DR. TONY KESS (Orcid ID : 0000-0002-1079-3791)
DR. SARAH J LEHNERT (Orcid ID : 0000-0002-3569-8299)
DR. KARA K.S. LAYTON (Orcid ID : 0000-0002-4302-3048)
DR. ANTHONY EINFELDT (Orcid ID : 0000-0003-4588-3479)
DR. DANIEL RUZZANTE (Orcid ID : 0000-0002-8536-8335)
DR. CAMERON NUGENT (Orcid ID : 0000-0002-1135-2605)

Article type : Original Article

Genomic basis of deep-water adaptation in Arctic Charr (*Salvelinus alpinus*) morphs

Kess, T.*[1], Dempson, J.B.[1], Lehnert, S.J.[1], Layton, K.[2], Einfeldt, A[3]., Bentzen, P.[3], Salisbury, S.J.[3], Messmer, A.M.[1], Duffy, S.[1], Ruzzante, D.E.[3], Nugent, C.M.[4], Ferguson, M.M.[4], Leong, J.S.[5], Koop, B.F.[5], O'Connell, M. F.[1], Bradbury, I.R[1].

[1]Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St. John's, Newfoundland and Labrador, Canada
[2]University of Aberdeen, School of Biological Sciences, Aberdeen, United Kingdom
[3]Dalhousie University, Department of Biology, Halifax, Nova Scotia, Canada
[4]University of Guelph, Department of Integrative Biology, Guelph, Ontario, Canada
[5]University of Victoria, Department of Biology, Victoria, British Columbia, Canada
*Corresponding author email: tony.kess@dfo-mpo.gc.ca

Abstract:

The post-glacial colonization of Gander Lake in Newfoundland, Canada, by Arctic Charr (*Salvelinus alpinus)* provides the opportunity to study the genomic basis of adaptation to extreme deep-water environments. Colonization of deep-water (> 50 m) habitats often requires extensive adaptation to cope with novel environmental challenges from high hydrostatic pressure, low temperature, and low light, but the genomic mechanisms underlying evolution in these environments are rarely known. Here, we compare genomic divergence between a deep-water morph adapted to depths of up to 288m and a larger, piscivorous pelagic morph occupying shallower depths. Using both a SNP array and resequencing of whole nuclear and mitochondrial genomes, we find clear genetic divergence (F_{ST} = 0.11 - 0.15) between deep and shallow water morphs, despite an absence of morph divergence across the mitochondrial genome. Outlier analyses identified many diverged genomic regions containing genes enriched for processes such as gene expression and DNA repair, cardiac function, and membrane transport. Detection of putative copy number variants (CNVs) uncovered 385 genes with CNVs distinct to piscivorous morphs, and 275 genes with CNVs distinct to deep-water morphs, enriched for processes associated with synapse assembly. Demographic analyses identified evidence for recent and local morph divergence, and ongoing reductions in diversity consistent with postglacial colonization. Together, these results show that Arctic Charr morph divergence has occurred through genome-wide differentiation and elevated divergence of genes underlying multiple cellular and physiological processes, providing insight into the genomic basis of adaptation in a deep-water habitat following postglacial recolonization.

Introduction

Understanding the relationship between genetic variation, adaptation, and the environment has been a fundamental goal of biology since the modern synthesis (Dobzhansky 1951), and advances in genome sequencing are now improving our ability to identify the molecular mechanisms that allow species to adapt to different habitats (Savolainen *et al.* 2013.). Extreme environments at species' biologically tolerable limits provide a unique opportunity to explore the genetic basis of adaptation, as organisms in these environments require extensive physiological and molecular changes to survive under these conditions (Rothschild & Mancinelli 2001). Recent genomic studies have begun to illuminate these molecular mechanisms (Wang & Guo 2019), including in fish adapted to extreme hydrogen sulfide concentrations (Tobler *et al.* 2018), and changes underlying human adaptation to altitude on the Tibetan plateau (Huerta-Sanchez *et al.* 2014). Genomic studies of colonization of extreme environments have found changes at individual genes (Ilardo & Nielsen 2018), as well as structural variation and changes in gene copy number (Rinker *et al.* 2019), indicating changes in different types of genetic variation may facilitate adaptation under extreme abiotic conditions. Identifying the genomic basis of these adaptations can uncover the molecular solutions that evolve in response to ecological challenges from extreme environmental stressors and provide insight into how biodiversity arises through genomic change.

Deep-water environments pose distinct challenges to survival, requiring novel phenotypic solutions in response to increased pressure, reduced temperature, and aphotic conditions. Increases in hydrostatic pressure and reductions in temperature can cause systemic effects on cardiac and nervous system function, as well as influence gene expression and cellular structure (reviewed in Brown & Thatje 2014; Pradillon & Gaill 2006). Light availability also changes with depth in the water column, affecting transmission and reception of visual signals between individuals (Endler 1992), and has been linked to repeated changes at opsin visual system genes (Schott *et al.* 2014; Musilova *et al.* 2019; Hill *et al.* 2019), and to diversification and speciation events in cichlid and stickleback lineages (Boughman 2001; Nagai *et al.* 2011: Seehausen *et al.* 2008). Recent studies have begun to provide insight into the molecular basis of adaptation in these habitats (Hahn *et al.* 2017; Wang *et al.* 2019) and more broadly provide valuable insight into the genomic basis of adaptation.

Arctic Charr (*Salvelinus alpinus*), the northernmost of all freshwater and diadromous fishes, is well adapted to extreme environments in cold Arctic waters, enabling postglacial (~10,000 years) colonization and evolution of extensive local adaptation across diverse northern habitats (Klemetsen 2010). Across the species range, Arctic Charr display repeated examples of morphs divergent in life history, morphology, and behaviour, often in sympatry (Jonsson *et al.* 1988; May-McNally *et al.* 2015; Michaud *et al.* 2008; Reist *et al.* 2013; Woods *et al.* 2013; Salisbury *et al.* 2018). Resource polymorphism, the ecological separation of distinct intraspecific morphs through different feeding or habitat use, has likely contributed to morph formation during recolonization through differences in depth preferences between morphs co-occurring with shifts in diet, and changes in body size and feeding morphology (Jonsson & Jonsson 2001; Skúlason & Smith 1995). The geologically brief period for colonization of multiple freshwater sites, and extensive diversity exhibited by Charr provide the opportunity to better understand how novel phenotypes arise during colonization of these habitats over short timescales (Klemetsen 2010).

The formation of divergent, depth-associated sympatric morphs in Gander Lake in Newfoundland, Canada, illustrates post-glacial divergence in low light, high pressure, and different thermal environments (Gomez-Uchida *et al.* 2008; Power *et al.* 2012). Gander Lake has a maximum depth of 288m, and has only recently de-glaciated within the past ~9000 – 13,000 years (O'Connel & Dempson 2002; Shaw *et al.* 2006). The deep-water (pale colouration, small, and benthic) and piscivorous (dark colouration, larger, and pelagic) morph pair found here exhibit separate exploitation of partially overlapping pelagic and deep-water niches with different thermal characteristics, potentially to minimize resource competition (Power *et al.* 2012). The deep-water morph inhabits one of the deepest aquatic environments colonized by Arctic Charr (Klemetsen 2010), at depths up to 280m (O'Connel & Dempson 2002; Power *et al.* 2005). This morph experiences low light, up to 28 atm pressure, and reduced temperatures (2–6 °C), likely requiring extensive adaption to cope with these extreme conditions. Genetic analysis of microsatellite markers has revealed strong nuclear differentiation between morphs (Gomez-Uchida *et al.* 2008), which contrasts recent identification of shared mitochondrial haplotypes, suggesting potential ecological divergence in sympatry (Salisbury *et al.* 2019), although the evolutionary history of morphs in this site remains unknown. Despite

phenotypic and ecological examination and known genetic divergence, the genomic basis of adaptations to extreme environmental variation in these morphs remains unclear (Power *et al.* 2005; Power *et al.* 2012). The recent availability of new genomic tools developed for *Salvelinus* species (Nugent *et al.* 2019) presents new opportunities to understand the genomic mechanisms that have enabled extensive Charr divergence across habitats such as Gander Lake, and the genomic basis of adaptation to extreme environments.

Here we use a single nucleotide polymorphism (SNP) array, resequencing of nuclear and mitochondrial genomes, and demographic analysis to investigate the genomic changes underlying deep-water and piscivorous morph divergence in Arctic Charr in Gander Lake. First, we quantify genome-wide divergence using a SNP array and whole genome resequencing. We carried out demographic reconstruction and historical admixture analyses to identify the evolutionary history of morph divergence. We then identify SNP outliers and candidate genes with morph-specific putative CNVs and investigate their potential functions using gene ontology and gene set enrichment. Genome-wide patterns of divergence in this system have yet to be investigated and can provide insight into the genomic architecture and molecular mechanisms underlying morph formation and deep-water adaptation.

Methods

Sampling

Artic Charr (piscivorous and deep-water morphs) were collected from Gander Lake (48°55 N, 54°35 W), Newfoundland, Canada (Figure 1A, morphs in Figure 1B) over three sampling periods (2001, n = 54; 2005 and 2006, n = 30); 2017, n = 61), using Lundgren small experimental gillnets of multiple mesh sizes ranging from 6.25-60 mm as outlined in Hammer and Filipsson (1985). Nets were set shallow (3.35 – 54.86 meters) to target piscivorous morphs and deep (167.64 – 281.94 meters) to target deep-water morphs. Typical soak times ranged from 16-24 hours. Clips were taken from the right pectoral fin to keep the left side physically intact and preserved in 95% ethanol. Video of deep-water morph presence at the lake bottom (Supplementary File 1, video) was also captured from Gander Lake using a remote operated vehicle in 2005.

Genotyping

Axiom array genotyping

DNA from 81 individuals (piscivorous = 31 [7 (2005/2006), 24 (2017)], deep-water = 50 [23 (2005/2006), 27 (2017)]) was extracted using the DNeasy 96 Blood and Tissue Kit (Qiagen). Genomic DNA quality was checked on 1% agarose gels and quantified using PicoGreen (ThermoFisher). These individuals were genotyped using a recently developed 87K Axiom SNP array (Nugent *et al.* 2019), and were genotyped among a larger set of genotyped Arctic Charr individuals from populations across Newfoundland and Labrador (Layton *et al.* 2021). We then selected polymorphic SNPs with high call confidence (n = 23,104), and retained SNPs within this set that aligned to contigs corresponding to chromosomal linkage groups, and that had a minimum allele frequency (MAF) of 0.01 among Gander Lake individuals and a genotyping rate > 0.95, resulting in 3677 SNPs used for population genetic comparisons. We selected this MAF threshold based on recent investigation of the influence of MAF on inference of population structure by Linck and Battey (2019). As this array was designed to include variable sites across multiple lineages (Nugent *et al.* 2019), most loci were monomorphic in Gander Lake and were dropped from analysis. Subsequent file conversions were conducted using *PGDSpider* (Lischer & Excoffier 2012), *plink 1.9* (Chang *et al.* 2015), *vcftools* (Danecek *et al.* 2011) and the *genepopedit* R package (Stanley *et al.* 2016).

Whole genome resequencing and processing

To characterize genomic divergence with higher resolution, we then sequenced whole genomes at low sequencing coverage for 10 deep-water and 10 piscivorous morphs, for analysis of both genotype likelihoods from each sequenced individual and pooling of individuals sequenced at low depth, as described in Therkildsen *et al.* (2017). Tissue samples from 10 of each morph collected in 2017, selected with equal sex ratios, and previously genotyped on the Axiom SNP array were selected for whole genome sequencing. Genomic DNA was extracted following Mayjonade *et al.* (2016) without grinding tissues and using Longmire's buffer (Longmire *et al*. 1997) for tissue lysis. Sequencing libraries were prepared following the protocol outlined by Therkildsen & Palumbi (2017)

with modifications. Briefly, Nextera DNA Flex Library Prep Kits (Illumina) were used at scaled down reaction volumes (0.13 X the standard Illumina protocol). Kapa Hi-Fidelity Library Amplification Kits (Roche) were used for library amplification at a 20 µl reaction volume with 4 µl Nextera Unique Dual Indexes Set A (Illumina). Libraries were quantified by Qubit (ThermoFisher) and checked for average fragment size on an Agilent Bioanalyzer. Libraries were normalized to equimolar concentrations and combined into a single pool for sequencing. Libraries for each individual were sequenced on an Illumina HiSeq4000 at Genome Quebec.

Sequenced libraries were analyzed for quality using FastQC, (Andrews 2010), and then trimmed to remove the leading 10 bases, adapter content, and 3' regions falling below a q score of 10 using *cutadapt* 2.1 (Martin 2011). Trimmed reads from each individual were aligned to the *Salvelinus* reference genome (GCA_002910315.2) using *bwa mem* (Li 2013). Following *Genomic Analysis Toolkit* (GATK 3.7) Best Practices recommendations (DePristo et al. 2011), duplicate reads from library preparation or sequencing artefacts were marked and removed using the *PicardTools* 2.20.6 *MarkDuplicates* function. De-duplicated reads were then sorted and realigned around potential insertions and deletions using *RealignerTargetCreator* and *IndelRealigner* functions in *GATK*.

