) proportion of heterozygotes (PropHet), (ii) F_{IS}, (iii) median of allele ratio for heterozygotes (MedRatio), (iv) median of SNP read depth for heterozygotes (MedDepthHet) and (v) median of SNP read depth for homozygotes (MedDepthHom). The parameters were calculated using a custom python script (available at https://github.com/enormandeau/stacks_workflow) parsing the filtered VCF file. The five parameters were plotted pairwise to visualize their distribution across all loci (see Figures S1 and S2). Based on the graphical demonstration by McKinney et al. (2017) and the approach proposed by Dorant et al. (2020), we considered different combinations of parameters and graphically fixed the cut-off of the four categories of SNPs (singleton SNPs, duplicated SNPs, high-coverage SNPs and low-confidence SNPs). We then kept one single marker per locus (the one with the highest minor allele frequency) and extracted the read depth of duplicated loci to construct the data set of putative CNVs using vcftools. For more simplicity, the term "putative CNV" is used to define all loci classified as "duplicated" and "high coverage". Following the procedure described in Dorant et al. (2020), read counts of putative CNV loci were normalized to account for differences in sequencing effort 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. This procedure was applied to the whole data set (i.e., the 1538 capelins from 35 spawning sites) to detect the putative CNVs present throughout the study area.

Using the capelin reference genome (Cayuela et al., 2020), we examined if the putative CNVs discovered were located within sequences of protein genes or within intergenic regions. Note that the genome is assembled at the scaffold level; nevertheless, contig order

was reconstructed via synteny analysis, which allowed the sorting of contigs into 24 orthologous chromosomes (Cayuela et al., 2020). For CNV candidates located in intergenic regions, we examined whether they corresponded to repeated elements including TEs. For that purpose, the reference genome was annotated for TEs (both DNA transposons and retrotransposons) and interspersed repeats (including satellites, simple repeats and low-complexity DNA sequences) using REPEATMASKER 4.1 (Smit et al., 2015) with the default options and the dfam database (Hubley et al., 2016) for the zebrafish (*Danio rerio*).

2.3 | Investigating CNV–fitness associations

To investigate the potential effect of CNVs on fitness-related traits, we examined how female gonadosomatic index correlates with normalized read depth of putative CNVs using a locus-by-locus GWAS-like approach. The gonadosomatic index, expressed as gonad mass as a percentage of total body mass, is widely used as a simple measure of the extent of reproductive investment and gonadal development (Brewer et al., 2008; Gunderson, 1997; Ressel et al., 2020). We used linear mixed models (LMMs) with restricted maximum likelihood optimization where the log-transformed gonadosomatic index was introduced as the response variable and the scaled normalized read depth was included as the explanatory variable. The spawning site was included as a random effect (i.e., random intercepts) to account for the nonindependency of observations across sites. One model was performed for each putative CNV locus. We used likelihood ratio tests (comparing the models with and without the explanatory term) to assess the significance of the CNV–fitness associations. The analysis was conducted in the R package lme4 (Bates et al., 2015). We considered a false discovery rate (FDR) of 0.10 and 0.05 to limit the risk of type I error when conducting multiple comparisons.

2.4 | Investigating CNV–temperature and CNV–lineage associations

To investigate the potential role of CNVs in local adaptation, we examined whether normalized read depth of putative CNVs correlates with sea water temperature, an environmental factor affecting growth and survival of embryos and larvae (Frank & Leggett, 1981; Leggett et al., 1984). Before examining CNV–temperature associations, we first verified that sea temperature was a major environmental factor affecting normalized read depth by quantifying the relative effect of temperature, salinity and chlorophyll α concentration—three potentially important selective factors for the capelin (Cayuela et al., 2020; Purchase, 2018)—on the read depth matrix. The analyses and the results are described in Supplementary material. The marine data layers for bottom annual temperature, salinity and chlorophyll concentration were downloaded from Bio-ORACLE (http://www.bio-oracle.org/) and the three variables at spawning sites were extracted using the R package sdmpredictors (Bosch et al., 2017).

After showing that temperature significantly explained the read depth matrix, and more importantly than the other two factors (Table S8), CNV–temperature associations were evaluated by combining locus-by-locus regressions and multivariate analyses. We ran LMMs with restricted maximum likelihood optimization where the log-transformed normalized read depth was incorporated as the response variable and temperature as the explanatory variable. The spawning site was included as a random effect (i.e., random intercepts) to account for the nonindependency of observations across sites. One model was performed for each putative CNV, and we used likelihood ratio tests and an FDR of 0.10 following the method of Benjamini and Hochberg (1995) to assess the significance of CNV–temperature associations.

We also used a partial redundancy analysis (pRDA, Legendre & Legendre, 2012) to detect multi-CNV candidates that were associated with temperature after controlling for the spawning sites, as commonly done for SNP data (Forester et al., 2018; Laporte et al., 2016; Le Luyer et al., 2017 for details). Global and marginal analyses of variance (ANOVAs) with 1000 permutations were performed to assess the significance of the models. Once CNVs were loaded against the pRDA axes, candidates for an association with temperature were determined as those exhibiting a loading >2.25 standard deviations from the mean loading ($p < .01$) (Forester et al., 2018). We retained the outliers that were detected by both LMMs and pRDA to reduce the number of false positives and represented the overlap with a Venn diagram.

We used a similar approach to explore the potential role of CNVs in the ongoing speciation process among the three aforementioned capelin lineages (Cayuela et al., 2020; Dodson et al., 2007). The procedure was performed separately for each pair of lineages. In LMMs, the lineage was introduced as a discrete explanatory variable with two modalities.

Next, for 20 CNV loci putatively associated with temperature (10 CNVs located in noncoding regions and 10 within protein coding genes; those with highest R^2 for each category), we examined whether discrete copy number categories could be drawn from the distributions of normalized read depth using a model-based unsupervised clustering approach implemented in the R library Mclust, version 5.4.4 (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 of read depth (that we term copy number groups) that probably reflect variation in copy number of any given locus.

2.5 | Inferring genetic structure based on CNVs

We documented population structure based on putative CNVs using a hierarchical clustering analysis (Langfelder & Horvath, 2012). We first calculated Bray–Curtis distance for each pair of individuals as commonly used in landscape genetic studies (Shirk et al., 2017). Then, we calculated the mean of the individual genetic distances for

TABLE 1 Candidate CNVs associated with gonadosomatic index

CNV ID	Contig	Chr	DNA repeated elements	Gene name	Gene Pfam	False discovery rate
50008_8	contig_5	6	None	XM_012828078.1	None	0.05
59193_13	contig_88	16	None	None	None	0.05
25048_60	contig_306	11	None	XM_012823774.1	PF00001; PF12369; PF13306; PF13855	0.05
48830_43	contig_5712	19	None	XM_012836008.1	PF04752	0.05
53320_53	contig_69	U	None	None	None	0.10
31892_56	contig_358	4	None	None	None	0.10

Note: The table shows the CNV name (CNV ID), the contig name and gene name from the Capelin genome assembly, the chromosome (Chr) name inferred using synteny analyses by Cayuela et al. (2020; U = unanchored contig), the type of repeated elements, and the gene Pfam. We retained protein-coding genes if the candidate CNV was within the gene sequence (i.e., XM_012828078.1) or within a 5-kbp window around the gene (XM_012836008.1). We also present the gene XM_012823774.1 (distance of ~20 kbp from the candidate 25048_60) because of its potential direct effect on the gonadosomatic index.

each pair of sites to obtain a square matrix of between-site genetic distances. With this distance matrix, we performed hierarchical clustering analyses using the R function *hclust*, based on Ward's minimum variance method (option *ward.D2*, Murtagh & Legendre, 2014). To evaluate the robustness of the dendrogram clusters, we performed 10,000 bootstraps using the R function *boot.phylo* implemented in the APE 5.3 package (Paradis et al., 2019). Dendrogram nodes were considered robust if their bootstrapping value was >0.80. Then, we examined how temperature of beach-spawning sites may affect the genetic structure inferred from putative CNVs within the NWA lineage. We performed two hierarchical clustering analyses: one based on all 6620 putative CNVs detected and the other with the 105 CNV candidates significantly associated with temperature (see Results). Then, we examined how lineage demographic divergence affects the genetic structure inferred from putative CNVs. To maintain a balanced number of sites in the three lineages (NWA, GRE and ARC), we randomly selected six spawning sites within the NWA lineage based on the clustering analysis performed for this lineage: three sites represented genetic cluster 1 (L2, L6 and L10) and three other sites represented cluster 2 (S1, S2 and S3) (see Results). Two hierarchical clustering analyses were conducted: one based on all 6620 putative CNVs and the other with the 175 CNV candidates associated with lineage divergence (see Results).

The clustering analysis for the 18 beach-spawning sites within the NWA lineage revealed two genetic clusters: cold sites from Labrador and the Atlantic coast of Newfoundland were grouped in cluster 1 whereas warm sites from the Gulf/Estuary of the St. Lawrence River were grouped in cluster 2 (see Results). Interestingly, spawning site S8 located in the St. Lawrence Estuary was assigned to cluster 1 (see Results). We thus hypothesized that if the individuals of site S8 are adapted to the cold waters of the Saguenay fjord (see Discussion), they should have higher normalized read depth than individuals from the warm waters of the nearby Gulf of St. Lawrence. We tested this hypothesis by building linear models where the log-transformed normalized read depth was included as the response variable and the type of site (S8 vs. the geographically close sites from the Gulf of St. Lawrence [S1, S2, S3, S4, S5 and S6]) as the explanatory variable. One model was performed for each putative CNV, and we used likelihood ratio tests and an FDR of 0.05 following the method of Benjamini and Hochberg (1995) to assess the significance of CNV–temperature associations.

