

A migration-associated supergene reveals loss of biocomplexity in Atlantic cod

Tony Kess^{1*}, Paul Bentzen², Sarah J. Lehnert¹, Emma V.A. Sylvester¹, Sigbjørn Lien³,
Matthew P. Kent³, Marion Sinclair-Waters⁴, Corey Morris¹, Paul Regular¹, Robert
Fairweather², Ian R. Bradbury^{1,2}

¹ Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St. John's,
Newfoundland, Canada

² Biology Department, Dalhousie University, Halifax, Nova Scotia, Canada

³ Centre for Integrative Genetics, Department of Animal and Aquacultural Sciences,
Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway

⁴ Department of Biosciences, University of Helsinki, Helsinki, Finland

*Correspondence

Tony.Kess@dfo-mpo.gc.ca

Abstract: Intraspecific phenotypic diversity is integral to ecological resilience and the provision of ecosystem services¹. Chromosome structural variation may underpin intraspecific diversity and complex phenotypes² by reducing recombination within supergenes containing linked, co-adapted alleles. Connecting ecologically-relevant phenotypes to genomic variation can enable more precise conservation of exploited marine species by protecting important genetic diversity^{3,4}. Here, using genome-wide association analysis of a 12K single nucleotide polymorphism (SNP) array we confirm that an ancient, derived chromosomal rearrangement consisting of two adjacent inversions is strongly associated with migratory phenotype and individual-level genetic structure in Atlantic cod (*Gadus morhua*) across the Northwest Atlantic. The presence of all identified migration-associated loci within this rearrangement indicates that pervasive variation in migration phenotype is in part controlled by a recombination-resistant supergene, facilitating fine-scale individual phenotypic variation within Northern cod. Furthermore, we reconstruct trends in effective population size over the last century, and find genomic signatures of population collapse, and different patterns of population expansion and decline among individuals based on supergene alleles. We demonstrate declines in effective population size consistent with the onset of industrialized harvest (post 1950) and substantially reduced effective size of individuals homozygous for the derived chromosomal rearrangement relative to heterozygous individuals or those homozygous for the ancestral version of this chromosomal region. These results illustrate how chromosomal structural diversity can mediate fine-scale genetic and phenotypic variation in a highly connected marine species, and suggest a loss of biocomplexity from a migration-associated supergene within Northern cod by overfishing.

Consideration of intraspecific variation is central to the effective management of natural resources and ecosystem services^{5,6}. Individual phenotypic and genetic variation can play a key role in dictating ecological composition and function, and accordingly perturbations to this variation have been shown to alter ecosystem structure⁷. Incorporating intraspecific diversity into management plans requires the identification of heritable phenotypes linked to ecological resilience and biocomplexity^{1,8}. Genomic analyses have increasingly revealed chromosome structural rearrangements (e.g., inversions, translocations) underpinning ecologically variable traits, supporting an important role for genomic architecture in promoting intraspecific ecological diversity^{9,10}. For example, chromosomal inversions may facilitate the evolution of complex phenotypes in sympatry through the formation of supergenes or clusters of linked adaptive loci in regions of reduced recombination⁹. Quantifying the relationship between genomic architectural variation and ecologically relevant phenotypes can provide deeper understanding of the genomic drivers of biocomplexity, enabling genomics-guided species management and precise measurement of human impacts within species¹¹. However, the relationship between genomic architecture and key ecological traits remains largely unknown in most exploited species⁴.

Atlantic cod (*Gadus morhua*) have an extensive history of exploitation and increasing evidence of genomic structural variation associated with ecological adaptation^{4,10,12}. The Northern cod population in northwest Atlantic waters around Newfoundland and Labrador have undergone multiple population declines, the most recent and drastic of which has been driven largely by overharvesting following adoption of industrial-scale fishing, in tandem with climate shifts¹³. In conjunction with recent declines in abundance, size and age at maturity have also declined within the Northern cod stock since the 1970's^{14,15}. Both the presence of a genetic basis for intraspecific variation in the species, and reports of changes in key life history traits suggests the potential that fishery-induced selection has occurred¹⁵.

Individual and regional genomic diversity that differentiates Atlantic cod populations by environment and migratory behavior has been identified across the species range in the North Atlantic¹². Recent genomic analyses have revealed large chromosomal rearrangements on four linkage groups (LG1, LG2, LG7 and LG12)^{10,12}.

Two adjacent inversions within LG1¹⁶ and the rearrangement on LG12 differentiate offshore migratory and coastal non-migratory populations^{4, 10, 16, 17}. The migratory phenotype exhibited by Northern cod may have imposed increased vulnerability to overfishing in these populations¹⁸. However, the link between these rearrangements, migratory phenotype, and the consequence of overharvest remains largely unexamined. Understanding the link between phenotypic variation, population stability, and genomic architectural variation is particularly important in Northern cod, which have undergone one of the most severe population crashes recorded in an exploited marine species, and have failed to recover to historical abundance despite a lengthy and continuing fishing moratorium¹⁹.