Population structure and genetic diversity

Axiom array SNPs

To quantify divergence between morphs, we estimated mean F_{ST} (Weir & Cockerham 1984) between piscivorous and deep-water morphs using *plink 1.9*, as well as F_{ST} for each SNP. We then filtered SNPs in *plink* to produce a putatively neutral set of 2430 unlinked SNPs to quantify population structure. We removed all SNPs with F_{ST} values above the 95th percentile, or that exhibited linkage disequilibrium (LD; r^2) greater than 0.5 with other SNPs in a 10 SNP window. To quantify admixture proportions for each individual we used sparse non-negative matrix factorization (*snmf*) (Frichot *et al.* 2014) in the *R* package *LEA* (Frichot & Francois 2015) with default conditions, testing *K* values from 1 to 5. We then used principal component analysis (PCA) in the R package *pcadapt* (Luu *et al.* 2017) to cluster individuals by genotype using K=2. We also used the *adegenet* R package (Jombart & Ahmed 2011) to conduct discriminant analysis of principal components (DAPC; Jombart

et al. 2010) without including prior information on morph identity, and a machine learning clustering approach that maximizes projected variation on two axes, t-distributed neighbour embedding (van der Maaten & Hinton 2008), to cluster individuals by genotype using the R package *Rtsne* (Krijthe 2015). All analyses described above were conducted on individuals from all sampling years with genotype data (2017, 2005/2006) combined.

To account for multiple sampling years in our dataset, we also conducted a partial redundancy analysis using the *rda* function in the *vegan* R package (Oksanen *et al.* 2011), using sampling year as a condition. Significance of the model was assessed using 999 permutations. We then conducted hierarchical analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) in Arlequin 3.5.2.1 (Excoffier & Lischer 2010), grouping individuals by morphs and by years within morphs. Mean observed (H_O) and expected (H_E) heterozygosity were estimated for each morph using all 3677 SNPs in the R package *diveRsity* (Keenan *et al.* 2013); significance was assessed across all SNPs using a Mann Whitney U test.

We estimated effective population size (N_e) separately for each morph with the neutral, physically-unlinked SNPs, using the linkage-based method in NeEstimator 2.0 (Do *et al.* 2014) and a minor allele frequency cut-off of 0.05.

Axiom array signal intensity

Direct quantification of signal intensity variation on SNP arrays can be used as distinct measures of genetic variation allows bypassing classification and genotyping of complex polymorphisms. This approach can enable characterizing a full range of polymorphisms such as regions with loss of heterozygosity (Peiffer *et al.* 2006) and copy number variants (CNVs, Redon *et al.* 2006) using a SNP array with a single metric (Myles *et al.* 2015, (Cantsilieris *et al.* 2013). Here, we used the *PennCNV-Affy* pipeline for Axiom arrays (Wang *et al.* 2007) to quantify signal between samples based on 13,664 loci. To quantify variance in signal intensity within and between morphs, we calculated V_{ST} (Redon *et al.* 2006), an analogue of F_{ST}, which describes the proportion of variance in signal intensity explained by population, using *calc_Vst* from WinXPCNVer (Lou *et al.* 2015). We also conducted a multivariate quantification of signal intensity associated with morph divergence using redundancy

analysis (Rao 1964) in the *vegan* R package (Oksanen *et al.* 2011) using morph as a predictor, conditioned on sampling period. Significance of the model was assessed using 999 permutations.

Whole genome resequencing

Using whole genome resequencing, we then estimated population differentiation using pooled sequences from piscivorous and deep-water morphs, as well as from individual genotype likelihoods. Pools were created by combining realigned reads from all individually barcoded deep-water and piscivorous morphs using *SAMtools merge* (v. 1.10, Li et al 2009). Aligned read depths per pool and per sample were assessed in 20Kb windows using *Mosdepth* 0.2.6 (Pedersen & Quinlan 2018). We then calculated mpileup files for both morph pools together using *SAMtools mpileup*. All population genetic analyses were limited to sequence from contigs representing the 39 *Salvelinus* chromosomes. We used *Popoolation2* 1.201 (Koffler et al. 2011) *mpileup2sync* to make a sync file using the combined morph mpileup files, and estimated averaged F_{ST} for all SNPs in 1 Kbp windows, using *fst-sliding.pl*, with the following parameters: *--min-count 2, --min-coverage 4, --max-coverage 1000, –pool-size 10, --window-size 1000, step-size 1000, --supress-noninformative.*

To account for potential biases in divergence estimates from pooled sequences, we also calculated genotype likelihoods from realigned reads for each individual using Analysis of Next Generation Sequencing Data (*ANGSD* 0.933, Korneliussen et al. 2014) and output genotype likelihoods as beagle format likelihood files for all SNPs passing quality filers (*-minMapQ* 30 *-minQ* 20 *-SNP_pval* 2e-6 *-uniqueOnly* 1 *-remove_bads* 1). Genotype likelihoods were used for inference of population structure using both principal component analysis and admixture modelling in *PCAngsd* 0.981 (Meisner & Albrecthsen 2019).

Demographic and evolutionary history

Mitochondrial genome assembly and haplotype networks

To identify the colonization history of Arctic Charr morphs, we constructed haplotype networks from mitochondrial genomes obtained from each of the 20 re-sequenced individuals. We first isolated mtDNA reads from each individual by aligning to the Arctic Charr mitochondrial

genome (GenBank accession: NC_000861.1; Doiron *et al.* 2002) using *Bowtie2* 2.3.4.2 (Langmead & Salzberg 2012). We trimmed Illumina adapter sequences and applied stringent trimming parameters for sequence quality at leading and trailing bases (Q score > 25) and over a sliding window of 4 nucleotides (average Q-score > 23) using *Trimmomatic* 0.38 (Bolger, Lohse, & Usadel, 2014). We then performed *de novo* assembly for the mitochondrial genome of each individual using *MIRA* 4.0.2 (Chevreax, Wetter, & Suhai 1999). To identify and correct assembly errors, for each individual we remapped mitochondrial reads to *de novo* assemblies with *Bowtie2*, and then corrected miscalled bases, filled gaps, and marked ambiguous bases with *Pilon* 1.22 (Walker *et al.* 2014). We annotated mitochondrial genomes using *MITOS2* (Bernt *et al.* 2013) and determined substitution states using the standard vertebrate mitochondrial code. We created an unrooted haplotype network using the integer neighbour joining net algorithm implemented in *PopART* (Leigh & Bryant 2015). To simplify the observed haplotype network, we also used *FITCHI* (Matschiner 2016) to visualize a network restricted to 12 segregating transversions. Maximum likelihood trees in Newick format for *FITCHI* input files were estimated using aligned mitogenomes for all individuals in *MEGAX* for *MacOS* (Stecher *et al.* 2020).

Demographic and evolutionary history reconstruction

We used *treemix* 1.13 (Pickrell & Pritchard 2012), to infer the evolutionary history of Gander morphs compared to Arctic Charr populations across the Newfoundland and Labrador range. We combined SNP array genotypes with genotype data from Layton *et al.* (2021), consisting of 128 individuals from five additional sites across Newfoundland and Labrador (Figure 4B). We then identified and removed outlier loci identified in *pcadapt* with K=3, and carried out LD pruning as was carried out on Gander array SNPs, resulting in 5889 putatively neutral, unlinked SNPs. We then used this combined dataset to construct a maximum likelihood tree of Arctic Charr population and morph divergence in *treemix.*

We then used two methods for demographic reconstruction restricted to Gander Lake morphs, based on the site frequency spectrum from whole genome re-sequencing data, and linkage disequilibrium among SNPs from the Axiom array. First, we used *stairwayplot2* (Liu & Fu 2020),

which estimates N_e at different time periods using the site frequency spectrum to infer estimated N_e fluctuation coinciding with coalescent events. We generated folded site frequency spectra separately for each morph using the *-dosaf* flag and *-realSFS* function in *ANGSD*, restricting export to the 39 chromosome-scale contigs and high quality uniquely mapping sites (*-minMapQ* 30 *-minQ* 20 *-uniqueOnly* 1 *-remove_bads* 1) and including invariant sites by removing SNP probability filters (*snp_pval*), and setting *-minInd* 10 to ensure no missing genotypes. We ran *stairwayplot2* separately on folded site frequency spectra for each morph, using default recommended parameters for folded site frequency spectra. We used a mutation rate of 1.0×10^{-8}, as used in other demographic reconstructions in salmonid species (Jacobs *et al.* 2020; Rougeux *et al.* 2017) and an average generation time of four years, assuming generation time from age at first reproduction, which has been shown to predict generation time and molecular evolution rates (Lehtonen & Lanfer 2014; Tstantes & Steiper 2009). We selected estimated maturation times from data in other lake-restricted, two-morph Arctic Charr systems (Jonsson & Hindar 1982, Jonsson *et al.*1988).

Next, we generated separate genotype files for each morph using *plink*, and removed all outlier SNPs and those with any missing genotypes, resulting in a set of 2757 SNPs for piscivorous morphs and 2558 SNPs for deep-water morphs. We then used *SNeP* 1.1 (Barbato *et al.,* 2015), which estimates recombination rate from physical distance, and then compares estimated linkage between markers at differing recombination rates to infer changes in effective population size at different time periods. We ran *SNeP* using the method described by Sved (1971), with the *-svedf* option. Because calculated N_e scales with sample size in this method (Barbato *et al.,* 2015), we quantified change in N_e relative to maximum inferred N_e, by dividing each N_e estimate by the maximum value.

Outlier detection

Axiom array SNPs

We used four outlier detection methods to identify SNPs that contributed significantly to morph divergence. First, we used *pcadapt* in R (Luu *et al.* 2017) to identify SNPs with significant correlation with the first principal component, which explained a large proportion of genetic variation (21.4%), and corresponded to divergence between morphs. We also used the modified F_{ST} outlier test

based on deviation from expected allele frequencies assuming admixture between K modelled ancestral populations in the *LEA* R package (Martins *et al.* 2016), using *snmf* with $K = 2$ to estimate genetic background., as well as the *FDIST* method (Beaumont & Nichols 1996) implemented in *Arlequin* 3.5.2.1. We used the *qvalue* R package (Storey 2015) to transform SNP *p* values from outlier detection software to False Discovery Rate (FDR, Storey & Tibshirani 2003) corrected *q* values, and selected SNPs with $q < 0.05$ as outliers in each method. Lastly, we selected markers exceeding the top 99th percentile of absolute values of canonical scores as outliers in RDA conditioned on sampling year, and categorized any SNP found using at least two methods as potential outliers.

Axiom array signal intensity

To detect divergent genomic regions using signal intensity data, we selected markers in the top 99th percentile of both V_{ST} values and RDA canonical axis loadings as signal intensity outliers differentiated between morphs.

Whole genome resequencing

We identified outlier regions in our whole genome resequencing data as windows that exhibited F_{ST} and PCA scores both exceeding the 99th percentile of genome-wide values for each statistic from pooled data and individual genotype likelihoods. We used the FastPCA selection scan method (Galinsky *et al.* 2016) implemented in *PCAngsd* to scan for genomic targets of selection from individual genotype likelihoods, averaged over 1 Kbp windows using the *winscan* function in the *windowscanr* R package (Tavares 2019). To compare locations of these regions with outlier SNPs and signal intensity outliers from the Axiom array, we binned outlier regions across genotyping methods into 100 Kbp windows using *windowscanr* and then tested for overlap between outlier sets. We selected this large window of potential overlap because both Axiom SNPs and signal intensity variation are sparse relative to genomic data (>100 Kbp between each marker) and may not correspond to the peak of a highly divergent region.

Structural variant detection

Using whole genome re-sequencing data, we carried out copy number variant (CNV) detection using *CNVnator* 0.4.1 (Abyzov *et al.* 2011), which infers presence of CNVs from regions with aberrant read coverage compared to averaged genome-wide read coverage. We carried out separate CNV calling, with 1 Kbp bin sizes, on combined bam files for each morph, and retained identified putative structural variants with greater than 50% of reads mapping to a single location (q0 > 0.5), as suggested for filtering parameters in Abyzov et al. (2011). We used the *bedtools* 2.29.2 (Quinlan & Hall 2011) *intersect* function to identify genes overlapping with each putative CNV in each morph separately. We then categorized putative CNVs distinct to each morph pool as those that were found in different genes. To determine the level of genomic differentiation associated with putative CNVs distinct to each morph, we then used *bedtools merge* to compare F_{ST} in 1 Kbp windows overlapping putative CNVs from pooled F_{ST} estimates.