3 | RESULTS

3.1 | Sequencing statistics and discovery of putative CNV

For the whole data set (1538 capelins from 35 spawning sites), GBS produced 897,382 ± 461,911 reads per sample on average before any quality filtering. The SNP calling process identified 642,098 SNPs that were successfully genotyped in at least 70% of the samples, with a low rate of missing data (median of 1.8%). The *HDplot* approach identified 280 duplicated markers and 14,319 SNP markers with high coverage distributed over 6620 putative CNVs. Moreover, 48,591 markers were classified as low confidence, and 578,819 markers were classified as singleton SNPs (without any filtering).

Among the 6620 putative CNVs, 1519 (22%) were located within protein-coding genes and 327 (5%) corresponded to repeated elements. Among the latter, 50% were retrotransposons (i.e., LRT and non-LRT retrotransposons; for the specific families of retrotransposons, see Table S2). Moreover, 31% of the repeated elements corresponded to interspersed elements including simple repeats, low-complexity regions and satellites (Table S3). The other repeated elements were DNA transposons (14%) and small RNAs (5%) (Table S3).

3.2 | Associations between CNVs and fitness proxy

We detected four and six CNV candidates associated with female gonadosomatic index with FDR of 0.05 and 0.10 respectively (Table 1). CNV candidates were located on chromosomes 6, 9, 11, 16 and 19 (Figure 2). Three candidates were located in intergenic

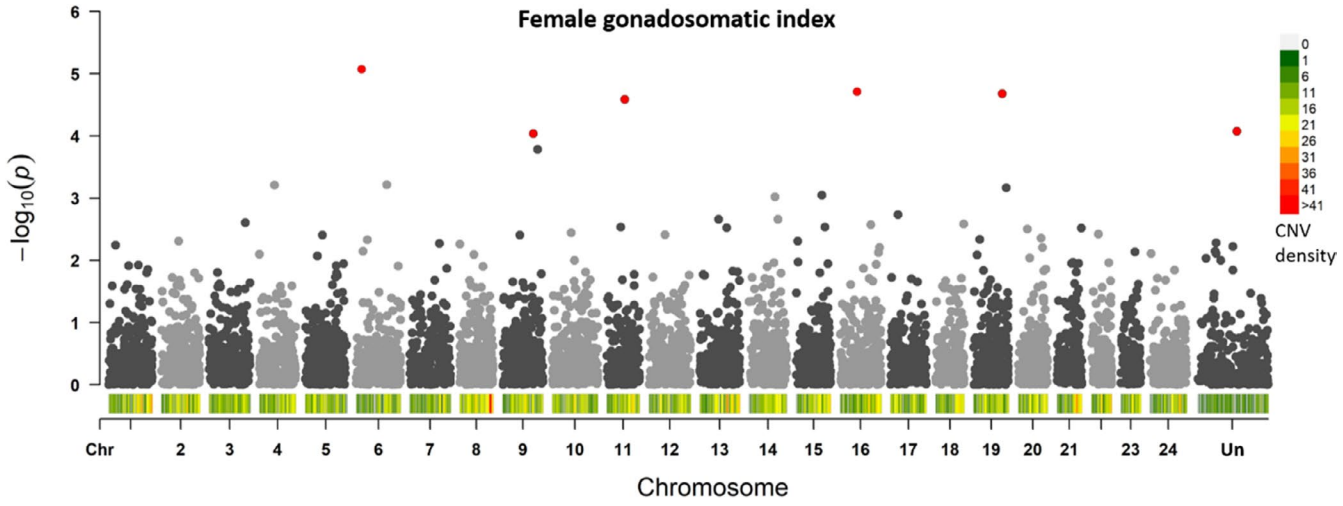

FIGURE 2 Manhattan plot showing the distribution of candidate CNVs associated with female gonadosomatic index. The *p*-values (−log[*p*-value]) of LRT tests associated with LMMs are shown on the Manhattan plot. Red dots indicate the candidate CNVs based on an FDR of 0.10 [Colour figure can be viewed at wileyonlinelibrary.com]

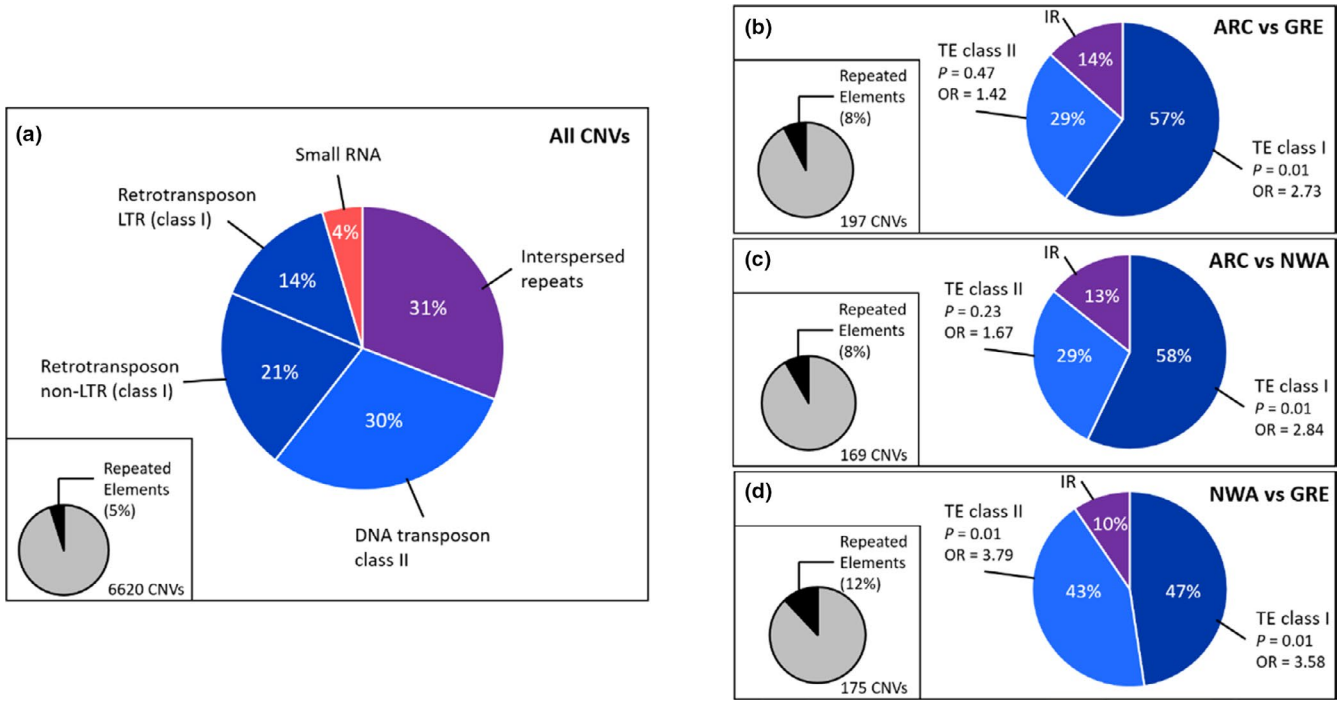

FIGURE 3 Repeated elements among the 6620 putative CNVs detected in the entire data set. (a) Type of repeated elements (i.e., interspersed repeats, DNA transposons, LTR and non-LTR retrotransposons, and small RNAs). Interspersed repeats include simple repeats, DNA satellites and low-complexity regions. (b–d) TEs in the set of candidate CNVs detected for each pair of lineages (ARC–GRE, ARC–NWA and NWA–GRE). The proportion of repeated elements among candidate CNVs is also presented, as well as the composition of three classes of repeated elements: TE class I (i.e., retrotransposons, LTR and non-LTR), TE class II (i.e., DNA transposons) and interspersed repeats (IR). The results of the Fisher tests performed to examine an excess of TE classes I and II in candidate CNVs are provided (*p*-value and odds ratio, OR) [Colour figure can be viewed at wileyonlinelibrary.com]

regions, far from protein-coding genes, whereas three others were within or close to the sequence of a gene (Table 1). The candidate 50008_8 was found within the sequence of a gene (Table 1) with unknown function. CNV 48830_43 was included within a 5-kbp window surrounding a gene involved in the regulation of various cellular processes (Table 1). The candidate 25048_60 was in an intergenic region at a distance of ~20 kbp from a gene (XM_012823774.1) that regulates the activation of the follicle-stimulating hormone (FSH) receptor (Table 1), which directly controls female oocyte production in vertebrates. The relationship between the gonadosomatic index and the normalized read depth of this candidate was negative (slope coefficient: −0.15 ± 0.03, R^2 = .11), suggesting that a high number of

copies of the gene (or its promotor) regulating the receptor of the FSH could decrease female fecundity.

3.3 | Associations between CNVs and temperature

Variance decomposition analyses showed that temperature explained at least five times more of read depth variation (2.53%) than salinity (0.36%) and chlorophyll concentration (0.55%) (Table S8). Therefore, further analyses were focused on temperature only. The pRDA (Figure 3a) built to detect CNVs associated with temperature was highly significant (df = 1, F-statistic = 26.79, p = .001), although the coefficient of determination was low (adjusted r^2 = .03). The pRDA detected 106 candidate CNVs associated with temperature, whereas LMMs identified 1932 candidate CNVs (Figure 3b). A total of 105 CNVs related to temperature were common to the two methods and thus considered as strong candidates (Figure 3b). Those CNVs were spread among all chromosomes except 16, 17 and 22 (Figure 3d).

For all 105 candidate CNVs, correlations between normalized read depth and temperature were negative (for the slope coefficients, see Table S3), indicating that the number of copies decreased with temperature in both coding and noncoding regions (Figure 3c). Twenty-eight candidate CNVs (23%) were located within the sequence of protein-coding genes (see Table S4). Fisher tests showed that those candidates were not present in excess within protein-coding genes (odds ratio = 1.12, p = .75). Furthermore, our analyses revealed that six of those candidate CNVs (6%) corresponded to repeated elements. Four of them were DNA transposons, one was a retrotransposon (LTR, Gypsy family) and one corresponded to simple repeats. The low number of repeated elements precluded rigorous statistical analysis of their composition or their potential excess in the representation.