Here we explore the genomic basis of individual variation within the Northern cod stock around Newfoundland and Labrador (Figure 1A, Supplementary Table 1) using a 12K single nucleotide polymorphism (SNP) array. We examine the role of chromosomal rearrangements associated with migration phenotype and temporal trends in effective population size over the last century. Within the Northern cod stock, principal component (PC) analysis using 6669 informative SNPs identified three distinct clusters along the first PC axis (Figure 1B). Outlier identification using the Mahalanobis distance for each SNP and the first PC revealed individual genetic structuring was greatest among SNPs localized between 9.89 and 27.22 Mbp on LG1, consistent with boundaries of the derived chromosomal rearrangement consisting of two adjacent inversions on LG1¹⁶. These inversions arose in the Eastern Atlantic 1.6 – 2.0 Mya, prior to the most recent colonization of the Northwest Atlantic^{10,16}, suggesting trans-Atlantic introduction of the derived LG1 rearrangement by migratory cod. The positions of outlier loci indicate observed clusters correspond to individuals homozygous for the ancestral collinear chromosome orientation, and those homozygous or heterozygous for the derived LG1 rearrangement haplotype (Figure 1C). The presence of two adjacent inversions may prevent multiple crossovers events, suppressing recombination along the entire rearrangement and enabling this genomic region to function as a single supergene^{10, 20}.

Population structure analysis was expanded geographically to include samples from throughout the Northwest Atlantic in waters around Newfoundland and Labrador, Southern Canada, the Laurentian Gulf, and coastal Labrador. We conducted separate

analyses using genome-wide SNPs from regions without known inversions ($n=2500$), and SNPs within the rearranged region on LG1 ($n=237$). Individual Neighbour-Joining trees revealed little structuring at neutral loci, and again uncovered the presence of three discrete groups using only SNPs within the LG1 rearranged region (Figure 1D). Bayesian clustering analysis of SNPs within the LG1 rearranged region also demonstrated ancestry proportions consistent with homozygous or heterozygous genotypes from ancestral or derived rearranged LG1 haplotypes (Figure 1E).

To quantify the relationship between chromosomal rearrangements in the Atlantic cod genome and migratory phenotype, we conducted genome-wide association analyses across all sampled populations. We assigned regional-level migratory and non-migratory phenotypes to individuals from each sampled location using previously collected migratory behaviour data²¹ (Supplementary Table 1), and detected migration-associated SNPs using random forest, partial redundancy analysis, and latent factor mixed models. We uncovered 21 SNPs (Supplementary Table 2) associated with migratory phenotype across all three association methods, all located within 9.89 and 27.22 Mbp in LG1, corresponding to the position of the rearrangement (Figure 2A, B, Supplementary Figure 1), and exhibiting high linkage disequilibrium ($r^2 = 0.82$, $sd = 0.09$). Additionally, we detect evidence of a selective sweep within the derived LG1 rearrangement (Figure 2C). A large suite of functional variants (>300) differentiated between coastal and migratory ecotypes in Norwegian waters have previously been identified within this rearrangement¹⁶. Several of these alleles are suggested to affect swim bladder function and muscular efficiency, facilitating vertical movement at variable depths in offshore sites and enhancing muscular capacity for strenuous migration^{10, 16}.

Northern cod experienced substantial overharvest, reflected in the occurrence of one of the biggest population crashes of any marine vertebrate that resulted in $>90\%$ decline in abundance during the 20th century^{14, 18, 22}. Despite the scale of stock collapse, previous neutral molecular marker-based attempts have failed to detect genetic signatures of this decline²³. Here, we find that genomic data can detect declines specific to groups with ancestral or derived copies of LG1. We separated individuals into LG1 homozygous derived, LG1 homozygous ancestral and heterozygous groups and then estimated trends in recent effective population size (N_e) in each group using genome-wide SNPs outside

of LG1 and LG12 rearrangements identified in Northern cod. We then compared proportional change in N_e relative to maximum observed effective size within each group during the past 150 years to observed and reconstructed stock abundance¹³. We found a similar starting effective population size at the 1860 time point for each group ($N_e = 1976 - 2629$, Supplementary Table 3), but observed different patterns of expansion and decline between LG1 homozygous ancestral, LG1 homozygous derived, and heterozygous groups (Figure 3A, B), suggesting different selective histories. Within LG1 homozygous ancestral and heterozygous groups, we observed an historical expansion of effective size, indicated by large N_e estimates with confidence intervals including negative values (i.e. infinite), coinciding with a period of population productivity from 1900 to approximately 1970¹³. From 1970 to present, we observe a shared pattern of decline but different contemporary effective sizes of each group. Although non-independence of SNPs may reduce confidence in exact estimates of N_e ²⁴, the magnitude of different contemporary N_e values between LG1 homozygous ancestral ($N_e = 9697$) and derived ($N_e = 165$) individuals indicates potential for loss of migratory phenotype due to overharvest. Comparing LG1 rearrangement groups by region revealed decline in LG1 homozygous ancestral and heterozygous individuals driven by decline in sites around Labrador, suggesting spatial variation in selection intensity (Supplementary Figure 2). The capacity of genomic data to detect recent collapse of Northern cod was also supported by allele frequency-based bottleneck tests.