Gene ontology and gene set enrichment

We extracted gene ontology information for the annotated *Salvelinus* genome and tested whether specific functions were enriched in outlier regions. We extracted gene information from 3947 1 Kbp outlier windows, and compared functional enrichment of these genes with a baseline of all genes from the 39 annotated chromosomes. Gene information was extracted using *bedtools intersect*. Using the *topGO* R package (Alexa & Rahnenfuhrer 2019), we then tested for overrepresentation of biological processes among outliers relative to the total set of genes, with a node size of 5 and the 'weight01' algorithm, and used an alpha value cutoff of 0.05. As suggested in the *topGO* documentation, we did not adjust *p* values for multiple testing, as GO terms analyzed using the "weight01" algorithm include some adjustment of significance values internally, and resulting *p* values are not independent, and FDR correction may result in loss of biologically meaningful signals. Next, we repeated GO term enrichment for all genes with putative CNVs that were distinct to each morph. To minimize redundancy in identified processes, we then ordered and visualized significantly enriched functions into higher order processes using *REVIGO* (Supek et al. 2011).

To test for evidence of shared patterns of differentiation among genes in distinct biological pathways, we used a gene-set enrichment approach, as described by Daub *et al.* (2013). This method

differs from gene-ontology enrichment by using statistics from all surveyed genes to detect polygenic patterns of differentiation that may be missed by gene ontology enrichment restricted to outlier genomic regions. We used the *polysel* gene set enrichment tools, implemented in R, from Daub *et al.* (2013). As gene set information was not available for *Salvelinus*, we downloaded KEGG pathway and gene set information from *Homo sapiens* from http://www.ncbi.nlm.nih.gov/biosystems, as used by Rougeux *et al.* (2019) in another salmonid species. We then restricted our analysis to conserved genes between each group by retaining genes within sets that possessed Entrez Gene Symbols and gene descriptions matching those in the *Salvelinus* genome annotation, and a minimum of 5 genes per set, resulting in 242 gene sets with 1041 genes. We used max observed F_{ST} from each gene to test for elevated F_{ST} among specific pathways, assuming a normal distribution based on plotting of gene scores of different pathway sizes, and an alpha significance cutoff of of 0.05, as in Rougeux *et al*. 2019.

Results

Genome-wide differentiation and population structure between morphs

We found high divergence between deep-water and piscivorous morphs across all measures of population structure. Genome-wide F_{ST} for all 3677 SNPs was 0.15, and two clusters exhibiting divergence between deep-water and piscivourous morphs were also identified using PCA (Figure 1C), t-SNE, and DAPC (Supplementary Figure 1). In PCA, we found that the first principal component explained a large proportion of genetic variance between individuals (21.4%) and corresponded to separation between the deep-water and piscivorous morphs, and that subsequent principal components each explained a low proportion of variation (<2.5%). Admixture comparison using *snmf* had highest support at $K = 2$ and revealed genetic separation between morphs (Figure 1D). AMOVA identified a very small effect of year ($F_{SC} = 0.007$) relative to morph ($F_{CT} = 0.25$) in explaining genetic variation, indicating temporal genetic variation is negligible compared to variation between morphs. Observed heterozygosity was significantly higher in the piscivorous morphs than the deep-water morphs ($H_{O\ \text{Piscivorous}} = 0.263$ vs $H_{O\ \text{Deep-water}} = 0.249$, W = 7,002,800, $p = 0.0077$), as was expected heterozygosity

($H_{E\ Piscivorous}$ = 0.258 vs $H_{E\ Deep\text{-}water}$ = 0.249, W = 6,947,100, p = 0.0399). Effective population size (N_e) was also greater in piscivorous morphs (N_e = 1916.1, CI = 1276.0 - 3826.7) than in the deep-water morphs (N_e = 867.0, CI = 766.9-996.7).

We found that signal intensity based on the Axiom array also clearly delineated piscivorous and deep-water morphs. Calculation of V_{ST} revealed moderate divergence (V_{ST} = 0.028), and we found that RDA of signal intensity separated morphs on the first canonical axis (Supplementary Figure 2) and explained 3.96% of the variation in signal intensity ($p < 0.001$).

We observed clear genomic divergence between morphs (F_{ST} = 0.1165, 1000 bootstrap 95% CI = 0.1161 – 0.1169) similar to that found on the Axiom array. Population structure identified from whole genome resequencing data using *PCAngsd* similarly supported findings from the Axiom array. In PCA, we observed clear separation of deep-water and piscivorous morphs on the first PC axis, which explained much of the observed genomic variation (17.3%, Supplementary Figure 3A), whereas all subsequent PC axes explained less than 4% of variation, and showed separation by individuals. Admixture analysis also suggested K=2 genetic backgrounds, with little evidence of shared genetic variation between them (Supplementary Figure 3B).

From read counts using pool sequencing data, we genotyped 4,995,630 SNPs in 652,855 1 Kbp windows, whereas genotype likelihoods uncovered 4,829,624 SNPs in 870,856 1 Kbp windows. Correlations between PC loading (genotype likelihoods, mean 1.4x coverage per individual) and F_{ST} (pool sequencing, mean 14x coverage per pool) in 583,232 windows shared between both methods was high ($r^2 = 0.86$, $p < 1.0 \times 10^{-15}$), indicating that the same overall pattern of genomic differentiation was found across genotyping approaches, despite different cut-offs and methodologies for inferring SNPs.

We found high correlation among 1 Kbp window estimates of F_{ST} and PCA scores from genotype likelihood data but found that few individual SNPs overlapped between the Axiom array and whole genome re-sequencing datasets genotyped in the same individuals (n = 9, pooled genotypes, n = 13, genotype likelihoods). We identified 68% of 1 Kbp windows (n = 2175/3200) were shared across array and pooled resequencing data, which were strongly correlated in F_{ST} scores ($r^2 = 0.68$, $p < 10^{-15}$, Supplementary Figure 4A). Both the number of shared 1 Kbp windows (2857/3200, 89%),

and correlation of F_{ST} scores from array data and PC scores from genotype likelihoods ($r^2 = 0.77$, $p < 10^{-15}$, Supplementary Figure 4B) were higher when comparing 1 Kbp windows from array and genotype likelihood data. We also found minimal impact of removal of repetitive regions and MAF filter selection on inferred patterns of population structure (Supplementary text, Supplementary Figure 5A, 5B).

Low differentiation of mitochondrial genomes among morphs

We found that morphs did not show evidence of mitochondrial divergence (Figure 2A), with haplotypes for both the piscivorous and deep-water morphs present within the same mitochondrial lineage, except for a single deep-water morph. We did not find evidence of codon-changing substitutions within sequenced mitochondrial genomes: of 107 substitutions identified, 80 occurred in coding regions, all of which were synonymous. Restricting our analyses to the 12 segregating transversions using *FITCHI*, we found a similar pattern, with mitogenome haplotypes shared between morphs, and a single deep-water individual exhibiting a highly diverged haplotype from the larger group of both morphs. (Supplementary Figure 6).

Estimated demographic histories suggest recent divergence

The estimated maximum likelihood tree of populations identified using *treemix* uncovered a more recent split between Gander Lake morphs relative to sites across Newfoundland and Labrador (Figure 2B). Declines in estimated N_e and similar demographic histories among both morphs were observed over recent (~120 years) and more distant (~200,000 years) time periods (Figure 2C). Reconstruction of demographic history using *stairwayplot2* revealed similar patterns of post-glacial decline in estimated N_e between morphs coinciding with the end of the Last Glacial Period (LGP, ~11,000 years ago). We found that estimated N_e was high during and prior to the LGP relative to more recent estimates of N_e. Linkage-based comparison of demographic histories of each morph over the past ~120 years using *SNeP* indicated similar patterns of decline among both piscivorous and deep-water morphs (Figure 2D).

Genome-wide distribution of outlier regions

Using the Axiom array, We found 131 outlier SNPs distributed across the genome that were identified by at least two methods (Figure 3A), consistent with a multilocus basis of morph divergence. Across all methods, 199 outliers were identified.: 179 outliers were uncovered with *snmf* and 151 with *pcadapt,* with 131 being common to both methods, whereas RDA and FDIST only detected 37 and three outliers, respectively (Figure 3B). Three outlier SNPs detected across all methods were found on two linkage groups (LG14 and LG5); SNPs on LG14 were ~1Mbp apart, and exhibited almost complete fixation between morphs ($F_{ST} > 0.98$).

Using Axiom array signal intensities, we found 197 outliers (137 RDA, 137 V_{ST}), of which 77 (39.0%) outliers were shared between both methods, and distributed across multiple chromosomes (Figure 3A). Comparisons with whole genome data (Figure 3C) indicated that signal intensity outliers exhibited elevated F_{ST} (0.386), significantly beyond the genomic baseline (0.117, Mann Whitney U test $p < 1 \times 10^{-15}$).

From whole genome resequencing data, 3947 1 Kb windows showed signatures of elevated PC score and F_{ST}, which we considered potential targets of divergent selection (Figure 3A). Using F_{ST} calculated from pools, we found 6529 1 Kbp windows of elevated divergence between morphs (mean outlier $F_{ST} = 0.894$) distributed across all 39 chromosomes. Using PCA scores based on genotype likelihoods, we identified 8,709 genomic windows representing greater divergence than the genomic background. Larger windows (100 Kbp) containing shared outliers from F_{ST} and PCA (whole genome data) overlapped with many windows containing SNP outliers and signal intensity outliers from the Axiom array (Figure 3A). Of 100 Axiom SNP outlier windows, 78 (78%) overlapped with regions exhibiting signals of both elevated F_{ST} and PCA scores. Among the 76 V_{ST} outlier windows, overlap was lower than observed for SNPs, with 22 windows (29%) overlapping with 100 Kbp windows with elevated PC and F_{ST} scores.

Detection of morph-specific CNVs

We uncovered 1294 (deletion = 1290, duplication = 4, genes = 1183) putative CNVs overlapping genes in piscivorous morphs and 1175 (deletion = 1171, duplication = 4, genes = 1073)

in deep-water morphs. Of these, we found 385 putative CNVs overlapping genes distinct to piscivorous morphs (deletion = 384, duplication = 1, genes = 385), and 275 putative CNVs overlapping genes distinct to deep-water morphs (deletion = 274, duplication = 1, genes = 275, Supplementary File 2). We identified two genes overlapping with putative CNVs that were distinct to morphs that were also signal intensity outliers, with deletions in a predicted cation-independent mannose-6-phosphate receptor gene and a predicted semaphorin-6B-like gene in piscivorous morphs. Additionally, we identify deletions in two predicted opsin genes (opsin-5-like, opsin 5); these deletions were distinct to deep-water morphs. We found similar genomic divergence among putative CNVs distinct to each morph relative to the genomic background (F_{ST} = 0.111, Figure 3C).

Enrichment of biological processes in outlier regions using gene ontology and gene set enrichment

We found significant enrichment of 147 biological processes among outlier genomic regions from whole genome data (Supplementary File 3, $p < 0.05$). Of the 3947 outlier windows from whole genome data, these overlapped with 502 distinct gene products with associated GO terms. Organization of the significantly enriched GO terms into higher order processes by *REVIGO* analysis identified representative processes primarily associated with immune and metabolic processes (Figure 4A, B cell receptor signaling pathway), gene regulation and DNA repair (histone H3-K4 methylation, RNA export from nucleus), and development (oocyte growth). Gene set enrichment uncovered 14 significantly enriched KEGG pathways ($p < 0.05$, Supplementary File 4). Although these pathways were not significant after FDR correction ($q > 0.2$), we retain significant pathways at $p < 0.05$ for comparison with enriched GO processes.

GO enrichment for putative CNVs similarly identified a large number of enriched processes (n = 382, $p < 0.05$, Supplementary File 5). Of the 660 distinct genes with putative CNVs unique to each morph, 571 of these overlapped with gene products with associated GO terms. We found genes with putative CNVs were also enriched for different higher-order functions than those identified with outlier loci, primarily associated with synapse assembly (Figure 4B).

Discussion

Extreme environments provide a unique opportunity to explore the genomic basis of adaptation to challenging environmental conditions (Tobler *et al.* 2018; Wang & Guo 2019). Deep-water environments require specific adaptations at molecular, cellular, and physiological levels, but the genomic mechanisms underlying such adaptations are not well characterized. In our analysis of deep-water and piscivorous Arctic Charr morphs in Gander Lake, we found that divergence and adaptation during colonization of Gander Lake has resulted in significant population structure, and involved multiple genomic regions and categories of genetic variation (i.e. both SNPs and putative CNVs). We uncovered differentiation among genes involved in gene expression, DNA repair, and neurological and sensory function, suggesting the importance of these mechanisms in Arctic Charr morph divergence and adaptation during colonization of Gander Lake.