The identification of discrete groups of copy number categories was examined using the model-based clustering procedure for 20 strong CNV candidates (10 loci located in noncoding regions and 10 within protein-coding genes; those with highest R^2 for each category) associated with temperature. This analysis showed that among these 20 candidate CNVs, up to three and six discrete categories can be delineated for CNVs located in coding and noncoding regions respectively (Figure S3), where each category may represent a specific number of copies for a given locus.

3.4 | Associations between CNVs and ancient lineages

The pRDA performed to identify CNVs associated with the three divergent capelin lineages was highly significant for all pairs of lineages: ARC–GRE (df = 1, F-statistic = 10.05, p = .001), ARC–NWA (df = 1, F-statistic = 17.03, p = .001), and NWA–GRE (df = 1, F-statistic = 18.01, p = .001). We detected 205 candidate CNVs associated with the divergence between ARC and GRE, 170 candidates between ARC and NWA, and 183 between NWA and GRE (Figure 4). Yet, the coefficient of determination was low in all cases (ARC–GRE: adjusted r^2 = .03; ARC–NWA: adjusted r^2 = .01; and NWA–GRE: adjusted r^2 = .01).

Using LMMs, we identified 761 candidate CNVs associated with the divergence between lineages ARC and GRE, 1034 candidates between ARC and NWA, and 884 between NWA and GRE (Figure 4). Among those, 197 were detected by both pRDA and LMMs for the pair of lineages ARC–GRE, 169 for the pair ARC and NWA, and 175 for the pair NWA and GRE, making them strong candidates associated with divergence between these lineages (Figure 4). Twenty-one of those CNVs were common to all three pairwise comparisons (Figure 4). Candidate CNVs were spread across the 24 chromosomes regardless of the pair of lineages considered (Figure 5).

Forty-six (23%), 43 (25%) and 30 (17%) candidate CNVs were located within protein-coding genes in the pairs of lineages ARC–GRE, ARC–NWA and NWA–GRE respectively (see Table S5). Fisher tests indicated that these candidates were not present in excess in the sequences of protein-coding genes (ARC–GRE: odds ratio = 0.75, p = .94, ARC–NWA: odds ratio =1.11, p = .30, NWA–GRE: odds ratio =1.02, p = .48).

Eight per cent of candidate CNVs were repeated elements for the ARC–GRE and ARC–NWA comparisons, and 12% for NWA–GRE (Figure 6b–d; see the detailed composition of those repeated elements in Table S6). Those repeated elements were mainly DNA transposons and retrotransposons (Figure 6). Fisher tests revealed that the candidate CNVs of the three lineage pairs were enriched for class I TEs (Figure 6). That is, both LRT and non-LRT retrotransposons were about three times more abundant in the set of candidate CNVs than in the entire set of 6620 putative CNVs. By contrast, DNA transposons were found in excess (about four times higher than expected) for the NWA–GRE lineage pair only (Figure 6).

3.5 | Hierarchical genetic structure based on CNVs

Our findings suggest that the significant association between CNVs and temperature could reflect adaptive genetic structure within the NWA lineage. Hierarchical clustering analyses (based on the 6620 putative CNVs) resolved two genetic clusters, one occurring in the northern part of the study area (cluster1) along the Atlantic coast and the other occupying the Gulf of St. Lawrence (cluster 2) (Figures 7a and 8). The CNV divergence between the two clusters was strongly supported by the data, as reflected by the 100% bootstrap support for the first node of the dendrogram (Figure 7a). Cluster 1 (nine spawning sites) occurs in relatively cold waters, with a mean bottom temperature of 1.71 ± 1.33°C (min: −0.13, max: 2.82). By contrast, cluster 2 (nine spawning sites) occurs in warmer waters with a mean bottom temperature of 4.23 ± 1.47°C (min: 2.09, max: 5.97). The high bootstrap support (>80%) at the shallower nodes of the dendrogram (Figure 7a) suggested a complex genetic substructure within the two clusters associated with geographical proximity and/or thermal similarity of spawning sites. The hierarchical clustering

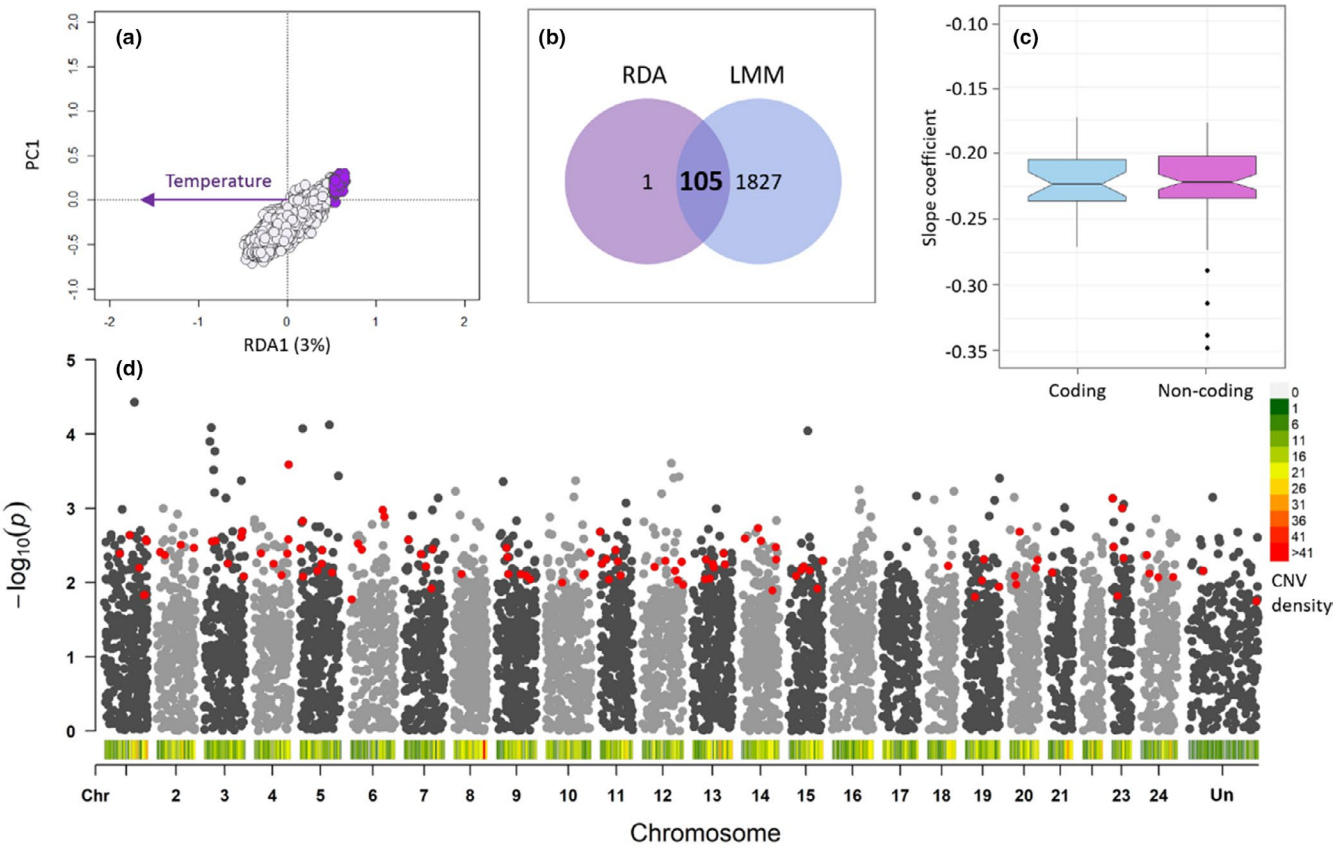

FIGURE 4 CNV–temperature associations in the Capelin. (a) pRDA including temperature: purple dots show candidate CNVs for temperature. (b) Venn diagram showing the number of candidate CNVs associated with temperature detected using the pRDA, LMMs and both methods combined (intercept). (c) Slope coefficients from LMMs for the effect of temperature in the 105 candidate CNVs: coefficients are always negative for both coding and noncoding regions (for regression outputs, see Table S3). (d) Manhattan plot showing the distribution of candidate CNVs associated with temperature along the Capelin genome. The *p*-values (−log[*p*-value]) of LRT tests associated with LMMs are shown on the Manhattan plot. Red dots indicate the candidate CNVs detected by both pRDA and LMMs [Colour figure can be viewed at wileyonlinelibrary.com]

analyses based on the whole set of putative CNVs (n = 6620) or the 105 candidates associated with temperature were very similar, suggesting that thermal selection acting on CNVs may drive the genetic structure within the NWA lineage.

Regardless the set of the loci used, the clustering analysis indicated that site S8 located in the Saguenay fjord in the upper part of the St. Lawrence Estuary was assigned to cluster 1 despite its closer geographical proximity to other sites within cluster 2 (Figure 8). Regression analyses showed that capelins from this site have a higher normalized read depth than those from spatially close sites within the Gulf of St. Lawrence for 104 of the 105 candidate CNVs associated with temperature (Table S7). This pattern supports the hypothesis that individuals from the Saguenay fjord are genetically more similar in terms of CNV variation to the putatively cold-adapted capelins from cluster 1 than the warm-adapted capelins from the Gulf of St. Lawrence.