Our results provide novel advances in understanding biocomplexity at a genomic level in marine species. We demonstrate that intraspecific diversity in migratory behavior is associated with a derived chromosomal rearrangement consisting of two adjacent inversions spanning LG1 and illustrate that this rearrangement enables extensive differentiation among individuals within Northern cod. Suppressed recombination between opposite orientations of the LG1 rearrangement can promote coordinated inheritance of migration-associated loci, allowing the two adjacent inversions to function as a single supergene protected from gene flow^{2, 9, 10, 20}. This supergene is associated with genetic signatures of extensive decline in homozygous individuals, potentially driven by overharvesting. Collapse of Northern cod has resulted in massive restructuring of coastal shelf ecosystems²⁵. Continued overharvesting of the derived, migration-associated LG1

rearrangement could lead to widespread altered ecological function and composition in the Northwest Atlantic through changing cod distribution and site persistence. Removal of this genomic diversity through overfishing may also reduce the buffering effect provided by phenotypic diversity within these populations, increasing risk of future collapse¹. Management of Northern cod stocks may thus require consideration of the genomic architectural diversity within these populations to ensure maintenance of biocomplexity across the Northwest Atlantic.

References

1. Schindler, D.E., *et al.* Population diversity and the portfolio effect in an exploited species. *Nature* **465**, 609–613 ((2010).
2. Kirkpatrick, M. The evolution of genome structure by natural and sexual selection. *Journal of Heredity* **108**, 3–11 (2016).
3. Rudman, S. M., *et al.* What genomic data can reveal about eco-evolutionary dynamics. *Nature Ecology & Evolution* **2**, 9-15 (2018).
4. Bradbury, I. R., *et al.* Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications*, **6**, 450–461 (2013).
5. Mimura, M., *et al.* Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, **10**, 121-139 (2017).
6. Des Roches, S., *et al.* The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, **2**, 57–64 (2018).
7. Farkas, T.E., Mononen, T., Comeault, A.A., Hanski, I. & Nosil, P. Evolution of camouflage drives rapid ecological change in an insect community. *Current Biology*, **23**, 1835–1843 (2013).
8. Hilborn, R., Quinn, T. P., Schindler, D. E., & Rogers, D.E. Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences USA*, **100**, 6564-6568 (2003).
9. Joron, M., *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, **477**, 203–206 (2011).
10. Sinclair-Waters, M., Bradbury, I. Morris, C.J., Lien, S. Kent, M., & Bentzen, P. Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic cod in the northwest Atlantic. *Molecular Ecology*, **27(2)**, 339-351 (2018)
11. Allendorf, F.W., Hohenlohe P.A., & Luikart, G. Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709 (2010).
12. Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., & Jentoft, S. Trans-oceanic genomic divergence of Atlantic Cod ecotypes is associated with large inversions. *Heredity*, **119(6)**, 418–428 (2017).

13. Rose, G.A. Reconciling overfishing and climate change with stock dynamics of Atlantic cod (*Gadus morhua*) over 500 years. *Canadian Journal of Fisheries and Aquatic Sciences*, **61**, 1553 – 1557 (2004).
14. Hutchings, J.A. Life history consequences of overexploitation to population recovery in Northwest Atlantic cod (*Gadus morhua*). *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 824–832 (2005).
15. Olsen, E.M., *et al.* Maturation trends indicative of rapid evolution preceded the collapse of Northern cod. *Nature*, **428**, 932–935 (2004).
16. Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., De Rosa, M. C., Righino, B., Johansen, T., Otterå, H., Sonesson, A., Lien, S. & Andersen, Ø. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, **25**, 2130-2143 (2016).
17. Sinclair-Waters, M., Bentzen, P., Morris, C.J., Ruzzante, D.E., Kent, M., Lien, S., & Bradbury, I. Genomic tools for management and conservation of Atlantic Cod in a coastal marine protected area. *Canadian Journal of Fisheries and Aquatic Sciences*, (2018). <https://doi.org/10.1139/cjfas-2017-0254>
18. Hutchings, J.A. Spatial and temporal variation in the density of northern cod and a review of the hypotheses for the stock collapse. *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 943–962 (1996).
19. Morgan, M.J., Koen-Alonso, M., Rideout, R.M., Buren, A.D., & Maddock Parsons, D. Growth and condition in relation to the lack of recovery of northern cod. *ICES Journal of Marine Science* 75(2), 631-641 (2018).
20. Hoffmann, A. A. & Rieseberg, L. H. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu. Rev. Ecol. Evol. Syst.* **39**, 21–42 (2008)
21. Robichaud, D., & Rose, G.A. Migratory behaviour and range in Atlantic cod: inference from a century of tagging. *Fish and Fisheries*, **5**, 185–214 (2004).
22. COSEWIC. Assessment and status report on the Atlantic Cod (*Gadus morhua*) in Canada. (2010).
23. Ruzzante, D.E., Taggart, C.T., Doyle, R.W., Cook, D. Stability in the historical pattern of genetic structure of Newfoundland cod (*Gadus morhua*) despite the catastrophic decline in population size from 1964 to 1994. *Conservation Genetics*, **2**, 257–269, (2001).
24. Waples, R. K., Larson, W. A., & Waples, R. S. Estimating contemporary effective population size in non-model species using linkage equilibrium across thousands of loci. *Heredity*, **117**, 233–240 (2016).
25. Frank, K. T., Petrie, B., Choi, J. S. & Leggett, W. C. Trophic cascades in a formerly cod-dominated ecosystem. *Science*, **308**, 1621–1623 (2005).