We found high divergence between deep-water and piscivorous morphs, revealing a genome-wide pattern of elevated differentiation between morphs, identified using both allele frequency and signal intensity information. These findings are concordant with other past estimates of divergence in this system using a small panel of microsatellites (Gomez-Uchida *et al.* 2008), and match recent observations of ecological and genomic divergence between morphs in other Arctic Charr systems (e.g. Guðbrandsson *et al.* 2019; Jacobs *et al*. 2020). Genomic differentiation observed here is greater than has been found in comparison of sympatric morphs in Arctic lakes in Norway (F_{ST} = 0.002–0.074; O'Malley *et al.* 2019), and is comparable to the higher bounds of divergence estimates across individual lakes in Labrador (Salisbury *et al.* 2020 F_{ST}: 0.08 – 0.20) and in Scotland and Siberia (F_{ST}: 0 – 0.4, Jacobs *et al.* 2020). This high degree of genomic differentiation between Gander morphs is similar to high divergence observed among abyssal morphs and other morph types identified in another deep-water lake in Norway (F_{ST}: 0.12 – 0.26, Østbye *et al.* 2020), supporting a genetic underpinning to the extensive and stable phenotypic divergence of deep-water morphs relative to other Arctic Charr morph systems (Klemetsen 2010).

Analysis of mitochondrial genomes and evolutionary history of Gander Lake morphs revealed local divergence and multiple colonizing lineages following deglaciation. In contrast to genomic divergence, mitochondrial whole genome analysis identified predominantly shared haplotypes between deep-water and piscivorous morphs matching a recent observation of shared D-loop

haplotypes corresponding almost exclusively to a single glacial refuge (Salisbury *et al.* 2019). Interestingly, we also found one deep-water individual that carried a highly diverged mitochondrial haplotype, reinforcing the observation that at least two lineages have colonized Gander Lake. Reconstruction of divergence from maximum likelihood trees of Arctic Charr populations in Newfoundland and Labrador revealed clear separation of Gander Lake morphs from other sites, suggesting local divergence following colonization.

We further explored the demographic history of each morph using site-frequency and linkage-based demographic history reconstruction methods, and found similar patterns of reduction in diversity in each morph following periods with glacial retreat, consistent with colonization of Gander Lake after the end of the Last Glacial Period (LGP, ~11 kya). Demographic history inference using coalescent methods may be affected by population substructure and migration between sampled demes, in addition to changes in census size reflected in N_e (Mazet *et al.* 2016; Orozco-terWengel 2016). Here, we found decreases in estimated N_e following the LGP, consistent with emergence of genetic sub-structuring between morphs following postglacial colonization. The timing of these estimates may be sensitive to specified generation times based on maturation rates, which exhibit considerable variation across Arctic Charr populations (Jonsson & Jonsson 2001). More recent ongoing declines in estimated Ne were also observed using an LD-based approach, consistent with recent identification of genomic vulnerability of southern Arctic Charr populations (Layton *et al.* 2021). The estimated demographic and evolutionary history, as well as the restricted occurrence of phenotypically similar deep-water morphs to deep-water habitats across the species range, which have not been identified in other lakes in Newfoundland (Klemetsen 2010), suggests some component of genomic divergence is likely a result of adaption *in situ*. Future investigations of local sources of ancestry across the genome (e.g. Foote *et al.* 2019) may provide further insight into the role of ancestral and post-colonization divergence in the emergence of piscivorous and deep-water morphs.

Putatively-adaptive loci were found distributed across multiple regions of several chromosomes. Models of local adaptation and early speciation with gene flow predict a subset of linked genomic regions may be selected and diverge against the genomic background (Feder *et al.* 2012; Via 2012; Wu 2001; Yeaman & Whitlock 2011), with transitions to more extensive genomic

divergence dependent on the strength of, and number of traits under selection (Feder *et al.* 2012). Secondary contact of allopatric populations may also produce heterogeneous patterns of divergence, with gene-flow resistant genomic regions displaying high divergence, and these genomic regions have been associated with ecological adaptation across other study systems (Feder *et al.* 2013, Scordato *et al.* 2017; Vijay *et al.* 2016). Our results from analyses of demography and mitochondrial genome divergence suggests both *in situ* divergence and contact events; in either of these scenarios outlier genomic regions are likely targets of selection during adaption to variable depth environments in Gander Lake. However, we also acknowledge that variable rates of recombination and linked selection may also play a role in shaping genomic divergence observed here (Burri *et al.* 201; Burri 2017). Our identification of a genome-wide distribution of outlier loci matches similar results from other recent investigations into genomic divergence across morphs that have also found outlier loci on several chromosomes (e.g. Guðbrandsson *et al.* 2019; Jacobs et al. 2020). Genomic parallelism between morph systems has generally been low, suggesting many potential targets of selection during local adaptation in Arctic Charr (Jacobs *et al.* 2020; Salisbury *et al.* 2020). We found that genes in highly divergent regions of the genome are overrepresented for functions with roles in DNA repair and gene regulation, (representative processes: histone H3-K4 methylation, RNA export from nucleus), and membrane function (lipoprotein biosynthesis). Gene set enrichment also uncovered the strongest signs of enrichment among KEGG pathways associated with similar biological processes including DNA repair (purine metabolism, e.g. Zhou *et al.* 2010), cardiac function (e.g., Arrhythmogenic right ventricular cardiomyopathy), and membrane function (SNARE interactions in vesicular transport, Kavalali 2002). Processes identified in gene-set enrichment were not significant after multiple test correction, which may highlight the challenge of identifying strong signals of polygenic adaptation in specific pathways, characterized by small shifts in allele frequency (Le Corre & Kremer 2012), when background genomic divergence is relatively high. Deep-water environments exhibit elevated hydrostatic pressure (Pradillon & Gaill 2006; Somero 1992), low light (Hahn *et al.* 2018), and reduced temperature (Brown & Thatje 2014). These environmental features can impact membrane integrity through reducing membrane fluidity, protein and enzyme activity through altered reaction kinetics, and nucleic acid stability and replication through alterations in nucleic acid

structures (Brown & Thatje 2014; Pradillon & Gaill 2006; Somero 1992, Yancey 2020). Enriched biological processes observed here are consistent with these environmental pressures, although developmental, dietary, and behavioural differences between morphs may also contribute to identified gene ontology enrichments.

We also detected a genome-wide distribution of putative CNVs distinct to each morph. Recent studies of ecological adaptation in aquatic species have uncovered a role for copy number variants in thermal adaptation (Dorant *et al.* 2020), and have been associated with divergence between marine and freshwater stickleback ecotypes (Lowe *et al.* 2018). We found enrichment for functions associated with synapse formation, which may underlie some combination of behavioural or sensory differences between morphs, and maintenance of neuronal function, which is impacted by temperature and pressure at high depths through disruption of membrane activity (Pradillon & Gaill 2006; Yancey 2020; Somero 1992). Consistent with changes to sensory system changes between morphs, we identified deletions in two predicted opsin genes, matching similar findings from other recent studies uncovering visual system gene loss in deep-sea and cave fishes (Wang *et al.* 2019; Wilkens 2020). We found similar patterns of genomic divergence of putative CNVs relative to the genomic background, suggesting polygenic divergence as opposed to high differentiation among putative CNVs. This finding matches the observation of polygenic divergence of CNVs in recently diverged cichlid species, and highlights the importance of additional characterization of structural variation to better understand the genomic bases of adaptation (McGee *et al.* 2020; Mérot *et al.* 2020). We found low overall correspondence between signal intensity outliers with putative CNVs detected using whole genome data, which may be a result of re-sequencing only a small sample of morphs and missing relevant CNV loci, or additional sequence variants at signal intensity outlier sites not detectable with *CNVnator* contributing to morph divergence. Future characterization of genomic variation in Arctic Charr using long read sequencing may also allow detection of structural variants including CNVs, as well as genomic regions differentiated by signal intensity variation, and advance understanding of the genomic basis of adaptation in Arctic Charr morphs.

There remain limitations in methods employed in this study that may introduce some uncertainty in patterns observed here. Low sample sizes with pooled whole genome re-sequencing

data have been shown to exhibit bias in F_{ST} (Anderson *et al.* 2014), and ascertainment bias on SNP arrays has also been shown to influence F_{ST} estimates unpredictably, with the influence of ascertainment unique to each panel and surveyed population (Bhatia *et al.* 2013). To address limitations of pool-sequencing F_{ST} bias, we also used methods based directly on genotype likelihoods (Korneliussen *et al.* 2014; Therkildsen & Palumbi 2017) to account for uncertainty in called genotypes from low coverage data, and find highly correlated patterns of divergence among whole genome F_{ST} and genotype-likelihood PCA scores. Comparison of the larger sample of individuals from the Axiom Array and the whole genome re-sequencing data revealed correlated patterns of divergence among 1 Kbp windows, and high overlap of outlier regions, although we found few individual SNPs shared between methods. Observation of few shared SNPs but correlated patterns across genomic regions is not necessarily surprising, as low sequencing coverage at individual SNPs and filtering parameters employed for whole genome re-sequencing methods may prevent identifying individual SNPs genotyped at high coverage on the Axiom array. Our identification of similar patterns of population structure and regions of elevated divergence suggests that ascertainment of array data and low sample size in resequencing data have not precluded identifying the genomic landscape of differentiation between morphs in this study. However, as sequencing continues to decrease in cost, further investigation of the genomic basis of depth adaptation in Arctic Charr morphs would benefit from additional population sampling and increased per-individual sequencing coverage to confirm genomic patterns identified here.

Detection of structural variants also remains challenging, especially in non-model systems (e.g. salmonids, Bertolotti *et al.* 2020), and our limited sample size and sequencing coverage per sample, and inherent limitations of bioinformatic predictions of CNVs from short read sequence data may prevent detection of all relevant structural variation in Arctic Charr morphs. Analysis by Lowe *et al.* (2018) of pooled samples with similar sample size and coverage indicate *CNVnator* exhibits high sensitivity to detect deletions, and recent comparisons suggest low coverage whole genome resequenecing outperforms SNP array signal intensity for CNV detection (Zhou *et al.* 2018). However, comparisons of different structural variant detection methods have found considerable differences among sensitivity of different pipelines dependent on variant type and sequencing approach,

indicating different types of structural variants may be more difficult to detect with methods used here (Zhang *et al.* 2019). *CNVnator* has shown high detection capacity for deletions across comparisons with other algorithms, but depth-based algorithms including CNVnator have shown low sensitivity for duplications relative to deletions (Lowe *et al.* 2018; Trost *et al.* 2018; Zhou *et al.* 2018), providing a potential explanation of the larger number of detected deletions relative to duplications in this study. Our investigation of CNVs and signal intensity variation may also be improved by additional sequencing coverage, as well as long read sequencing to more accurately characterize structural variants and other genomic regions that are difficult to survey with short read sequencing.

Conclusion

The extreme deep-water environment and recent deglaciation history of Gander Lake provides a unique opportunity to investigate the genomic basis of morph divergence and adaptation. Our results suggest multiple types of genetic variation (i.e. both SNPs and putative CNVs) and molecular processes underlie ecological divergence in this system and have facilitated adaptation of Arctic Charr morphs after colonization of Gander Lake. These results suggest morph divergence and adaptation has occurred through alterations in biological processes involved in gene expression, DNA repair, membrane function, heart development, and synapse assembly. As genomic resources are adopted in the study of ecological adaptation in other extremophile species, comparison with genes and processes identified here may aid in identifying common substrates for adaptation across extreme environments. Overall, our work highlights the need to consider the diverse array of molecular solutions that can evolve in response to extreme environments when studying the genomic basis of adaptation.

Acknowledgments

Thanks to Parks Canada, the Nunatsiavut Government, the NunatuKavut Community Council, the Sivunivut Inuit Community Corporation, the Innu Nation, the Labrador Hunting and Fishing Association, and Newfoundland DFO Salmonids section, for support of this study and tissue collection. Significant portions of lab work for genotyping were conducted by staff at Aquatic Biotechnology Lab at the Bedford Institute of Oceanography. Computing resources for bioinformatic

analyses were provided by Compute Canada. Support for this study was provided by the Ocean Frontier Institute, a Genomics Research and Development Initiative (GRDI) Grant, a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant and Strategic Project Grant to I.R.B., the Weston Family Award for research at the Torngat Mountains Base Camp and an Atlantic Canada Opportunities Agency and Department of Tourism, Culture, Industry and Innovation grant allocated to the Labrador Institute.

Data availability

Raw reads used in this study are available on the NCBI short read archive (Accession PRJNA626339). Genomic data from the Axiom SNP array used in this study is available at: https://doi.org/10.5061/dryad.gf1vhhmkh

Scripts used in this study can be found at: https://github.com/TonyKess/Gander_charr

Figures

Figure 1. (A) Map of Gander Lake in Newfoundland, Canada, where (B) piscivorous morph and deep-water morphs of *Salvelinus alpinus* Arctic Charr occur in sympatry. (C), genomic divergence between morphs inferred from principal component scores (PC) from PC analysis and (D) ancestry proportions from sparse non-negative matrix factorization of 2430 putatively neutral, unlinked Axiom array SNPs

Figure 2. Morph demographic histories suggesting postglacial local divergence inferred from (A) Haplotype network from mitochondrial whole genomes, (B) treemix maximum likelihood trees visualizing separation of Arctic Charr populations and morphs (C) demographic history estimated from the site frequency spectrum using stairwayplot2 and (D) estimated from linkage disequilibrium using SNeP.