Lastly, we investigated whether the extent of differentiation among the three divergent capelin lineages based on CNVs contributed to the genetic structure across the study area. Hierarchical clustering analysis based on the whole set of putative CNVs did not show any genetic structure associated with lineages (Figure 7b). The high bootstrap support for the first node (100) indicated the existence of two genetic clusters encompassing various sites from the three lineages (Figure 7b). Although bootstrap support of the shallower nodes was sometimes low, the classification suggested that geographical proximity and/or environmental similarity could partially influence the observed genetic structure (Figure 7b). Indeed, NWA sites belonging to genetic cluster 1 (L2, L10 and L6) and 2 (S1, S2 and S3) grouped within their respective branches of the dendrogram. Furthermore, geographically close sites of the GRE lineage (G1 and G2) were also positioned on the same branch; a similar observation could be made for G3 and G4. By contrast, the hierarchical analysis based on the 175 candidate CNVs associated with lineage divergence accurately classified the sites of the three lineages (Figure 7b). Those results clearly showed that geography and environmental variation within the distribution range of each lineage (especially temperature in NWA) are associated with broad variation in copy number that overrides the genetic signal driven by lineage divergence previously observed with SNP markers (Cayuela et al., 2020). In this previous study, SNPs showed a pronounced genetic

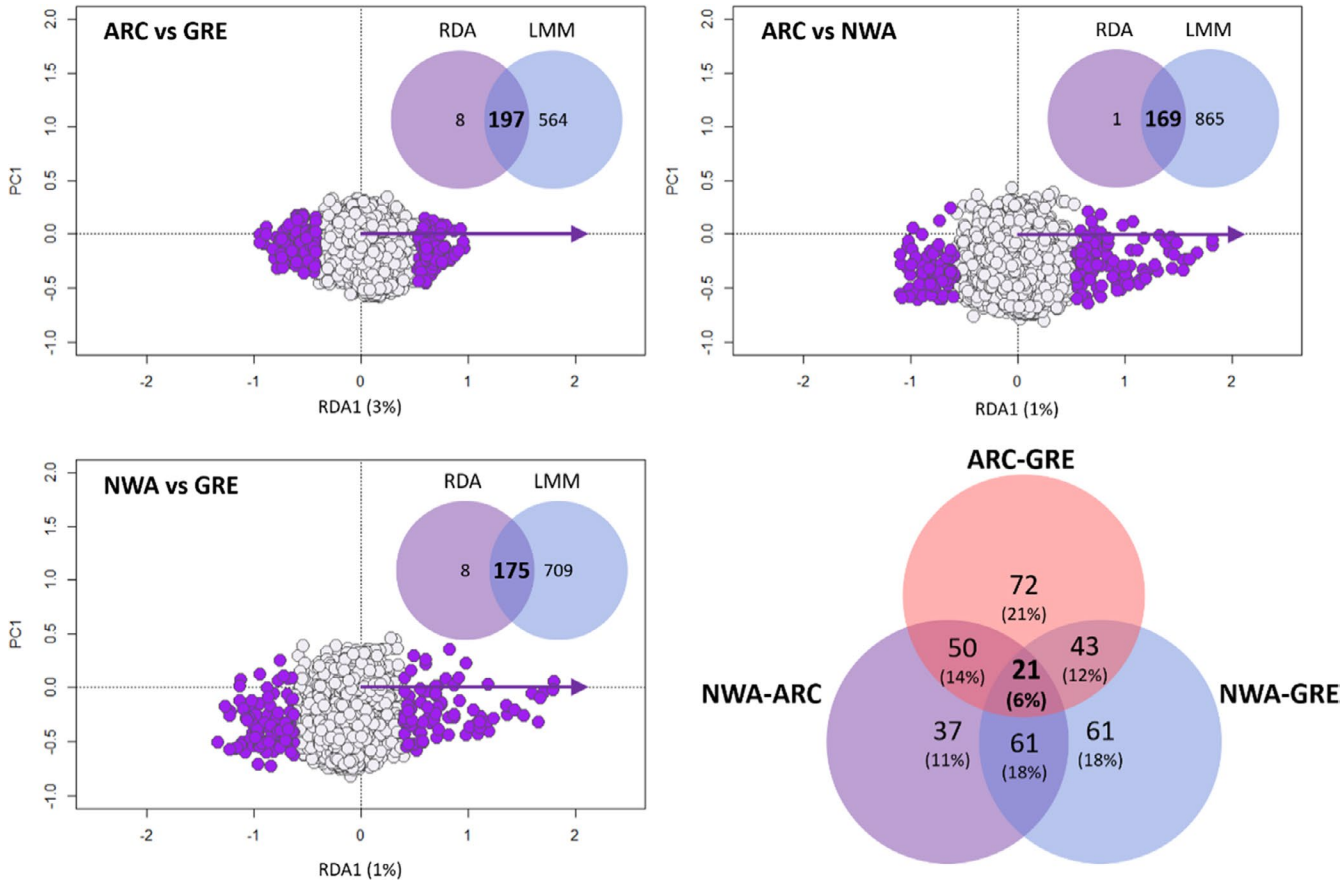

FIGURE 5 CNV–lineage associations in the Capelin. pRDAs include lineage as an explicative variable: purple dots show candidate CNVs. Venn diagrams for each of the pRDAs show the number of candidate CNVs associated with lineages detected using the pRDA, LMMs and both methods combined. The Venn diagram on the right shows the number of candidate CNVs shared between or private to each pair of lineages [Colour figure can be viewed at wileyonlinelibrary.com]

differentiation among lineages and extremely weak genetic structure within lineages.

4 | DISCUSSION

Our findings support our working hypothesis that CNVs play functional roles pertaining to fitness-related traits and local adaptation, and underly fine-scale genetic structure in capelin from the Northwest Atlantic. We found associations between CNVs and the gonadosomatic index, suggesting that copy number variation could affect female fitness by modulating oocyte production and thus fecundity. Second, we detected 105 candidate CNVs associated with temperature, of which 28 (20%) corresponded to genomic regions located within sequences of protein-coding genes. A salient observation was that for all 105 candidates, the normalized read depth was negatively correlated with temperature, supporting the hypothesis that fish using "cold water" spawning habitats have more gene copies than their counterparts from warmer spawning habitats. Third, we discovered 175 CNVs associated with the divergence of the three ancient capelin lineages. We found that 17–33% of the candidate CNVs were located within sequences of protein-coding genes, which might lead to the accumulation of genetic incompatibilities and thus could at least partly contribute to their reproductive isolation. Furthermore, we detected an excess of TEs (especially retrotransposons) in candidate CNVs, suggesting that these SVs have rapidly accumulated during the lineage divergence process. Lastly, clustering analyses revealed genetic structure within capelin lineages resulting from the effects of geography and temperature on local variation in copy numbers. Overall, our results underline the importance of considering CNVs in population genomics studies, as further discussed below.

4.1 | HDplot, a robust approach for CNV discovery in nonmodel species

In this study, we used *HDplot*, an approach developed by McKinney et al. (2017) as modified by Dorant et al. (2020). Simulations by and empirical data from McKinney et al. (2017) demonstrated the robustness of this approach using a set of simple summary statistics for SNP data by showing that it correctly identified duplicated sequences with >95% concordance for loci of known copy number. Moreover, our study adds to that of Dorant et al. (2020) in showing

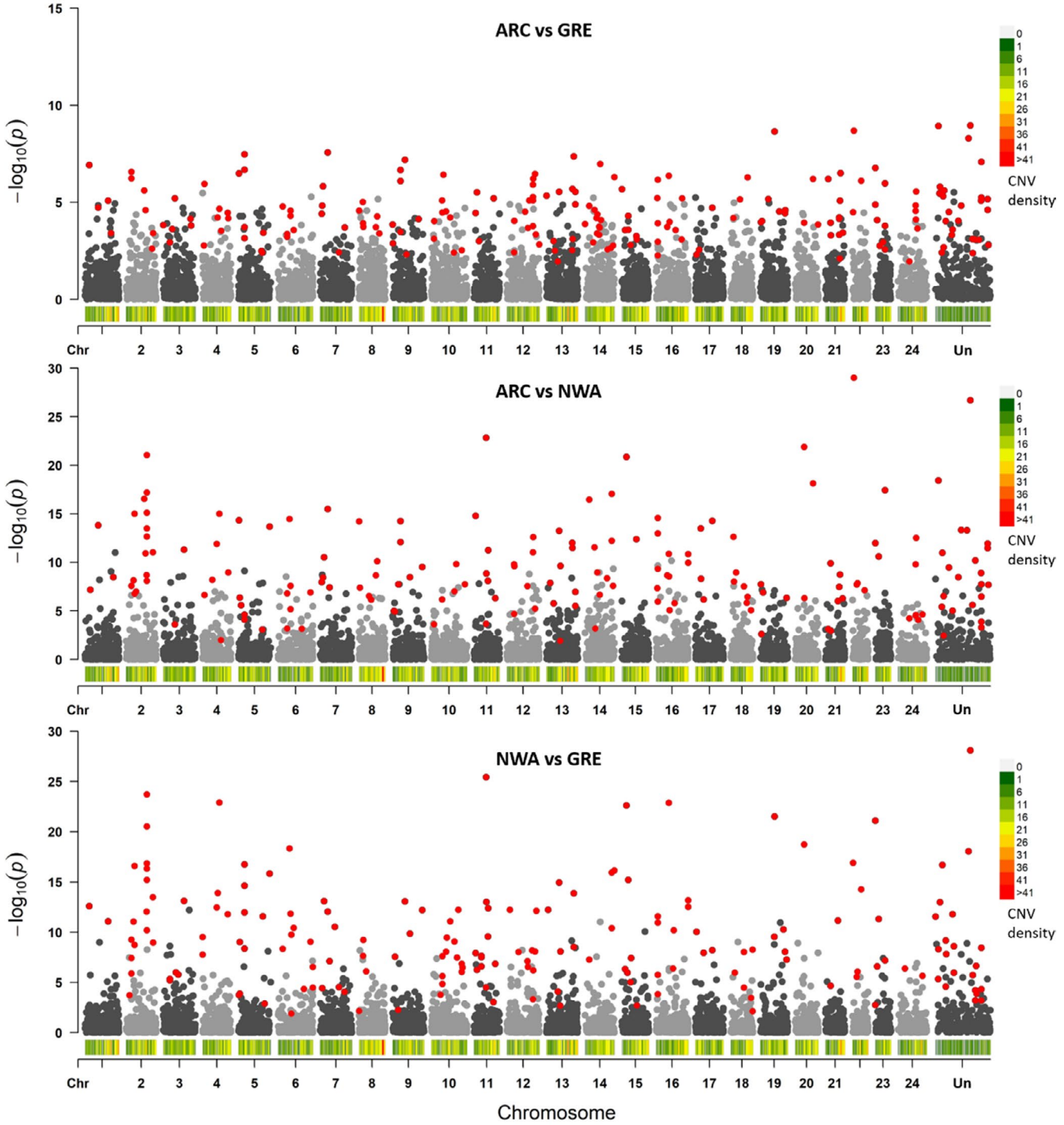

FIGURE 6 Manhattan plot showing the distribution of candidate CNVs associated with the three pairs of capelin lineages (i.e., ARC–GRE, ARC–NWA and NWA–GRE). The *p*-values (−log[*p*-value]) of LRT tests associated with LMMs are shown on the Manhattan plot. Red points indicate the candidate CNVs determined using both pRDA and LMMs [Colour figure can be viewed at wileyonlinelibrary.com]

that normalized read depth is a reliable proxy for the number of copies of a given duplicated locus.