Methods

Sampling and Genotyping: We combined Atlantic cod (*Gadus morhua*) genotype data from three studies conducted by Sinclair-Waters *et al.*^{10,17} and Berg *et al.*¹² (Supplementary Table 1). The combined dataset comprised 511 individuals in 24 sample groups collected between 2001 and 2015 from across 18 sites the Northwestern Atlantic. Sampling site details, sample collection and preparation methods are described in Sinclair-Waters *et al.*^{10,17}, and Berg *et al.*¹². All sampled individuals were genotyped using a 12K SNP array developed for Atlantic cod at the Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences in Ås, Norway²⁶. Following correction for strand flips, we retained a subset of 6669 polymorphic SNPs with complete mapping information across all datasets for further analysis.

Genetic structure detection: To characterize genomic sources of individual variation within Northern cod, we used the *pcadapt*²⁷ R package to conduct principal component analysis (PCA) on the 330 individuals from sampling sites around Newfoundland and Labrador. Using 6030 SNPs with minor allele frequency greater than 0.025, we carried out PCA retaining the first PC axis to identify the largest source of individual genomic variation. We plotted *p* values from transformed Mahalanobis distances measuring significance of correlation with each SNP and this PC axis across each linkage group using the R package *ggman*²⁸.

Range-wide phylogeny and genetic structure analysis: To identify intraspecific variation in LG1 rearrangement frequency relative to the genome-wide average across the northwestern Atlantic, we then calculated individual neighbor-joining trees using all 511 samples. We calculated separate trees for the 237 SNPs with map coordinates within the LG1 rearrangement and for 2500 randomly selected SNPs outside of the known chromosomal rearrangements found in Atlantic cod on linkage groups 1, 2, 7, and 12¹². Trees were calculated using POPULATIONS v1.2.33²⁹, based on Cavalli-Sforza and Edwards³⁰ chord distances with 1,000 bootstrap replicates on loci. Neighbour-joining trees were visualized using FigTree v1.74³¹. Subsampling of genome-wide and inversion loci was conducted using the *genepopedit* R package³².

Next, we conducted Bayesian clustering analysis of all 511 sampled individuals with three replicate Markov-Chain Monte Carlo runs on LG1 rearrangement genotypes in

STRUCTURE v2.3.4³³. We implemented these runs using *parallelstructure*³⁴ in R, assuming K=2 clusters, and conducting 100,000 burn-ins and 500,000 iterations. We visualized individual genetic ancestry proportions using CLUMPAK³⁵. Conversion of genotypes to STRUCTURE format was carried out using PGDSpider³⁶.

Genome-wide association analysis of migratory phenotype: We assigned migratory behavioural phenotypes to all individuals in each sampling site from previously identified behavioural classes categorized in Atlantic cod by Robichaud and Rose²¹ (Supplementary Table 1). When multiple locations within the range of a sampling site exhibited variable phenotypes, we assigned the most frequent phenotype identified across these locations. Due to potential variation in sampling and behavioural measurement methods between studies, we binned three migratory (M) behavioural classes together (accurate homers, AH, inaccurate homers, IH, and dispersers, D) as migratory (M), and categorized sedentary (S) individuals as non-migratory (NM). We then conducted genome-wide association using two polygenic methods, random forest (RF)³⁷ and partial redundancy analysis (RDA)³⁸ and a single locus method using latent factor mixed models (LFMM)³⁹.

To find polygenic associations with migratory phenotype, we first used random forest³⁷ classification, a non-parametric, decision tree-based algorithm. We produced a matrix with individual migratory phenotypes as the response variable, and SNP genotypes at each locus as predictor variables. Random forest was run using the R package *randomForest*⁴⁰ after imputing missing genotypes using the *rfImpute* function. We ran 5000 trees as predictor rank stabilized was ensured with this quantity of trees. We set the mtry parameter to default following testing of optimal mtry parameter values using the *tuneRF* function. Stratified sampling was applied to ensure unbiased representation of migratory phenotype class within each tree. Out-of-bag mean decrease in accuracy was calculated across all trees for each SNP as an estimate of predictor importance. We retained all SNPs with mean decrease in accuracy greater than 0.0015; values below this limit exhibited a sharp drop in mean decrease in accuracy, indicating relatively limited utility of these SNPs in classifying migratory and non-migratory individuals.