Figure 3. Genome wide distribution of outlier regions identified from smoothed genome-wide distribution of F_{ST} calculated in 1 Kbp windows between deep-water and piscivorous morphs and outlier regions identified as exceeding the 99th percentile for both F_{ST} and PCA scores (blue), Axiom array outlier SNPs (red), and V_{ST} of Axiom array signal intensity outliers (orange). (B) Venn diagram of overlap of Axiom array outlier SNPs identified across detection methods, and (C) comparison of F_{ST} across the genome with outliers and putative copy number variants distinct to each morph.

Figure 4. Enriched gene ontology terms for A) 502 gene products in 3947 1 Kb outlier windows from whole genome data, and B) 571 gene products with associated GO terms from putative copy number variants distinct to each morph.

References

Abyzov, A., Urban, A.E., Snyder, M., Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research*, **21(6)**, 974-984.

Alexa, A., & Rahnenfuhrer, J. (2019). topGO: Enrichment Analysis for Gene Ontology. R package version 2.36.0.

Anderson, E.C., Skaug, H.J., Barshis, D.J. (2014). Next-generation sequencing for molecular ecology: a caveat regarding pooled samples. *Molecular Ecology*, **23(3)**, 502-512.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Barbato, M., Orozco-terWengel, P., Tapio, M. & Bruford, M.W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, **6**, 109.

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

Bertolotti, A.C., Layer, R.M., Gundappa, M.K. *et al.* (2020). The structural variation landscape in 492 Atlantic salmon genomes. *Nature Communications,* **11,** 5176. doi.org/10.1038/s41467-020-18972-x

Bhatia, G., Patterson, N., Sankararaman, S., Price, A.L. (2013). Estimating and interpreting FST: the impact of rare variants. *Genome Research,* **23(9)**, 1514-1521.

Boughman, J.W.(2001). Divergent sexual selection enhances reproductive solation in sticklebacks. *Nature*, **411**, 944-948.

Brown, A., & Thatje, S. (2014). Explaining bathymetric diversity patterns in marine benthic invertebrates and demersal fishes: physiological contributions to adaptation of life at depth. *Biological Reviews*, **89(2)**, 406-426.

Brunner, P.C., Douglas, M.R., Osinov, A., Wilson, C.C., Bernatchez, L.(2001). Holarctic phylogeography of Arctic Charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution,* **55**, 573-586.

Burri R. (2017). Dissecting differentiation landscapes: a linked selection's perspective. *Journal of Evolutionary Biology*, **30(8)**, 1501–1505.

Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L. Suh, A., *et al.* (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research*, **25**, 1656–1665.

Cantsilieris, S., Baird, P.N., & White, S.J. (2013). Molecular methods for genotyping complex copy number polymorphisms. *Genomics*, **101(2)**, 86-93.

Chang, C., Chow, C., Tellier, L., Vattikuti, S., Purcell, S. M., Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.

Cossins A.R., & MacDonald, A.G. (1984). Homeoviscous theory under pressure: II. The molecular order of membranes from deep-sea fish. Biochimica et Biophysica Acta (BBA)-Biomembranes, 776(1),144-150.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, **27,** 2156-2158.

Daub, J.T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., & Excoffier, L. (2013). Evidence for polygenic adaptation to pathogens in the human genome. *Molecular Biology and Evolution*, **30(7)**, 1544-1558.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., McKenna, A. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43(5)**, 491.

Do, C., Waples, R.S. , Peel, D., Macbeth, G.M., Tillett, B.J., Ovenden, J.R. (2014). NeEstimator v2.0: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, **41**, 209–214.

Dobzhansky, T. (1951). Genetics and the Origin of Species. Columbia University Press, New York, NY.

Doiron, S., Bernatchez, L., Blier, P.U. (2002). A comparative mitogenomic analysis of the potential adaptive value of arctic Charr mtDNA introgression in brook Charr populations (*Salvelinus fontinalis* Mitchill). *Molecular Biology and Evolution*, **19(11)**, 1902-1909.

Dorant, Y., Cayuela,, H., Wellband, K., et al. (2020). Copy number variants outperform SNPs to reveal genotype–temperature association in a marine species. *Molecular Ecology*, **29**, 4765– 4782.

Endler, J.A.. (1992) Signals, signal conditions, and the direction of evolution. *The American Naturalist*, **139**, 125-153.

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

Excoffier, L., Smouse, P.E., & Quattro, J.M.(1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 179–191.

Feder, J.L., Egan, S.P., Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28(7)**, 342-50.

Feder, J.L., Flaxman, S.M., Egan, S.P., Comeault, A.A., Nosil, P. (2013). Geographic mode of speciation and genomic divergence. Annual Review of Ecology, *Evolution, and Systematics*, **44**, 73–97.

Foote, A.D., Martin, M.D., Louis, M., et al. (2019). Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Molecular Ecology*, **28(14)**, 3427-3444.

Frichot, E. & Francois, O. (2015). LEA: An R Package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**, 925–929.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, **196**, 973–983.

Galinsky, K.J., Bhatia, G., Loh, P.R., Georgiev, S., Mukherjee, S., Patterson, N.J., Price, A.L. (2016). Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *The American Journal of Human Genetics*, **98(3)**, 456-472.

Gomez-Uchida, D., Dunphy, K.P., O'Connell, M.F., Ruzzante, D. E. (2008). Genetic divergence between sympatric Arctic Charr *Salvelinus alpinus* morphs in Gander Lake, Newfoundland: Roles of migration, mutation and unequal effective population sizes. *Journal of Fish Biology*, **73(8)**, 2040–2057.

Guðbrandsson, J., Kapralova, K.H., Franzdóttir, S.R., Bergsveinsdóttir, T.M., Hafstað, V., Jónsson, Z.O., Snorrason, S.S., Pálsson, A. (2019) Extensive genetic divergence between recently evolved sympatric Arctic Charr morphs. *Ecology and Evolution,* **19**,10964-10983

Hahn, C., Genner, M.J., Turner, G.F., Joyce, D.A.. (2017). The genomic basis of cichlid fish adaptation within the deepwater "twilight zone" of Lake Malawi. *Evolution Letters*, **1**, 184–198.

Hill, J., Enbody, E.D., Pettersson, M.E., Sprehn, C.G., Bekkevold, D., Folkvord, A., Laikre, L., Kleinau, G., Scheerer, P., Andersson, L. (2019). Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proceedings of the National Academy of Sciences*, **116(37)**,18473-18478.

Huerta-Sánchez, E., Jin, X., Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, **512(7513)**,194-197.

Ilardo, M., & Nielsen, R. (2018). Human adaptation to extreme environmental conditions. *Current Opinion in Genetics & Development*, **53,** 77-82.

Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N.V, Alekseyev, S., Hooker, O., Leong, J., Minkley, D.R., Rondeau, E., Koop, B., Adams, C., Elmer, K.R. (2020) Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. *PLoS Genetics*, **16(4)**: e1008658.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

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

Jonsson, B., & Jonsson, N. (2001). Polymorphism and speciation in Arctic Charr. *Journal of Fish Biology*, **58**, 605-638.

Jonsson, B., & Hindar, K.. (1982). Reproductive strategy of dwarf and normal Arctic charr (Salvelinus alpinus) from Vangsvatnet Lake, western Norway. *Canadian Journal of Fisheries and Aquatic Sciences*, **39(10),** 1404-13.

Jonsson, B., Skúlason, B., Snorrason, S.S., Sandlund, O.T., Malmquist, H.J., Jónasson, P.M., Gydemo, R. & Linden, T. (1988). Life history variation of polymorphic Arctic Charr (*Salvelinus alpinus*) in Thingvallavatn, Iceland. *Canadian Journal of Fishery and Aquatic Sciences*, **45**,1537-1547.

Kavalali, E.T. (2002). SNARE interactions in membrane trafficking: a perspective from mammalian central synapses. *Bioessays*, **24(10)**, 926-936.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., Prodöhl, P.A. (2013). diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, **4(8)**, 782-788.

Klemetsen, A. (2010). The Charr problem revisited: exceptional phenotypic plasticity promotes ecological speciation in postglacial lakes. *Freshw. Rev*., **3**, 49–74.

Klemetsen, A. (2013) The most variable vertebrate on earth. *Journal of Ichthyology*, **53**, 781-791.

Knudsen, R., Johnsen, H., Sæther, B.S. Siikavuopio, S.I. (2015). Divergent growth patterns between juveniles of two sympatric Arctic Charr morphs with contrasting depth gradient niche preferences. *Aquatic Ecology*, **49**, 33–42.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R.V., Nolte, V., Futschik, A., Kosiol, C., Schlötterer, C. (2011). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE,* **6(1)**, e15925.

Korneliussen, T.S., Albrechtsen, A., Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, **15(1)**, 356.

Krijthe, J.H. (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation, URL: https://github.com/jkrijthe/Rtsne

Layton, K.K.S., Snelgrove, P.V.R., Dempson, J.B. et al. (2021). Genomic evidence of past and future climate-linked loss in a migratory Arctic fish. *Nat. Clim. Chang.,***11**, 158–165 (2021).

Lehtonen, J., & Lanfear, R. (2014). Generation time, life history and the substitution rate of neutral mutations. *Biology Letters*, **10(11)**, 20140801.

Le Corre. V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21(7)**, 1548-1566.

Leigh, J.W., & Bryant, D. (2015). PopART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6(9)**, 1110-1116.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, **1303**, 3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics.* **25(16)**, 2078-2079.

Linck, E., & Battey, C.J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, **19(3),** 639-647.

Lischer, H.E.L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28(2)**, 298-299.

Liu, X., & Fu, Y.X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology,* **21**, 280. https://doi.org/10.1186/s13059-020-02196-9

Liu, S., Lorenzen, E.D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., Zhou, L., Korneliussen, T.S., Somel, M., Babbitt, C., Wray, G. (2014). Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*, **157(4)**, 785-794.

Longmire, J. L., Maltbie, M., Baker, R. J. (1997). Use of "lysis buffer" in DNA isolation and itsimplications for museum collections. *Occasional Papers, The Museum of Texas Tech University*, **163**, 1–3.

Lou, H., Lu, Y., Lu, D., Fu, R., Wang, X., Feng, Q., Wu, S., Yang, Y., Li, S., Kang, L., Guan, Y. (2015). A 3.4-kb copy-number deletion near EPAS1 is significantly enriched in high-altitude Tibetans but absent from the Denisovan sequence. *The American Journal of Human Genetics*, **97(1)**, 54-66.

Lowe, C.B., Sanchez-Luege, N., Howes, T.R., Brady, S.D., Daugherty, R.R., Jones, F.C., Bell, M.A., Kingsley, D.M. (2018). Detecting differential copy number variation between groups of samples. *Genome Research*, **28(2)**, 256-265.

Luu, K., Bazin, E., & Blum, M.G.B. (2017). pcadapt: An R package for performing genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **17**, 67–77.

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., Durbin, R. (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, **2(12)**,1940-1955.

Matschiner, M. (2016) Fitchi: haplotype genealogy graphs based on the Fitch algorithm. *Bioinformatics*, **232(8)**, 1250-1252.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, **17(1)**,10-12.

Martins, H., Caye, K., Luu, K., Blum, M.G.B., François, O. (2016). Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. *Molecular Ecology*, **25(20)**, 5029–5042.

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., Muños, S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques,* **61(4)**, 203-205.

May-McNally, S.L., Quinn, T.P., Woods, P.J. & Taylor, E.B. (2015). Evidence for genetic distinction among sympatric ecotypes of Arctic char (*Salvelinus alpinus*) in south-western Alaskan lakes. *Ecology of Freshwater Fish,* **24**, 562–574.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., Chikhi, L. (2016). On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity*, **116(4)**, 362-71.

Mérot C., Oomen, R.A., Tigano, A., Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, **35(7)**, 561-572.

Meisner, J., Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, **210(2)**, 719-731.

McGee, M.D., Borstein, S.R., Meier, J.I. et al. (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature,* **586**, 75–79.

Michaud, W., Power, M., & Kinnison, M. (2008). Trophically mediated divergence of Arctic Charr (Salvelinus alpinus L.) populations in contemporary time. *Evolutionary Ecology Research* **10**, 1051-1066.

Moore, J.S., Bajno, R., Reist, J.D., Taylor, E.B. (2015). Post-glacial recolonization of the North American Arctic by Arctic char (*Salvelinus alpinus*): genetic evidence of multiple northern refugia and hybridization between glacial lineages. *Journal of Biogeography*, **42(11)**, 2089-2100.