Admittedly, this approach has the same technical limits as RAD-seq (i.e., reduced representation of the genome) and other methods of CNV detection based on SNP arrays and whole-genome sequencing (e.g., mappability issues, GC content, PCR duplicates, DNA library quality; Teo et al., 2012). In particular, the calling step of putative CNVs could be influenced by the quality of the reference genome on which RAD-seq reads are anchored. We therefore encourage further simulation studies to compare the efficiency of RAD-seq and whole-genome (read-pair, read-depth, split-read, and *de novo* assembly approaches; Pirooznia et al., 2015) CNV analyses, considering contrasted levels of reference genome quality. Overall, despite its technical limitations, RAD-seq CNV analysis provides a

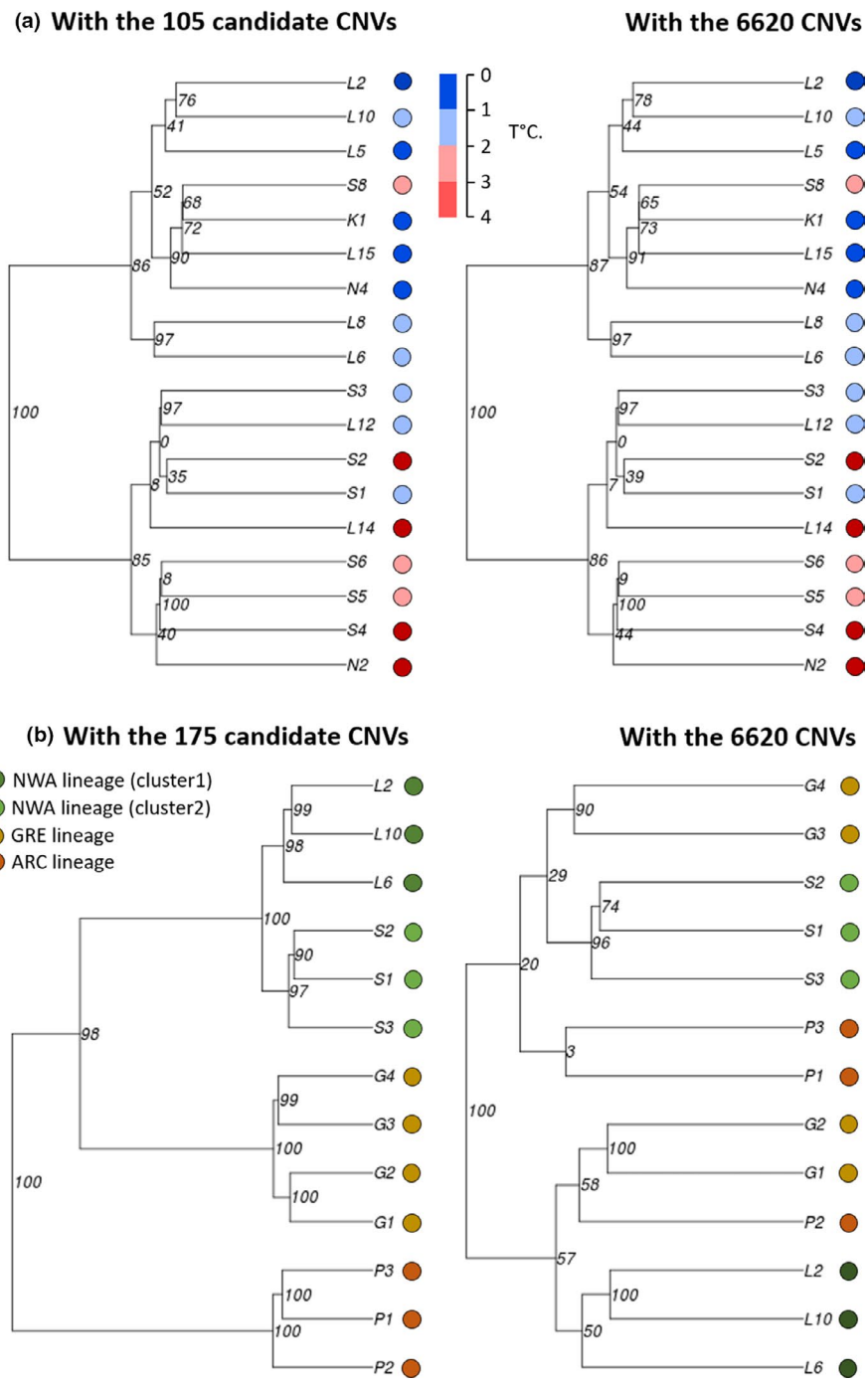

FIGURE 7 Dendrograms from hierarchical clustering analysis based on Ward's minimum variance method revealing genetic structure based on CNVs. To evaluate the robustness of the dendrogram clusters, we performed 10,000 bootstraps and dendrogram nodes were considered to be robust if their bootstrap value was >0.80. (a) We investigated how temperature of beach-spawning sites may affect the genetic structure inferred from CNVs within the NWA lineage. We performed two hierarchical clustering analyses: one based on the 6620 putative CNVs and the other with the 105 candidate CNVs associated with temperature. (b) We examined how lineage demographic divergence affects the genetic structure inferred from CNVs. To have a relatively balanced number of sites in the three lineages, we selected six spawning sites within the NWA lineage based on the clustering analysis performed for this lineage (a): three sites represented genetic cluster 1 (L2, L6, L10) and three sites the cluster 2 (S1, S2, S3). Two hierarchical clustering analyses were conducted: one based on the 6620 putative CNVs and the other with the 175 candidate CNVs associated with lineage divergence. See Figure 1 for sampling locations [Colour figure can be viewed at wileyonlinelibrary.com]

robust, novel tool for improving our knowledge on the evolutionary role of duplicated regions and TEs in nonmodel species by permitting studies involving hundreds of individuals or more from dozens of populations (Tigano, 2020).

4.2 | CNVs as a potential determinant of fitness variation

Our study revealed significant associations between the gonadosomatic index and six candidate CNVs, of which one was positioned within a coding gene with unknown functions. Interestingly, another candidate was found within an intergenic region of ~30 kbp of chromosome 11 at distance of ~20 kbp of the gene regulating the receptor of the FSH, a glycoprotein hormone secreted by the pituitary that stimulates early phases of gametogenesis. In female fish, FSH regulates both the secretion of oestradiol and the incorporation of vitellogenins into the oocytes (Yaron & Levavi-Sivan, 2011), and thus determines the number, the size and the energy content of eggs that a female can produce. This hormone could therefore be contributing to variation in the gonadosomatic index. The gonadosomatic index was negatively correlated with the normalized read depth of this candidate CNV, suggesting that a high number of copies of the gene (or its promotor) regulating the receptor of the FSH could decrease

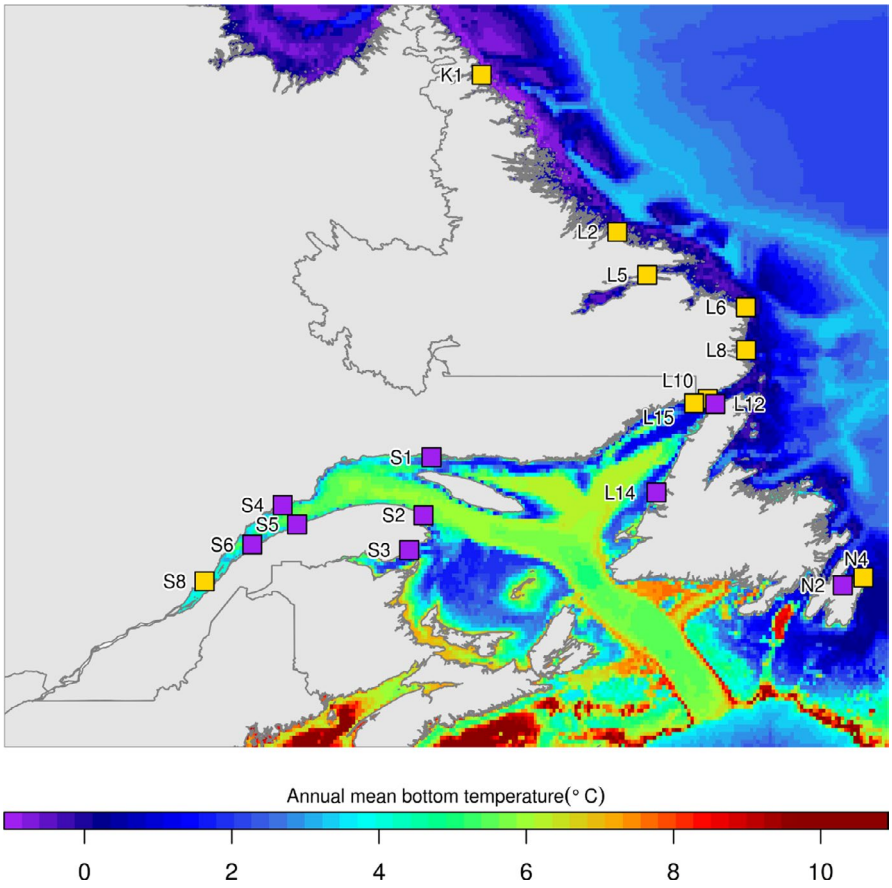

FIGURE 8 Map showing the genetic structure inferred by CNVs for the 18 beach-spawning sites of the NWA lineage. Yellow and purple squares show the spawning sites grouped within genetic clusters 1 and 2 respectively. Annual mean bottom temperature was downloaded from Bio-ORACLE (http://www.bio-oracle.org/) [Colour figure can be viewed at wileyonlinelibrary.com]

female fecundity. This duplicated region is thus an interesting candidate potentially involved in the endocrine control of a major component of fitness in capelin and perhaps other fishes.