We then carried out a constrained ordination using partial redundancy analysis. This method maximizes the proportion of variance explained in linear combinations of

response variables by linear combinations of predictor variables⁴¹. To detect polygenic associations, we used a matrix of diploid genotypes for each individual at each locus as dependent variables, and migratory phenotype as the predictor variable and carried out redundancy analysis using the *vegan*⁴² R package. Missing genotype calls were imputed for each individual using LinkImpute⁴³. To account for population structure due to geography, we conditioned the relationship between genotype and migratory phenotype by latitude and longitude. Significance of genotype-phenotype association was calculated with 999 permutations ($p < 0.001$). We categorized SNPs with absolute values of canonical redundancy axis scores exceeding the 99% percentile on the first axis (RDA1) as significantly associated with migratory phenotype. Linkage disequilibrium was calculated between migration-associated SNPs identified across all three methods in PLINK v 1.90⁴⁴ using an r^2 cut-off of 0.2.

Detection of SNP association in LFMM was conducted by using a matrix factorization algorithm to calculate loadings of K latent factors, similar to a reduced set of principal components, that each describe a different source of population structure³⁹. We empirically estimated the value of K with sparse nonnegative matrix factorization⁴⁵ using the *sNMF* function on all 6669 SNPs in the *LEA* R package⁴⁶. We then used the *lfmm* function in *LEA* to identify SNP association with migratory phenotype using K=4 latent factors. We converted p values for all loci to q values to control for false discovery rate⁴⁷ and retained loci with $q < 0.05$.

Detection of selective sweeps: To identify genomic regions exhibiting signatures of selection on migratory phenotype within Northern cod, we calculated composite likelihood ratio⁴⁸ (CLR) across LG1 using the SweeD⁴⁹ software package. We separated individuals from Newfoundland and Labrador sampling sites into LG1 homozygous ancestral and LG1 homozygous derived, rearranged classes based on individual PC scores. We then calculated CLR for windows of 20 SNPs within derived ($n = 41$) and ancestral ($n = 136$) individuals, assuming a folded site-frequency spectrum, and selected regions exhibiting CLR exceeding the 90% percentile as potential targets of selection.

Contemporary and historical effective population size estimation: To detect the recent genetic history of Northern cod, we used the linkage disequilibrium (LD) based method implemented in LinkNe⁵⁰ to estimate effective population size (N_e). We

calculated N_e across a 150 year period in the 330 Northern cod samples from Newfoundland and Labrador. We separated LG1 homozygous derived, heterozygous ($n = 153$) and LG1 homozygous ancestral individuals based on PC score. Prior to N_e estimation, we filtered genome-wide SNP datasets for these individuals to remove SNPs within the LG1 and LG12 rearrangements identified within Northern cod, resulting in a final dataset of 6270 SNPs. These regions were removed as selection and reduced recombination within rearrangements regions may skew LD inference of historical N_e . For each group, we calculated N_e separately in LinkNe in bins of 0.05 Morgans for all SNPs with minor allele frequency values greater than 0.05. We binned N_e estimates into generations calculated from recombination rate, assuming a mean generation time of 6 years¹⁵. We assigned a contemporary sampling date of 2010, corresponding to an intermediate estimate of sampling dates from all Northern cod samples. To compare changes in effective population size between groups that exhibited large differences in N_e , we calculated relative N_e by scaling N_e values within each group by the maximum observed N_e for that group. We then plotted relative N_e estimates as proportions against estimates of cod abundance measured from 1983 to 2004, and historical reconstructions of stock biomass from 1850 to present¹².

To account for the possibility of observed decline in LG1 homozygous derived individuals driven by small sample size, we generated 10 subsamples of 41 LG1 homozygous ancestral individuals and conducted replicate effective size estimation. We observed similar starting effective sizes across replicates (mean $N_e = 2577.449$, $sd = 170.1235$) and identified expansions coinciding with stock recovery in all 10 samples. We did not observe patterns of decline matching the pattern observed in the sample of homozygous derived individuals, indicating this decline is likely not an artefact of small sample size (Supplementary Figure 3). We identified a recent, post-1980 decline across 3 of the 10 subsamples, which could indicate spatial heterogeneity in recent decline patterns across LG1 homozygous ancestral Northern cod samples (Supplementary Figure 3). To test this hypothesis, we split LG1 homozygous ancestral, LG1 homozygous derived, and heterozygous groups into population groups corresponding to coastal Labrador samples, offshore samples, and samples from coastal Newfoundland. We then conducted LinkNe analysis on these samples, revealing recent decline of LG1

homozygous ancestral and heterozygous individuals within the coastal Labrador grouping (Supplementary Figure 2). Together, these results indicate that small sample sizes have likely not produced false signals of decline, and demonstrate spatial variation in cod stock decline across Newfoundland and Labrador.