Myles, S., Mahanil, S., Harriman, J., Gardner, K.M., Franklin, J.L., Reisch, B.I., Ramming, D.W., Owens, C.L., Li, L., Buckler, E.S., Cadle-Davidson, L. (2015). Genetic mapping in grapevine using SNP microarray intensity values. *Molecular Breeding*, **35(3)**, 88

Nagai, H., Terai, Y., Sugawara, T., Imai, H., Nishihara, H., Hori, M., Okada, N. (2011). Reverse evolution in RH1 for adaptation of cichlids to water depth in Lake Tanganyika. *Molecular Biology and Evolution*, **28(6)**, 1769–1776.

Nugent ,C.M., Leong, J.S., Christensen, K.A., Rondeau, E.B., Brachmann, M.K., Easton, A.A., et al. (2019) Design and characterization of an 87k SNP genotyping array for Arctic Charr (Salvelinus alpinus). *PLoS ONE*, **14(4)**, e0215008.

O'Connell, M.F., & Dempson, J.B. (2002) The biology of Arctic Charr, *Salvelinus alpinus*, of Gander Lake, a large, deep, oligotrophic lake in Newfoundland, Canada. *Env Biol Fish*, **64**,115–126.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H., Wagner, H. (2011) vegan: community ecology package. R package version 1.17-4. URL http://CRAN. R-project. org/package= vegan. 2010 Mar 17.

Østbye, K., Hagen Hassve, M., Peris Tamayo, A.M., Hagenlund, M., Vogler, T., Præbel K. (2020). And if you gaze long into an abyss, the abyss gazes also into thee": four morphs of Arctic charr adapting to a depth gradient in Lake Tinnsjøen. *Evolutionary Applications*, **13(6)**, 1240-1261.

O'Malley, K.G., Vaux, F., Black, A.N. (2019). Characterizing neutral and adaptive genomic differentiation in a changing climate: The most northerly freshwater fish as a model. *Ecology and Evolution*, **9(4)**. 2004-2017.

Orozco-terWengel, P. (2016). The devil is in the details: the effect of population structure on demographic inference. *Heredity*, **116(4)**, 349.

Pedersen, B.S., & Quinlan, A.R. (2018) Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics,* **34(5),** 867-8.

Peiffer, D.A., Le, J.M., Steemers, F.J., Chang, W., Jenniges, T., Garcia, F., Haden, K., Li, J., Shaw, C.A., Belmont, J., Cheung, S.W., Shen, R.M., Barker, D.L., Gunderson, K.L. (2006). High-resolution genomic profiling of chromosomal aberrations using infinium whole-genome genotyping. *Genome Research*, **16**, 1136–1148

Pickrell, J.K., Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, **8(11)**, 1-7.

Power, G. (2002). Charrs, glaciations and seasonal ice. Ecology, behaviour and conservation of the Charrs, genus *Salvelinus*. 17–35. Dordrecht, the Netherlands: Springer.

Power, M., O'Connell, M.F., Dempson, J.B.(2005). Ecological segregation within and among Arctic char morphotypes in Gander Lake, Newfoundland. *Environmental Biology of Fishes*, **73(3)**, 263–274.

Power, M., O'Connell, M.F., & Dempson, J.B. (2012). Determining the consistency of thermal habitat segregation within and among Arctic Charr morphotypes in Gander Lake, Newfoundland. *Ecology of Freshwater Fish*, **21**, 245–254.

Pradillon, F., & Gaill, F. (2006). Pressure and life: some biological strategies. *Rev. Environ. Sci. Biotechnol*., **6**, 181-195.

Quinlan, A.R., & Hall I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research. Sankhyā: *The Indian Journal of Statistics*, *Series A*, **1**, 329-358.

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. (2006). Global variation in copy number in the human genome. *Nature*, **444(7118)**, 444.

Reist, J.D., Power, M. & Dempson, J.B. (2013) Arctic charr (*Salvelinus alpinus*): a case study of the importance of understanding biodiversity and taxonomic issues in northern fishes. *Biodiversity*, **14**, 45–56.

Richards, E.J., Poelstra, J.W., Martin, C.H. (2018). Don't throw out the sympatric speciation with the crater lake water: fine-scale investigation of introgression provides equivocal support for causal role of secondary gene flow in one of the clearest examples of sympatric speciation. *Evolution Letters*, **2(5)**, 524-540.

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.

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.

Rougeux C., Bernatchez, L.,& Gagnaire, P.A.. Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, **9(8)**, 2057-2074.

Rothschild, L.J., & Mancinelli, R.L. (2001). Life in extreme environments. *Nature,* **409(6823)**, 1092-1101.

Salisbury, S.J., Booker, C., McCracken, G.R., Keefe, D., Knight, T., Perry, R., & Ruzzante, D.E. (2018). Genetic divergence among and within landlocked and anadromous Arctic char (Salvelinus alpinus) populations in Labrador, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **75(8)**, 1256-1269.

Salisbury, S.J., McCracken, G.R., Keefe, D., Perry, R., Ruzzante, D.E. (2019). Extensive secondary contact among three glacial lineages of Arctic Char (*Salvelinus alpinus*) in Labrador and Newfoundland. *Ecology and Evolution*, **9(4)**, 2031-2045.

Salisbury, S.J., McCracken, G.R., Perry, R., Keefe, D., Laton, K.K.S., Kess, T., Nugent, C.M., Leong, J.S., Bradbury, I. R., Koop, B.F., Ferguson, M.M., Ruzzante, D.E. 2020. Limited genetic parallelism underlies recent, repeated incipient speciation in geographically proximate populations of an Arctic fish (*Salvelinus alpinus*). *Molecular Ecology,* **29(22)**, 4280-4294.

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

Scordato, E.S., Wilkins, M.R., Semenov, G., Rubtsov, A.S., Kane, N.C., Safran, R.J. (2017). Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Molecular Ecology*, **26(20)**, 5676-91.

Sébert, P. (1997). Pressure effects on shallow-water fishes. *Fish Physiology*, **16**, 279-323.

Seehausen, O., Terai, Y., Magalhaes, I.S., Carleton, K.L., Mrosso, H.D.J, Miyagi, R., van der Sluijs, I., Schneider, M.V., Maan, M.E., Tachida, H., et al. (2008). Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620-626.

Shaw, J., Piper, D.J. W., Fader, G.B.J., King, E.L., Todd, B.J., Bell, T., … Liverman, D.G.E. (2006). A conceptual model of the deglaciation of Atlantic Canada. *Quaternary Science Reviews*, **25(17–18)**, 2059–2081.

Skúlason, S., & Smith, T.B. (1995). Resource polymorphism in vertebrates. *Trends in Ecology and Evolution*, **10**, 366-370.

Somero, G.N. (1992). Adaptations to high hydrostatic pressure. *Annu. Rev. Physiol*, **54**, 557–577.

Stanley, R.R., Jeffery, N.W., Wringe, B.F., DiBacco, C., & Bradbury, I.R. (2016). genepopedit: A simple and flexible tool for manipulating multilocus molecular data in R. *Molecular Ecology Resources*, **17(1)**, 12-18.

Stecher, G., Tamura, K. & Kumar, S. (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS, *Molecular Biology and Evolution*, **37(4)**, 1237–1239.

Storey, J.D., Bass, A.J., Dabney, A., Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. R package version 2.10.0.http://github.com/jdstorey/qvalue

Storey J.D., & Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* , **100**, 9440–9445.

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, **6(7)**, e21800.

Sved, J.A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, **2(2)**, 125-141.

Tavares, H. (2019). windowscanr: Apply functions using sliding windows. R package version 0.1. https://github.com/tavareshugo/WindowScanR

Therkildsen, N.O., Palumbi, S.R. (2017). Practical low-coverage genome-wide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in non-model species. *Molecular Ecology Resources*, **17(2)**, 194-208.

Tobler, M., Kelley, J.L., Plath, M., Riesch, R. (2018). Extreme environments and the origins of biodiversity: adaptation and speciation in sulphide spring fishes. *Molecular Ecology*, **27(4)**, 843-859.

Trost, B., Walker, S., Wang, Z., Thiruvahindrapuram, B., *et al.* (2018) A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. *The American Journal of Human Genetics,* **102(1)**, 142-55.

Tsantes, C., & Steiper, M.E. (2009). Age at first reproduction explains rate variation in the strepsirrhine molecular clock. *Proceedings of the National Academy of Sciences*, **106(43)**, 18165-18170.

van der Maaten, L.J.P. & Hinton, G.E. (2008). Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, **9**, 2579-2605.

Via, S. (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367(1587)**, 451-460.

Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P., Wolf, J.B. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, **7**, 13195

Wang, Y., & Guo, B. (2019) Adaption to extreme environments: a perspective from fish genomics. *Reviews in Fish Biology and Fisheries*, **29**, 735–747.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H., Bucan M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research*, **17(11)**, 1665-1674.

Wang, K., Shen, Y., Yang, Y., Gan, X., Liu, G., Hu, K., Li, Y., Gao, Z., Zhu, L., Yan, G., He, L.. (2019) Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation. *Nature Ecology & Evolution*, **3(5)**, 823.

Weir, B.S., & Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wilkens, H. (2020) The role of selection in the evolution of blindness in cave fish. *Biological Journal of the Linnean Society*. **130(3)**, 421-432.

Woods, P.J., Young, D., Skúlason, S., Snorrason, S.S. & Quinn, T.P. (2013). Resource polymorphism and diversity of Arctic Charr *Salvelinus alpinus* in a series of isolated lakes. *Journal of Fish Biology,* **82**, 569– 587.

Wu, C.I. (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology,* **14(6)**, 851-65.

Yancey, P.H. (2020). Cellular responses in marine animals to hydrostatic pressure. *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, **333(6)**, 398-420.

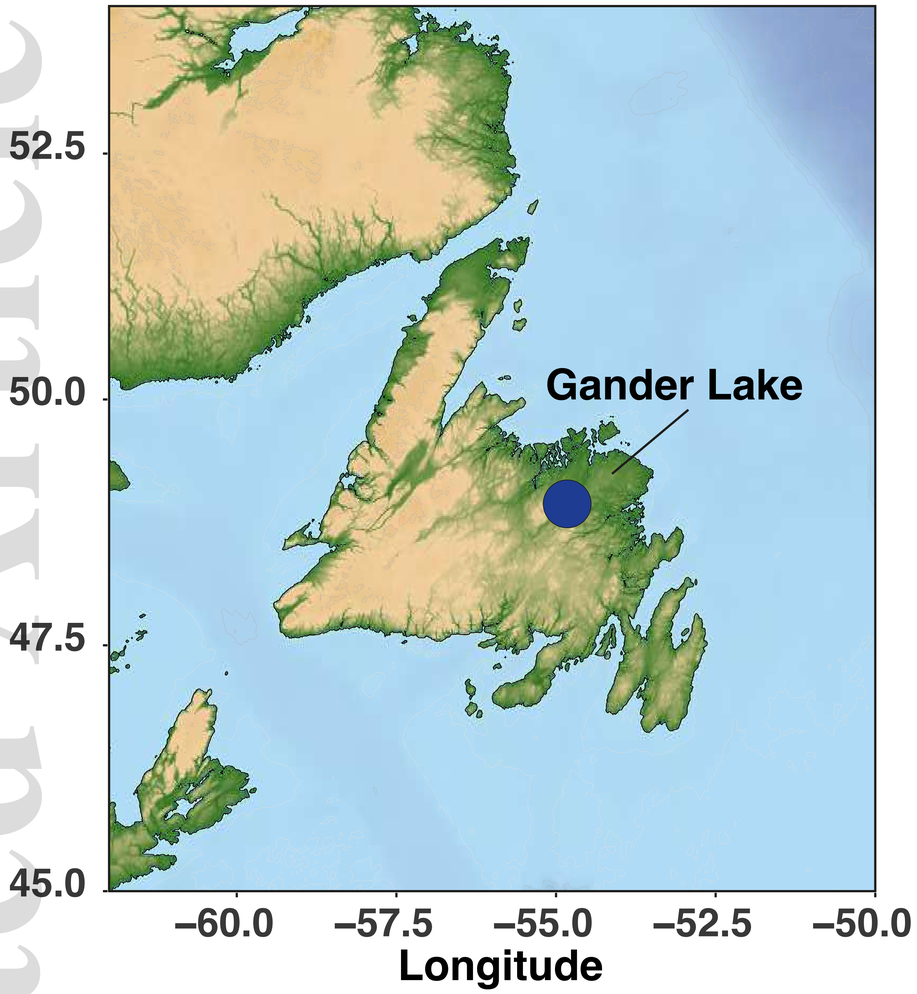
Yeaman, S., & Whitlock, M.C. (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution,* **65(7)**,1897-1911.

Zhang, L., Bai, W., Yuan, N., Du, Z. (2019). Comprehensively benchmarking applications for detecting copy number variation. *PLoS Computational Biology*, **15(5)**, e1007069.

Zhou, W., Yao, Y., Scott, A.J., Wilder-Romans, K., Dresser, J.J., Werner, C.K., Sun, H., Pratt, D., Sajjakulnukit, P., Zhao, S.G., Davis, M., et al. (2010). Purine metabolism regulates DNA repair and therapy resistance in glioblastoma. *Nature Communications*, **11**, 3811.