4.3 | CNVs as a potential mechanism promoting thermal adaptation

Our study highlighted CNV–temperature associations, suggesting a possible contribution of these SVs in thermal adaptation. A total of 105 candidate CNVs were detected using a conservative approach combining regression and redundancy analyses. The normalized read depth of all duplicated regions was negatively correlated with temperature, indicating that capelin populations associated with colder waters tend to be characterized by a higher number of copies in various regions of the genome. Seventy-two of the candidate CNVs were located within intergenic regions and could influence gene expression if they include regulatory regions or belong to large duplicated DNA fragments encompassing one or more genes (Kondrashov, 2012; Qian & Zhang, 2014). Moreover, 28 of the candidates were located within sequences of protein-coding genes regulating central nervous system development (e.g., neurogenesis and visual perception) and various molecular processes (e.g., DNA binding and protein receptor). Interestingly, CNV loci were also detected within genes involved in cell protection (XM_012826890.1; autophagy regulation) and intracellular response to environmental stress (XM_012838233.1; Eukaryotic translation initiation factor 2-alpha kinase 3). The latter candidate gene (XM_012838233.1) regulates the production of AMP-activated protein kinase, an enzyme playing a key role in regulating oxidative stress, eliminating cellular injury, and maintaining or re-establishing intracellular homoeostasis in the face of environmental stresses (Wang et al., 2012). Overall, our results are congruent with the findings of a previous study showing that variation in gene copy number is involved in adaptation to cold waters in other (polar) marine fishes (Chen et al., 2008).

We also showed that five other candidate CNVs were TEs (both DNA transposons and retrotransposons) whose normalized read depth, a proxy for their accumulation in the genome, was negatively correlated with temperature. This result is congruent with the hypothesis that TEs can be reactivated in response to external stress, which might favour local adaptation under several environmental conditions (Chuong et al., 2017; McClintock, 1950). Thermal stress may indeed cause the derepression of several families of TEs whose activation can in turn induce structural variation, providing a selective advantage in the face of specific climatic conditions (González et al., 2010). Although this mechanism has been investigated in organisms such as fruit flies and diatoms (González et al., 2010; Pargana et al., 2020), it remains largely unexplored in vertebrates. Nevertheless, recent studies suggest that TE accumulation is highly variable among bony fishes (Shao et al., 2019; Yuan et al., 2018). Moreover, a phylogenetic analysis performed on 39 teleost species highlighted an unexpected

clusterization of REX3 retroelements isolated from species living in cold waters compared with those isolated from species living in warmer waters, suggesting a possible selective role of temperature on this specific TE (Carducci et al., 2019). In our study system, the accumulation of TEs in the genome of individuals from colder waters could also possibly result from thermal stresses endured during their lifetime. Alternatively, the variation observed at the population level in copy number of TEs between cold and warm environments could be the outcome of selective processes and reflect local thermal adaptation, such as the one we proposed above for non-TE CNVs associated with temperature. However, the GBS data provided in this study are not sufficient to formulate robust conclusions concerning the functional role of TEs. This issue could be better addressed through common garden experiments to investigate the influence of thermal stress on TE activation and its consequences on individual fitness (e.g., mortality rate) measured during early developmental stages (e.g., embryo and larvae).

4.4 | CNVs as potential drivers of reproductive isolation among nascent species

Our results suggest that CNVs could also be involved in the ongoing speciation process of capelin lineages by enhancing the rate of lineage divergence via the duplication or deletion of genomic regions. Indeed, the accumulation rate of CNVs is 1.5–2.5 times higher than that of SNPs in vertebrates (Paudel et al., 2015; Sudmant et al., 2013). Thus, candidate CNVs associated with lineage differentiation might have played a role in the rapid accumulation of genetic incompatibilities. Although we did not detect any enrichment for CNV presence within protein-coding genes, we showed that 17–33% of the candidate CNVs are located within gene sequences and could therefore lead to the alteration of gene expression via dosage effects, the disruption of genic functions, and the neofunctionalization of genes (Kondrashov, 2012; Qian & Zhang, 2014). In turn, those changes could have contributed to the emergence of reproductive barriers, possibly early in the divergence process of capelin lineages.

Our analyses also revealed that among the candidate CNVs, retrotransposons were three to four times more abundant than expected by chance in explaining the differentiation between the three lineages of capelin from the North Atlantic. In contrast, DNA transposons were found in excess for the lineage pair NWA–GRE only. This result suggests strongly that TEs contribute to the structural differentiation of genomes among capelin lineages. Here, it is noteworthy that retrotransposons are expected to accumulate faster in terms of copy number compared to DNA transposons based on their respective mode of transposition (i.e., *copy and paste* for retrotransposons vs. *cut and paste* for DNA transposons) (Calos & Miller, 1980). These observations raise the hypothesis that the accumulation of retrotransposons could have contributed to the emergence of reproductive barriers, which could at least partly explain the apparent lack of admixture among capelin lineages in the absence of any physical or distance barrier between them (Cayuela et al., 2020).

4.5 | CNVs as DNA markers to resolve local adaptive structure

In comparison with cal area (Cayuela et al., 2020), our results revealed that CNVs and SNPs show very contrasted patterns of spatial structure, probably due to the different processes underlying their respective evolution. SNP analyses highlighted a pronounced genetic divergence among the three capelin lineages that diverged between 1.8 and 3.5 Mya and a weak intralineage genetic structure (Cayuela et al., 2020). This divergence resulted from the high differentiation of ~3000 SNPs (34–107 were fixed depending on the pairs being compared), probably due to the combined effects of genetic drift, selection and reproductive isolation between lineages (Cayuela et al., 2020). At the intralineage level, very large N^e and high gene flow among spawning sites weakens genetic structure, which was mainly revealed by a large, ancient chromosomal rearrangement in the NWA lineage (Cayuela et al., 2020). This rearrangement led to the co-occurrence of three haplogroups present in almost all beach-spawning sites at relatively stable frequency.

CNVs showed an opposite pattern to that observed with SNPs as the intralineage differentiation exceeded the interlineage structure. At the interlineage level, divergence is associated with 175 CNVs that probably differentiated during the allopatric phase of geographical isolation. However, the signal of lineage divergence is erased by site-specific changes in copy number within lineages, possibly caused by local variation in environmental conditions (e.g., temperature). This pattern could result from the fact that CNVs have a higher mutation rate than SNPs (Paudel et al., 2015; Redon et al., 2006; Sudmant et al., 2013), which may favour rapid evolution under novel environmental conditions (Kondrashov, 2012; Qian & Zhang, 2014). Rapid variation could also be induced by environmental stresses in the case of TEs (Chuong et al., 2017; McClintock, 1950).

Our data suggest that CNV characteristics make them more prone to reveal fine-scale adaptive structure compared to SNPs in a biological system characterized by high gene flow, which is congruent with the conclusions of Dorant et al. (2020) in their study on American lobster. Indeed, we showed that capelin experiencing the colder waters of Labrador and the Atlantic coast of Newfoundland (cluster 1) were generally characterized by a high number of copies for all CNVs associated with temperature, which could give them a selective advantage in colder environments (Chen et al., 2008). By contrast, individuals inhabiting the warmer waters of the Gulf of St. Lawrence (cluster 2) were systematically characterized by a lower copy number than their counterparts in cluster 1, resulting in pronounced genetic structure within the NWA lineage. On the eastern coast of Newfoundland, temperature variation caused by marine current and landscape is associated with a fine-scale genetic structure (Figure 8) between sites separated from each other by short marine distance (i.e., <30 km). Individuals from the warm spawning

site N2 (5.5°C, one of the warmest sites in the study area; Table S1) were assigned to cluster 2 whereas individuals from the cold site N4 (−0.4°C, the coldest in the study area) were attributed to cluster 1.

The reliability of our analyses based on CNVs is further supported by the unique case of the Saguenay fjord population, which reproduces on beach-spawning sites (S8 in our study) in the St. Lawrence Estuary, a few kilometres from the mouth of the fjord (Colbeck et al., 2011; Kenchington et al., 2015). The clustering analysis showed that S8 grouped with the cold spawning sites of cluster 1 despite being spatially proximal to sites in cluster 2. This intriguing result is congruent with those of two previous studies (using microsatellites and AFLP [amplified fragment length polymorphism] markers; Colbeck et al., 2011; Kenchington et al., 2015) showing that the individuals from the Saguenay fjord are more genetically related to those from Labrador and the northern part of Newfoundland (cluster 1) than to those of the Gulf of St. Lawrence (cluster 2). The Saguenay fjord, where the capelin population probably completes its life cycle outside of the breeding period (Lazartigues et al., 2016), is deep (max. ~250 m), has low bottom temperature (between −2 and 1.5°C depending on the place and the season, Galbraith et al., 2018), and is a refuge for many Arctic and sub-Arctic species of fish and invertebrates (Drainville, 1970; Judkins & Wright, 1974). Our results support the hypothesis formulated by Colbeck et al. (2011) and Kenchington et al. (2015) that the Saguenay fjord population could be locally adapted to cold waters. Our analyses showed that those individuals have a higher number of copies (for 104 CNVs of the 105 candidates associated with temperature) than those reproducing in the neighbouring beach-spawning sites in the Gulf of St. Lawrence.