Bottleneck detection: To provide secondary genetic confirmation of observed stock collapses observed in abundance data and using effective size reconstruction, we also tested for presence of recent genetic bottlenecks in LG1 homozygous ancestral, heterozygous, and LG1 homozygous derived groups. We analyzed 6270 genome-wide SNPs excluding LG1 and LG12 rearrangements in each group using the BOTTLENECK⁵¹ software to detect bottlenecks assuming an infinite alleles model. We inferred bottlenecks from presence of a mode-shift in allele frequency toward reduction in the number of low frequency alleles, or significant sign and standardized differences test results. Bottlenecks were observed using all three methods in homozygous groups (sign test, standardized differences test $p < 0.0005$), but allele frequency mode shift was not detected in heterozygotes. This observation may be attributed to historical scenario sensitivity of the different tests used in the BOTTLENECK software, or may reflect differences due to recent admixture within heterozygous individuals.

References:

26. Kent, M.P., *et al. in prep.*
27. Luu, K., Bazin, E., & Blum, M. G., *pcadapt*: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, **17**, 67-77 (2017).
28. Rajagopal, V.M., ggman: R package for Manhattan Plots. GitHub repository, <https://github.com/drveera/ggman> (2017).
29. Cavalli-Sforza, L.L. & Edwards, A.W.F. Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics* **19**, 233-257 (1967).
30. Langella, O. POPULATIONS 1.2.28. Population genetic software (individuals or populations distances, phylogenetic trees). Available at <https://launchpad.net/~olivier-langella/+archive/ubuntu/ppa>. (2012).
31. Rambaut, A. FigTree v1.4. Available at <http://tree.bio.ed.ac.uk/software/figtree/>. (2012).
32. Stanley, R. R., Jeffery, N. W., Wringe, B. F., DiBacco, C., & Bradbury, I. R. genepopedit: A simple and flexible tool for manipulating multilocus molecular data in R. *Molecular Ecology Resources*, **17(1)**, 12–18 (2016).

33. Pritchard J. K., Stephens M., & Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959 (2000).
34. Besnier, F., & Glover, K.A. ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PLoS One*, **8**, e70651 (2013).
35. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179–1191 (2015).
36. Lischer, H.E.L., & Excoffier, L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299 (2012).
37. Breiman, L. Random forests. *Machine Learning*, **45**, 5–32 (2001).
38. Rao, C.R. The use and interpretation of principal component analysis in applied research. *Sankhya: The Indian Journal of Statistics, Series A*, **26**, 329–358 (1964).
39. Frichot, E., Schoville, S. D., Bouchard, G., & François, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–1699 (2013).
40. Liaw, A., & Wiener, M. Classification and regression by randomForest. *R News*, **2**, 18–22 (2002).
41. Lasky, J. R., Des Marais, D. L., McKay, J., Richards, J. H., Juenger, T. E., & Keitt, T. H. Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529 (2012).
42. Oksanen, J., *et al.* vegan: Community Ecology Package. R package version 1.17 - 6. <http://CRAN.R-project.org/package=vegan>. (2011).
43. Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G., & Myles, S. LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3: Genes, Genomes Genetics*, **5**, 2383–2390 (2015).
44. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., & Lee, J.J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4** (2015).
45. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics*, **196**, 973–983 (2014).
46. Frichot, E., & François, O. LEA: A R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6**, 925–929 (2015).
47. Benjamini, Y., & Hochberg, Y. 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 (1995).
48. Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., & Bustamante, C. Genomic scans for selective sweeps using SNP data. *Genome Research* **15**: 1566–1575 (2005).
49. Pavlidis, P., Zivkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*. **30**, 2224–2234 (2013).

50. Hollenbeck, C.M., Portnoy, D.S., & Gold, J.R. A method for detecting recent changes in contemporary effective population size from linkage disequilibrium at linked and unlinked loci. *Heredity*, **117**, 207–216 (2016).
51. Piry, S., Luikart, G., & Cornuet, J.M. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503 (1999).

Acknowledgements

We thank C. Hollenbeck for assistance with LinkNe software. We also thank L. Weir, D. Heath, and R. Waples for helpful comments on the manuscript, and the Centre for Integrative Genetics (CIGENE) for genotyping of Atlantic cod used in this study.

Author Contributions

TK, SJL, IRB, and PB designed the study. TK, SJL and ES performed statistical analyses. PB, IRB, MSW, RF, MK, and SL provided molecular data for the study. PR provided cod abundance survey data. TK drafted the manuscript and all authors contributed to writing and editing of the final manuscript.