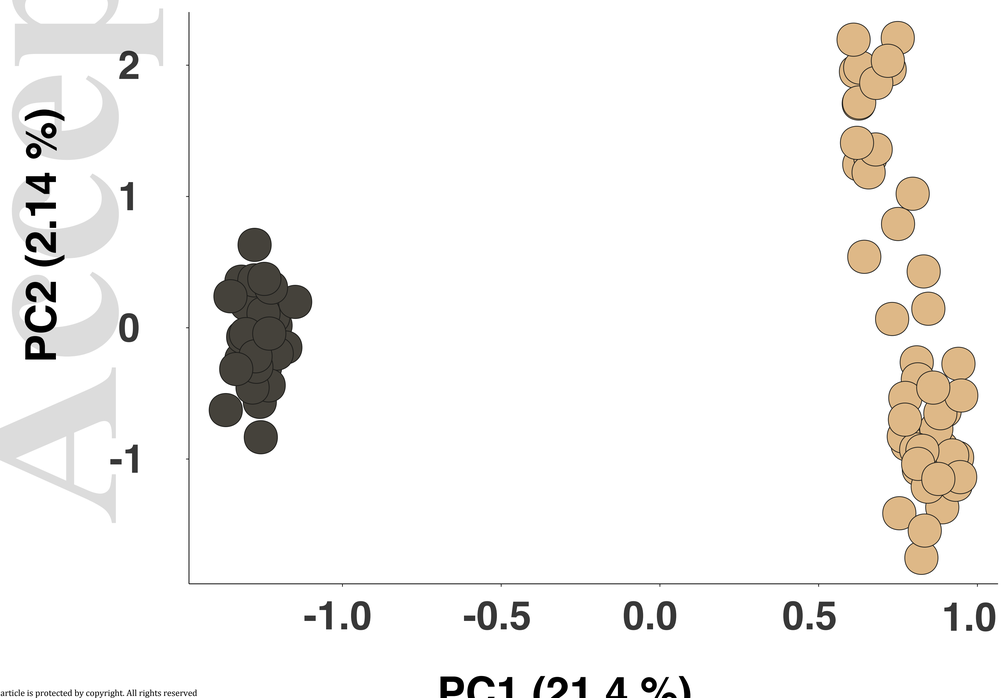
Zhou, B., Ho, S.S., Zhang, X., Pattni, R., Haraksingh, R.R., Urban, A.E. (2018). Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. *Journal of Medical Genetics*, **55(11)**, 735-43.