The strong influence of temperature on the fine-scale distribution of the two genetic clusters suggests that copy number differentiation between clusters 1 and 2 probably does not result from the presence of demographic subunits or glacial sublineages. This conclusion is also supported by SNP data that did not reveal any genetic substructure within the NWA lineage (Cayuela et al., 2020). However, the occurrence of complex substructure (i.e., clustering of spatially close spawning sites) within both genetic clusters suggest that other environmental or geographical factors could cause additional local variation in copy number.

4.6 | Inferring genetic structure using CNVs: research avenues and implications for conservation and management

The study of Dorant et al. (2020) and ours showed that putative CNVs discovered from RAD-seq data represent an under-explored type of DNA marker that can be investigated in population genomics studies of nonmodel species and using large data sets. Our analyses unveiled complex hierarchical patterns of structure determined by temperature and local geography that differ from patterns based on SNPs. These two types of markers thus appear highly complementary to document genetic structure, particularly in systems with large N^e and high migration rates, which is the case for the majority of marine species. In these systems, CNVs may be better suited to resolving fine-scale spatial structure driven by contemporary evolutionary processes than SNPs, which could be more efficient to capture large-scale structure resulting from historical and demographic processes. This is consistent with recent studies showing that variation in CNV showed more population-specific structure than SNPs or deletions in humans (Sudmant et al., 2015; Yang et al., 2018) or that revealed pronounced copy-number variation within domestic species (Serres-Armero et al., 2017). With increasing possibilities to investigate all types of genetic variation, not only SNPs but also CNVs or other kinds of SVs, we envisage that our approach based on the pioneering work of McKinney et al. (2017) will be extended to investigate a broader range of organisms occupying various habitats (i.e., freshwater and terrestrial environments) and characterized by a broad range of demographic characteristics (i.e., small N^e, reduced migration). This would allow further investigation of the environmental and demographic circumstances under which SVs provide a different understanding of population genetic structure compared to SNPs.

At the applied level, CNV analyses could help to improve the definition of evolutionarily significant units, distinct population segments and management units (Allendorf et al., 2010) by considering an additional type of genetic polymorphism. Our findings show that the study of those SVs from RAD-seq data is a cost-effective way to determine conservation units and demographic clusters for species whose population genetic structure is weak and difficult to resolve. In the case of the capelin, we showed that the spawning sites of the NWA lineage can be separated into two genetic clusters that could be considered as different management units by fishery managers and that were undetected using SNPs (Cayuela et al., 2020). Overall, our results and those of Dorant et al. (2020) outline that CNV markers could contribute to improve the spatial delineation of marine protected areas and fishery stocks at a time when global seafood production has reached approximatively 90 million tonnes per year (wild caught) and many species are declining due to overfishing (Asche, 2018; FAO, 2018).

ACKNOWLEDGMENTS

We thank biologists and technicians of the Department of Fisheries and Oceans Canada for their help as well as all everyone who contributed to sampling throughout the study area. We are also grateful to Associate editor Paul Hohenlohe and three anonymous reviewers for their constructive comments on a previous version of the manuscript. This research was funded by a Strategic Project Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to L. Bernatchez, M. Clément and P. Sirois, a financial contribution of Ressources Aquatiques Québec and was also supported by in-kind contribution from many other organizations: Department of Fisheries and Oceans Canada, Nunatsiavut Government, NunatuKavut Community Council, Labrador Fishermen's Union Shrimp Company, Department of Fisheries and Aquaculture – Government of Newfoundland and Labrador, World Wildlife Fund Canada, St. Lawrence Global Observatory, Parc

Marin du Saguenay–Saint-Laurent, and the Greenland Institute of Natural Resources. Hugo Cayuela was supported as a postdoctoral researcher by the Vanier-Banting postdoctoral fellowship programme and the Swiss National Science Foundation (SNF grant no. 31003A_182265).

AUTHOR CONTRIBUTIONS

H.C. performed the statistical analyses and wrote the paper. Y.D. and E.N. contributed to the bioinformatics and statistical analyses. L.B., M.C. and P.S. initiated the project, and L.B. conceptualized and coordinated the work. S.G.-H. performed gonadosomatic index measurements. C.M. and M.L. contributed to the writing of the article. All authors read and edited the final version of the manuscript.

DATA AVAILABILITY STATEMENT

Raw sequencing data for GBS libraries are available under accession no. PRJNA631144. Files of normalized read depth for the 6620 putative CNVs and environmental data used in CNV analysis are available on Dryad (https://doi.org/10.5061/dryad.msbcc2fx7).

ORCID

Hugo Cayuela https://orcid.org/0000-0002-3529-0736
Yann Dorant https://orcid.org/0000-0002-7295-9398
Claire Mérot https://orcid.org/0000-0003-2607-7818

REFERENCES

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709.

Asche, F. (2018). Impacts of climate change on the production and trade of fish and fishery products. The State of Agricultural Commodity Markets (SOCO) 2018 - Background Paper. Rome, FAO.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300.

Böhne, A., Brunet, F., Galiana-Arnoux, D., Schultheis, C., & Volff, J. N. (2008). Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research*, *16*, 203–215.

Bosch, S., Tyberghein, L., & De Clerck, O. (2017). *sdmpredictors: An R package for species distribution modelling predictor datasets*. Marine Species Distributions: From data to predictive models, 49.

Brewer, S. K., Rabeni, C. F., & Papoulias, D. M. (2008). Comparing histology and gonadosomatic index for determining spawning condition of small-bodied riverine fishes. *Ecology of Freshwater Fish*, *17*, 54–58.

Calos, M. P., & Miller, J. H. (1980). Transposable elements. *Cell*, *20*, 579–595.

Carducci, F., Biscotti, M. A., Forconi, M., Barucca, M., & Canapa, A. (2019). An intriguing relationship between teleost Rex3 retroelement and environmental temperature. *Biology Letters*, *15*, 20190279.

Casacuberta, E., & González, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular Ecology*, *22*, 1503–1517.

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

Cayuela, H., Rougemont, Q., Laporte, M., Mérot, C., Normandeau, E., Dorant, Y., Tørresen, O. K., Hoff, S. N. K., Jentoft, S., Sirois, P., Castonguay, M., Jansen, T., Praebel, K., Clément, M., & Bernatchez, L. (2020). Shared ancestral polymorphism and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Molecular Ecology*, *29*, 2379–2398.

Chain, F. J., & Feulner, P. G. (2014). Ecological and evolutionary implications of genomic structural variations. *Frontiers in Genetics*, *5*, 326.

Chen, Z., Cheng, C. H. C., Zhang, J., Cao, L., Chen, L., Zhou, L., Jin, Y., Ye, H., Deng, C., Dai, Z., Xu, Q., Hu, P., Sun, S., Shen, Y., & Chen, L. (2008). Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 12944–12949.

Christiansen, J. S., Præbel, K., Siikavuopio, S. I., & Carscadden, J. E. (2008). Facultative semelparity in capelin *Mallotus villosus* (Osmeridae)-an experimental test of a life history phenomenon in a sub-arctic fish. *Journal of Experimental Marine Biology and Ecology*, *360*, 47–55.

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics*, *18*, 71.

Colbeck, G. J., Turgeon, J., Sirois, P., & Dodson, J. J. (2011). Historical introgression and the role of selective vs. neutral processes in structuring nuclear genetic variation (AFLP) in a circumpolar marine fish, the capelin (*Mallotus villosus*). *Molecular Ecology*, *20*, 1976–1987.

Dion-Côté, A.-M., Renaut, S., Normandeau, E., & Bernatchez, L. (2014). RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Molecular Biology and Evolution*, *31*, 1188–1199.

Dodson, J. J., Tremblay, S., Colombani, F., Carscadden, J. E., & Lecomte, F. (2007). Trans-Arctic dispersals and the evolution of a circumpolar marine fish species complex, the capelin (*Mallotus villosus*). *Molecular Ecology*, *16*, 5030–5043.

Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., Normandeau, E., Rochette, R., & Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype-temperature association in a marine species. *Molecular Ecology*, *29*, 4765–4782. https://doi.org/10.1111/mec.15565.

Drainville, G. (1970). The Saguenay fjord: The ichthyological fauna and the ecological conditions. *Le Naturaliste Canadien*, *97*, 623–666.

FAO (2018). *The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals*. Rome.

Farslow, J. C., Lipinski, K. J., Packard, L. B., Edgley, M. L., Taylor, J., Flibotte, S., & Bergthorsson, U. (2015). Rapid Increase in frequency of gene copy-number variants during experimental evolution in Caenorhabditis elegans. *BMC Genomics*, *16*, 1044.

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

Fraley, C., Raftery, A., Scrucca, L., Murphy, T. B., Fop, M., & Scrucca, M. L. (2012). *Package 'mclust'*. Retrieved from http://ftp.uwsg.indiana.edu/CRAN/web/packages/mclust/mclust.pdf

Frank, K. T., & Leggett, W. C. (1981). Prediction of egg development and mortality rates in capelin (*Mallotus villosus*) from meteorological, hydrographic, and biological factors. *Canadian Journal of Fisheries and Aquatic Sciences*, *38*, 1327–1338.

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., Aburatani, H., Jones, K. W., Tyler-Smith, C., Hurles, M. E., Carter, N. P., Scherer, S. W., & Lee, C. (2006). Copy number variation: New insights in genome diversity. *Genome Research*, *16*, 949–961.

Galbraith, P., Bourgault, D., & Belzile, M. (2018). Circulation et renouvellement des masses d'eau du fjord du Saguenay. *Le Naturaliste Canadien*, *142*, 36–46.