Figure Captions

Figure 1. Map of Atlantic cod (*Gadus morhua*) sampling locations in the northwest Atlantic and genetic structuring across all sampled sites and within Northern cod.

A) Sampling sites across North America colour-coded by population **B)** Genetic structure within Northern cod visualized using the first principal component (PC) calculated from 6030 SNPs in *pcadapt*. Colours correspond to sample groups taken from sites within the greater Northern cod population **C)** Genomic sources of individual level-genetic variation Northern cod assessed by significance of correlation (p value) of each SNP locus with the first PC axis obtained from transformed Mahalanobis distances. SNPs within the LG1 rearrangement are coloured in red. **D)** Individual neighbour joining trees for 511 individuals from all sampling locations calculated using 2500 randomly selected SNPs outside of Northern cod inversions (left) and 237 SNPs in the LG1 rearranged region(right) **E)** Bayesian clustering results from STRUCTURE for 237 SNPs in the LG1 rearranged region calculated among all 511 individuals from all sampling locations for $K=2$ genetic clusters.

Figure 2. Genome-wide association of 6669 SNPs with migration phenotype assigned by Robichaud and Rose²¹, across all sampling locations, and selective sweeps within LG1.

A) Mean decrease in accuracy per SNP in Random Forest prediction of migratory phenotype, categorizing SNPs with mean decrease in accuracy > 0.015 as significantly associated. SNPs within rearrangements are coloured in red. **B)** Absolute values of scores for each SNP along the first redundancy canonical axis (RDA1) in partial redundancy analysis of migratory phenotype conditioned by geographic location, treating all SNPs in the 99th percentile as significantly associated **C)** Selective sweeps in LG1 homozygous ancestral and LG1 homozygous derived individuals detected by calculating composite likelihood ratio (CLR) per 20 SNP window with regions exceeding the 90th percentile considered likely candidates for selection

Figure 3. Comparison of recent relative effective population size (N_e) in LG1 homozygous ancestral, LG1 homozygous derived and heterozygous individuals and measured and historically reconstructed stock abundance of Northern cod.

A) Recent effective population size relative to maximum effective population size within each group of LG1 homozygous ancestral, LG1 homozygous derived and heterozygous Northern cod groupings calculated using LinkNe, plotted with reconstructed estimates of historical biomass measured from 1850 – present¹² relative to maximum value. Confidence intervals with arrows indicate very large N_e . **B)** Recent proportional effective population size relative to maximum effective population size of LG1 homozygous ancestral, homozygous LG derived and heterozygous Northern cod calculated using LinkNe compared to measurement of stock abundance from 1983 – 2004 relative to maximum value.