A

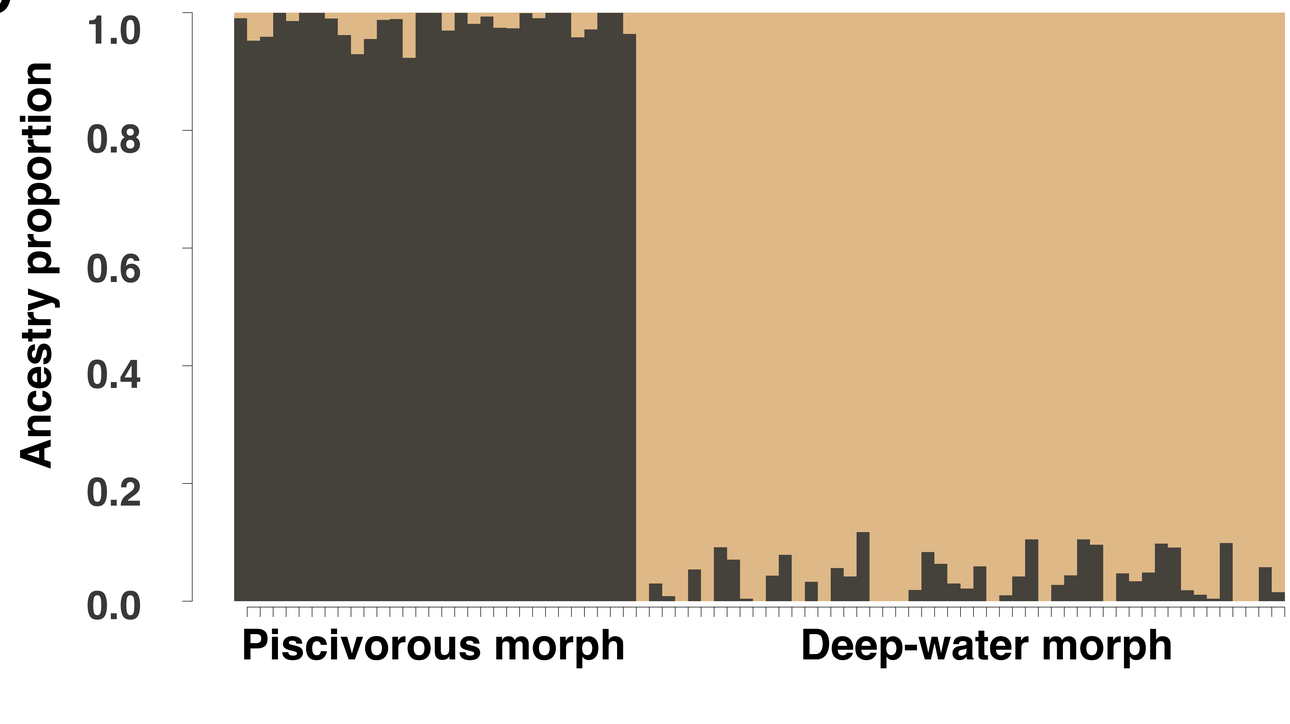


B

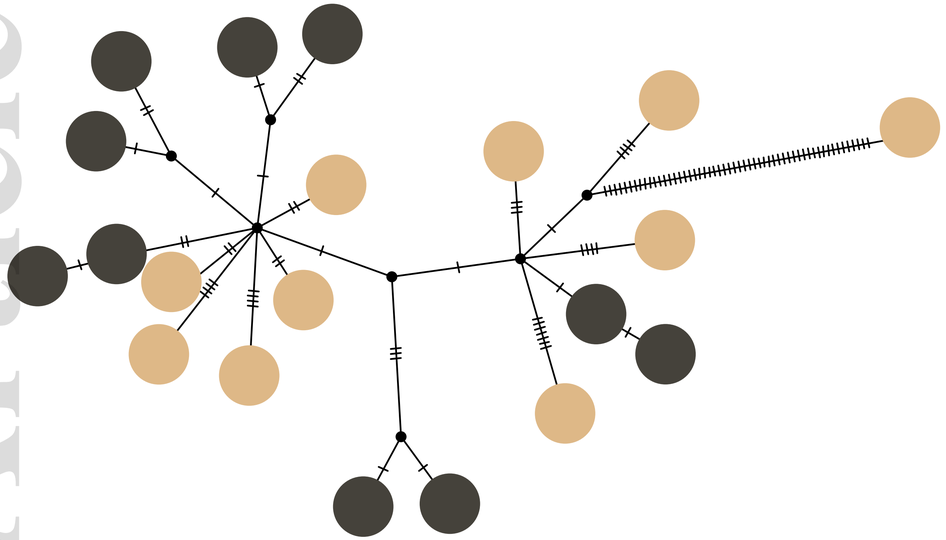
C



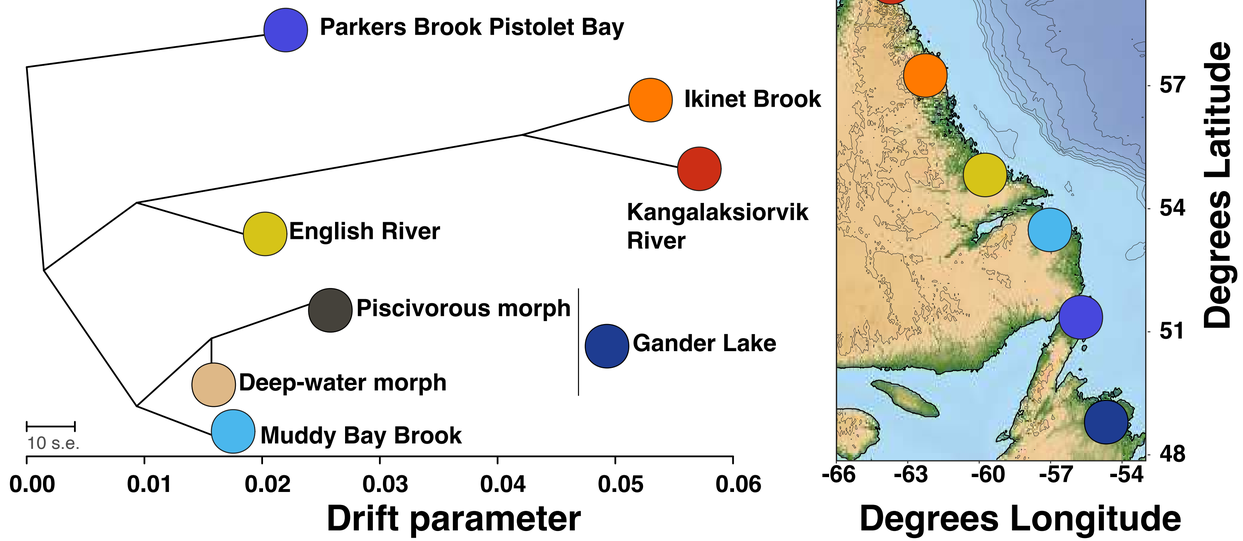
D



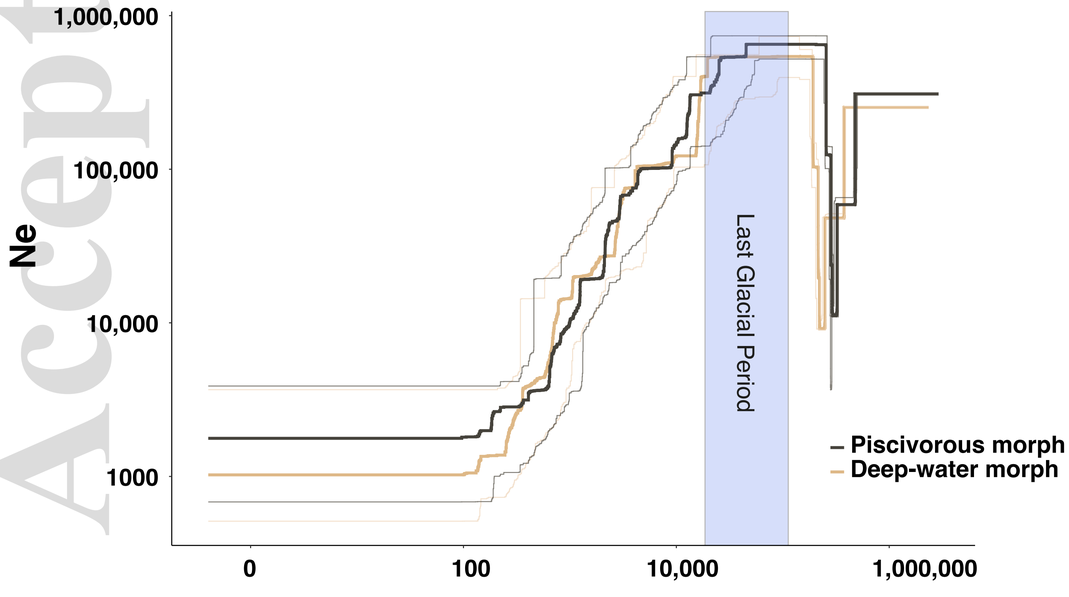
A



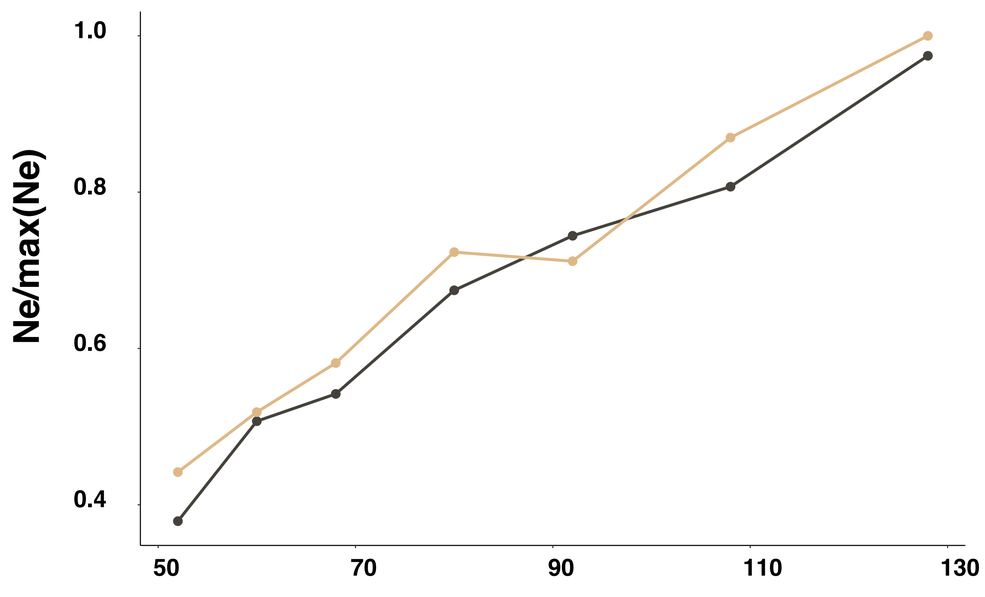
B



C

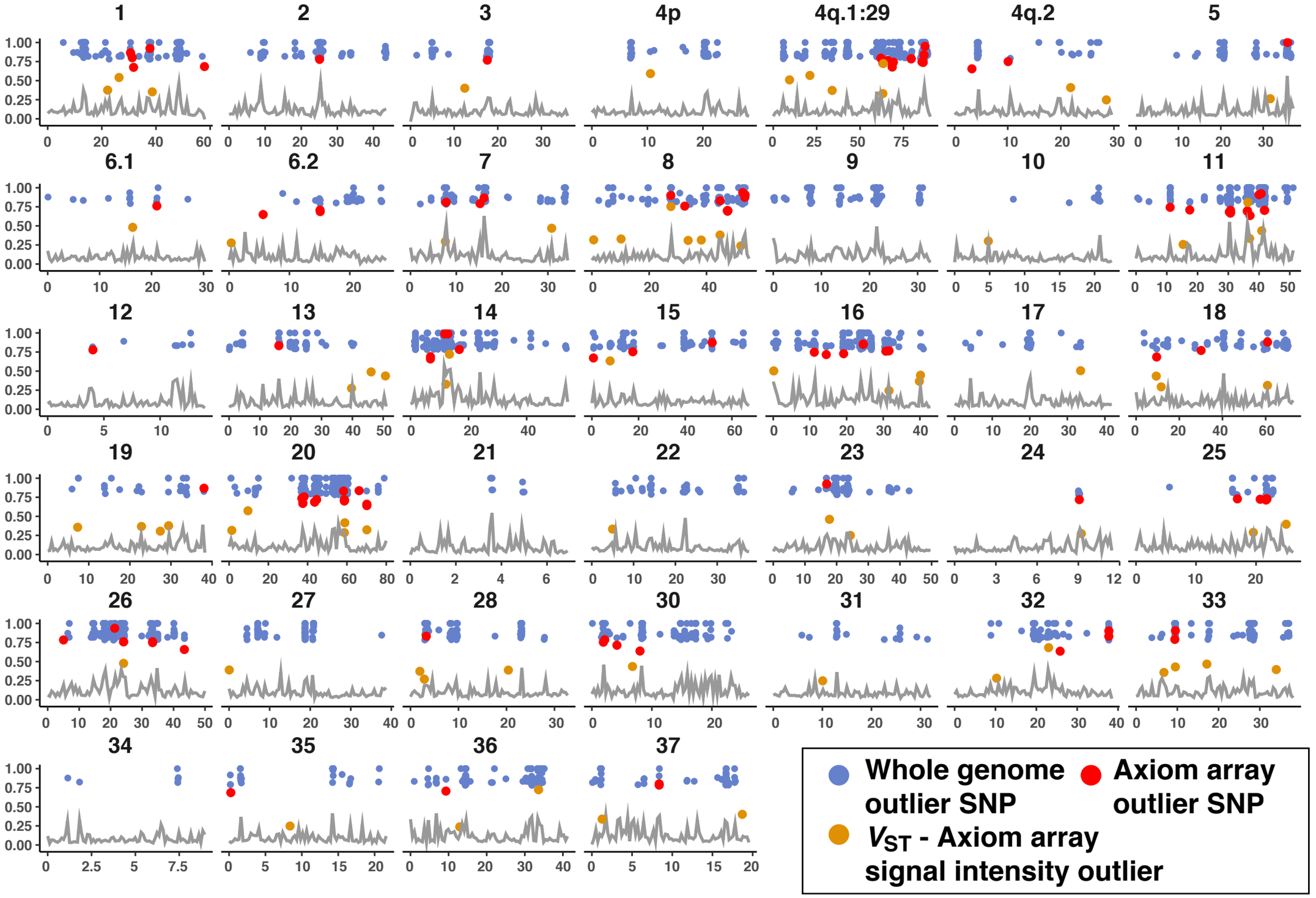


D

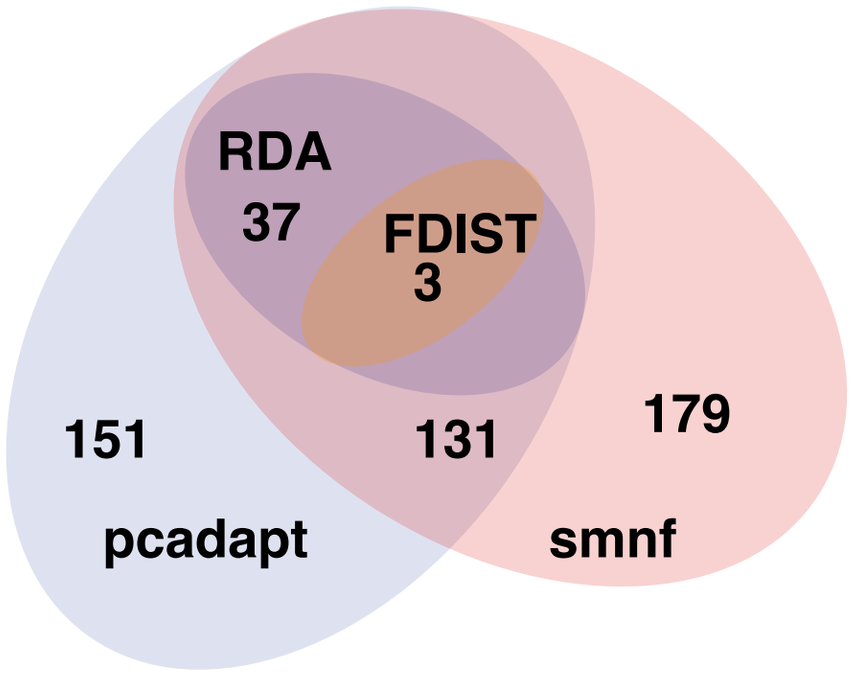


Years before present

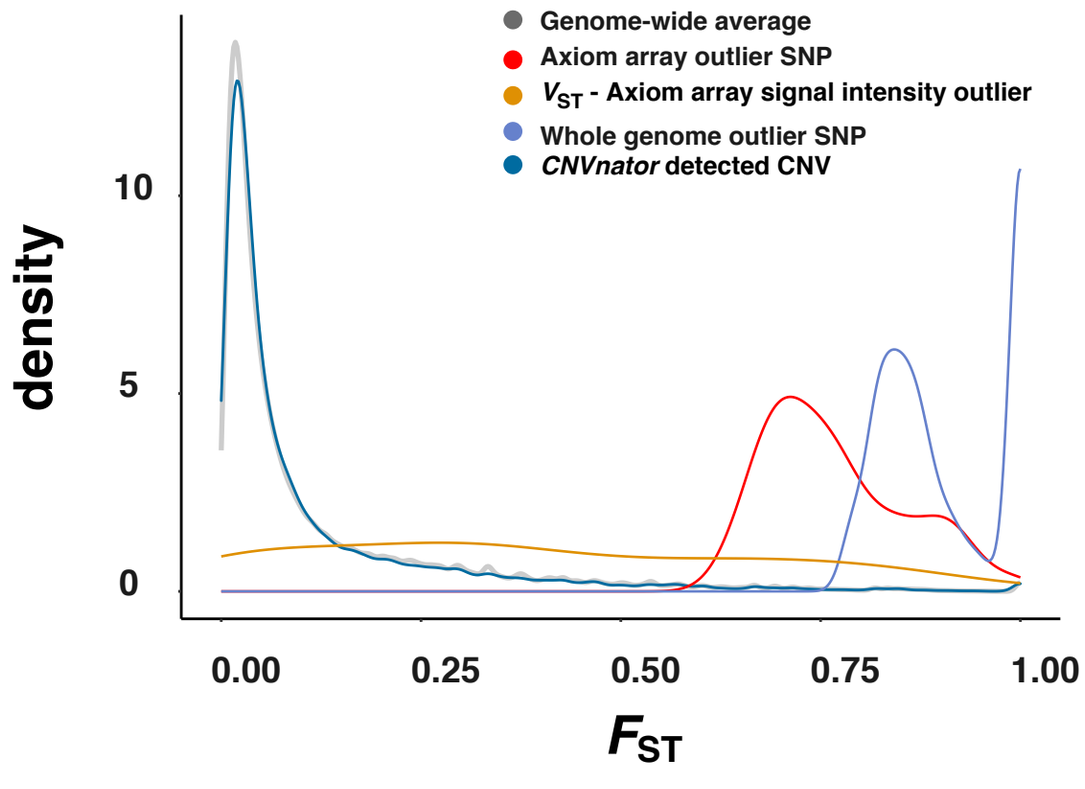
A



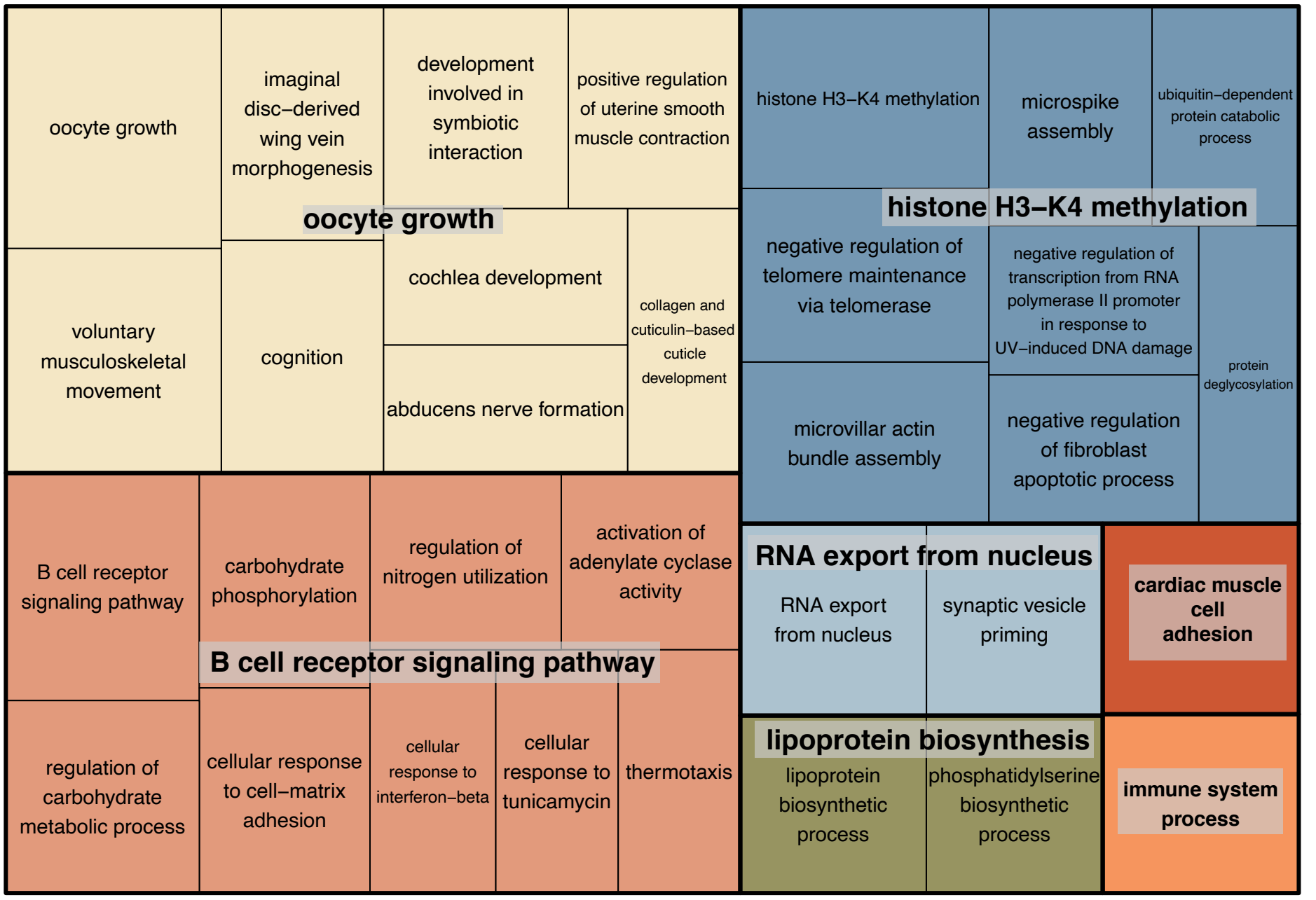
B



C



A



B