González, J., Karasov, T. L., Messer, P. W., & Petrov, D. A. (2010). Genome-wide patterns of adaptation to temperate environments

associated with transposable elements in Drosophila. *PLoS Genetics*, *6*, e1000905.

Gunderson, D. R. (1997). Trade-off between reproductive effort and adult survival in oviparous and viviparous fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, *54*, 990–998.

Hardie, D. C., & Hebert, P. D. N. (2003). The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome*, *46*, 683–706.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nature Reviews Genetics*, *10*, 551–564.

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, G. E., Maes, G. E., Dioperen, E., Carvalho, G. R., & Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, *11*, 123–136.

Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., Hand, B. K., Hohenlohe, P. A., Kardos, M., Koop, B., Sethuraman, A., Waples, R. S., & Luikart, G. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, *11*, 1197–1211.

Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. A., & Wheeler, T. J. (2016). Dfam database of repetitive DNA families. *Nucleic Acids Research*, *44*, 81–89.

Hull, R. M., Cruz, C., Jack, C. V., & Houseley, J. (2017). Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biology*, *15*, e2001333.

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*, 161–171.

Judkins, D. C., & Wright, R. (1974). New records of mysids oreo mysis-Nobilis G O Sars and Mysis-Litoralis (Banner) in Saguenay Fjord (St. Lawrence Estuary). *Canadian Journal of Zoology*, *52*, 1087–1090.

Katju, V., & Bergthorsson, U. (2013). Copy-number changes in evolution: Rates, fitness effects and adaptive significance. *Frontiers in Genetics*, *4*, 273.

Kenchington, E. L., Nakashima, B. S., Taggart, C. T., & Hamilton, L. C. (2015). Genetic structure of capelin (*Mallotus villosus*) in the Northwest Atlantic Ocean. *PLoS One*, *10*, e0122315.

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 5048–5057.

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., & Koonin, E. V. (2002). Selection in the evolution of gene duplications. *Genome Biology*, *3*, research0008-1.

Langfelder, P., & Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software*, *46*, i11.

Laporte, M., Le Luyer, J., Rougeux, C., Dion-Côté, A.-M., Krick, M., & Bernatchez, L. (2019). DNA methylation reprogramming, TEs derepression and postzygotic isolation of nascent species. Science. *Advances.*, *5*(10), eaaw1644.

Laporte, M., Pavey, S. A., Rougeux, C., Pierron, F., Lauzent, M., Budzinski, H., Labadie, P., Geneste, E., Couture, P., Baudrimont, M., & Bernatchez, L. (2016). RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, *25*, 219–237.

Lazartigues, A. V., Plourde, S., Dodson, J. J., Morissette, O., Ouellet, P., & Sirois, P. (2016). Determining natal sources of capelin in a boreal marine park using otolith microchemistry. *ICES Journal of Marine Science*, *73*, 2644–2652.

Le Luyer, J., Laporte, M., Beacham, T. D., Kaukinen, K. H., Withler, R. E., Leong, J. S., Rondeau, E. B., Koop, B. F., & Bernatchez, L. (2017). Parallel epigenetic modifications induced by hatchery rearing in a Pacific salmon. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 12964–12969.

Leaché, A. D., & Oaks, J. R. (2017). The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *48*, 69–84.

Legendre, P., & Legendre, L. F. (2012). *Numerical ecology*. Elsevier.

Leggett, W. C., Frank, K. T., & Carscadden, J. E. (1984). Meteorological and hydrographic regulation of year-class strength in capelin (*Mallotus villosus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *41*, 1193–1201.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology*, *11*, 2453–2465.

Lynch, M., & Force, A. G. (2000). The origin of interspecific genomic incompatibility via gene duplication. *The American Naturalist*, *156*, 590–605.

Makałowski, W., Gotea, V., Pande, A., & Makałowska, I. (2019). Transposable elements: Classification, identification, and their use as a tool for comparative genomics. *Evolutionary Genomics* (pp. 177–207). Humana.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, 10–12.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*, *36*, 344–355.

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*, 656–669.

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology and Evolution*, *35*(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002.

Morin, P. A., Luikart, G., Wayne, R. K., & the SNP workshop group. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, *19*, 208–216.

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification*, *31*, 274–295.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375–402.

Paradis, E., Blomberg, S., Bolker, B., Brown, J., Claude, J., Cuong, H. S., & Desper, R. (2019). Package 'ape'. Analyses of Phylogenetics and Evolution, version, 2.

Pargana, A., Musacchia, F., Sanges, R., Russo, M. T., Ferrante, M. I., Bowler, C., & Zingone, A. (2020). Intraspecific Diversity in the Cold Stress Response of Transposable Elements in the Diatom *Leptocylindrus aporus*. *Genes*, *11*, 9.

Paudel, Y., Madsen, O., Megens, H. J., Frantz, L. A., Bosse, M., Crooijmans, R. P., & Groenen, M. A. (2015). Copy number variation in the speciation of pigs: A possible prominent role for olfactory receptors. *BMC Genomics*, *16*, 330.

Pirooznia, M., Goes, F. S., & Zandi, P. P. (2015). Whole-genome CNV analysis: Advances in computational approaches. *Frontiers in Genetics*, *6*, 138.

Purchase, C. F. (2018). Low tolerance of salt water in a marine fish: New and historical evidence for surprising local adaption in the well-studied commercially exploited capelin. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*, 673–681.

Qian, W., & Zhang, J. (2014). Genomic evidence for adaptation by gene duplication. *Genome Research*, *24*, 1356–1362.

Raymond, M., Callaghan, A., Fort, P., & Pasteur, N. (1991). Worldwide migration of amplified insecticide resistance genes in mosquitoes. *Nature*, *350*, 151–153.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., González, J. R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., … Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature*, *444*, 444–454.

Ressel, K. N., Bell, J. L., & Sutton, T. M. (2020). Distribution and life history of spawning Capelin in subarctic Alaska. *Transactions of the American Fisheries Society*, *149*, 43–56.

Ricci, M., Peona, V., Guichard, E., Taccioli, C., & Boattini, A. (2018). Transposable elements activity is positively related to rate of speciation in mammals. *Journal of Molecular Evolution*, *86*, 303–310.

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

Schlötterer, C., & Pemberton, J. (1998). The use of microsatellites for genetic analysis of natural populations—a critical review. In R. DeSalle, & B. Schierwater (Eds.), *Molecular approaches to ecology and evolution* (pp. 71–86). Birkhäuser.

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stökl, J., Schultner, E., Wurm, Y., Smith, C. D., Yandell, M., Heinze, J., Gadau, J., & Oettler, J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, *5*, 5495.

Serrato-Capuchina, A., & Matute, D. R. (2018). The role of transposable elements in speciation. *Genes*, *9*, 254.

Serres-Armero, A., Povolotskaya, I. S., Quilez, J., Ramirez, O., Santpere, G., Kuderna, L. F. K., Hernandez-Rodriguez, J., Fernandez-Callejo, M., Gomez-Sanchez, D., Freedman, A. H., Fan, Z., Novembre, J., Navarro, A., Boyko, A., Wayne, R., Vilà, C., Lorente-Galdos, B., & Marques-Bonet, T. (2017). Similar genomic proportions of copy number variation within gray wolves and modern dog breeds inferred from whole genome sequencing. *BMC Genomics*, *18*, 977.

Shao, F., Han, M., & Peng, Z. (2019). Evolution and diversity of transposable elements in fish genomes. *Scientific Reports*, *9*, 1–8.

Shirk, A. J., Landguth, E. L., & Cushman, S. A. (2017). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology Resources*, *17*, 1308–1317.

Smit, A. F. A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0*. Retrieved from http://www.repeatmasker.org

Spielmann, M., Lupiáñez, D. G., & Mundlos, S. (2018). Structural variation in the 3D genome. *Nature Reviews Genetics*, *19*, 453–467.

Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*, *24*, 2241–2252.

Sudmant, P. H., Huddleston, J., Catacchio, C. R., Malig, M., Hillier, L. W., Baker, C., Mohajeri, K., Kondova, I., Bontrop, R. E., Persengiev, S., Antonacci, F., Ventura, M., Prado-Martinez, P., Great Ape Genome Project, Marques-Bonet, T., & Eichler, E. (2013). Evolution and diversity of copy number variation in the great ape lineage. *Genome Research*, *23*, 1373–1382.

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.

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*, 2711–2718.

Tigano, A. (2020). A population genomics approach to uncover the CNVs, and their evolutionary significance, hidden in reduced-representation sequencing data sets. *Molecular Ecology*, *27*(9), 2215–2233. https://doi.org/10.1111/mec.14584

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*, 507384.

van't Hof, A. E., Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., Hall, N., Darby, A. C., & Saccheri, I. J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, *534*, 102–105.

Wang, S., Song, P., & Zou, H. (2012). AMP-activated protein kinase, stress responses and cardiovascular diseases. *Clinical Science*, *122*(12), 555–573.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology and Evolution*, *33*, 427–440.

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*, 1203–1209.

Yang, X., Song, Z., Wu, C., Wang, W., Li, G., Zhang, W., Wu, L., & Lu, K. (2018). Constructing a database for the relations between CNV and human genetic diseases via systematic text mining. *BMC Bioinformatics*, *19*, 528.

Yaron, Z., & Levavi-Sivan, B. (2011). Endocrine regulation of fish reproduction. *Encyclopedia of Fish Physiology: From Genome to Environment*, *2*, 1500–1508.

Yuan, Z., Liu, S., Zhou, T., Tian, C., Bao, L., Dunham, R., & Liu, Z. (2018). Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genomics*, *19*, 141.

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics*, *10*, 451–481.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Cayuela H, Dorant Y, Mérot C, et al. Thermal adaptation rather than demographic history drives genetic structure inferred by copy number variants in a marine fish. *Mol Ecol.* 2021;30:1624–1641. https://doi.org/10.1111/mec.15835