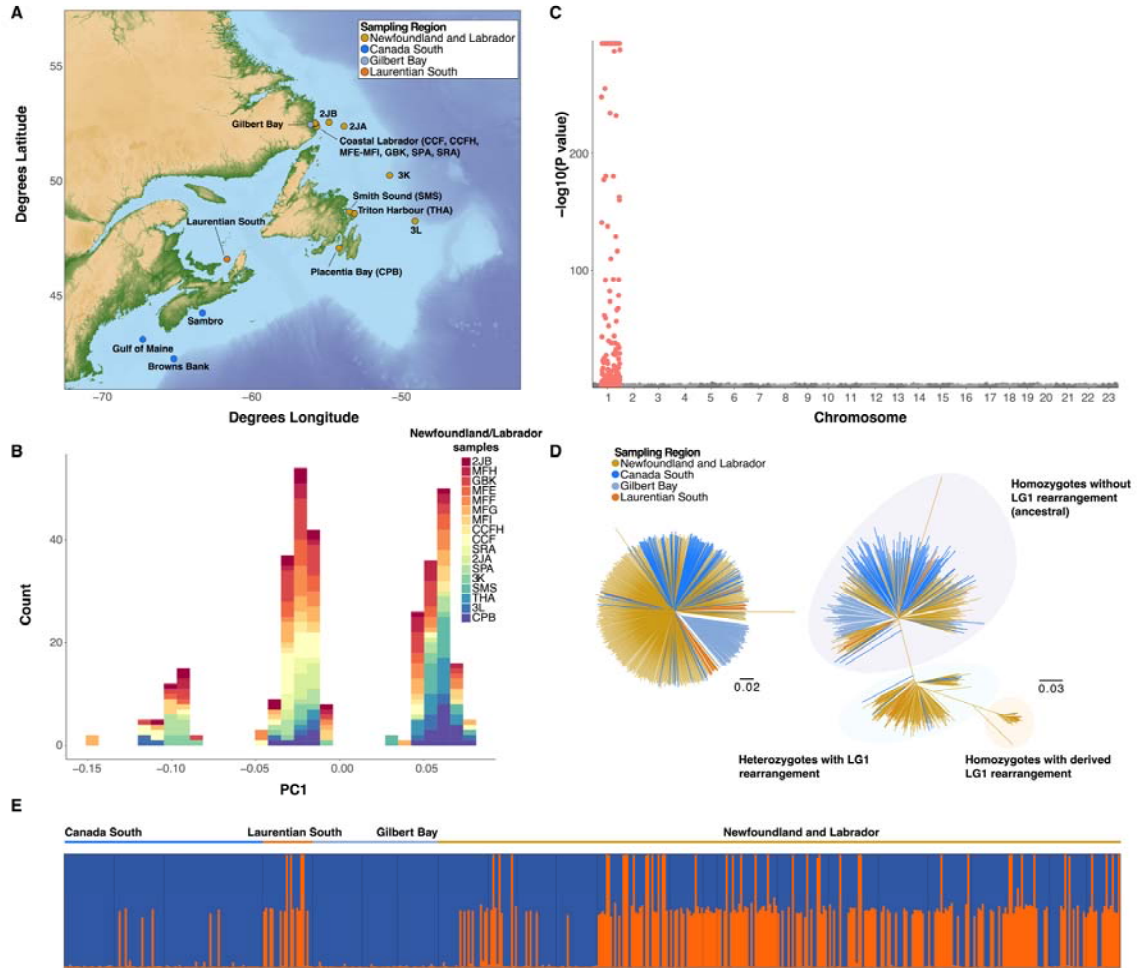
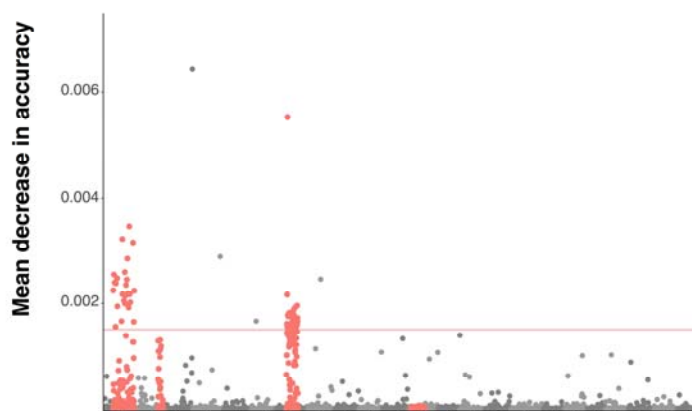
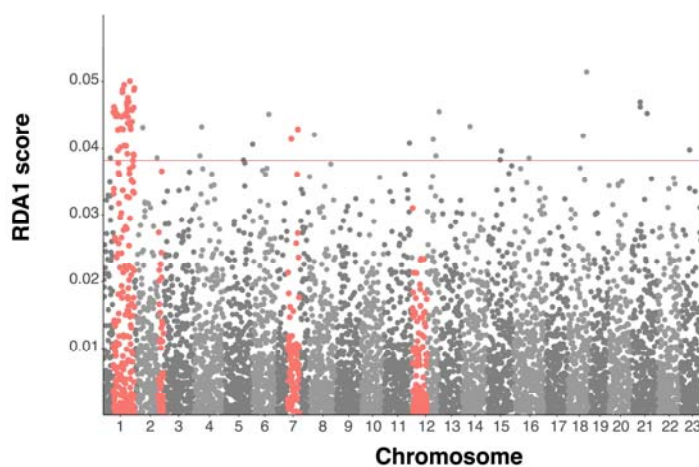


Figure 1

A Random forest



B Partial redundancy analysis



C LG1 sweep detection

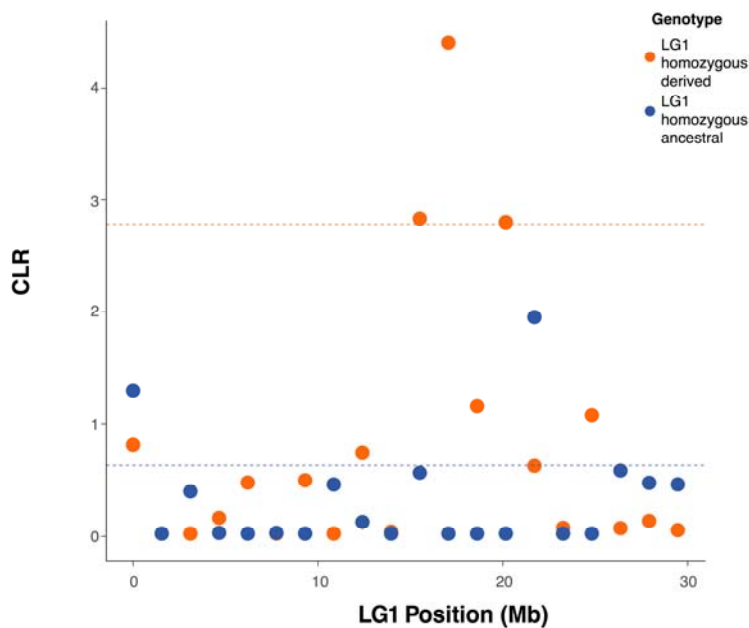


Figure 2

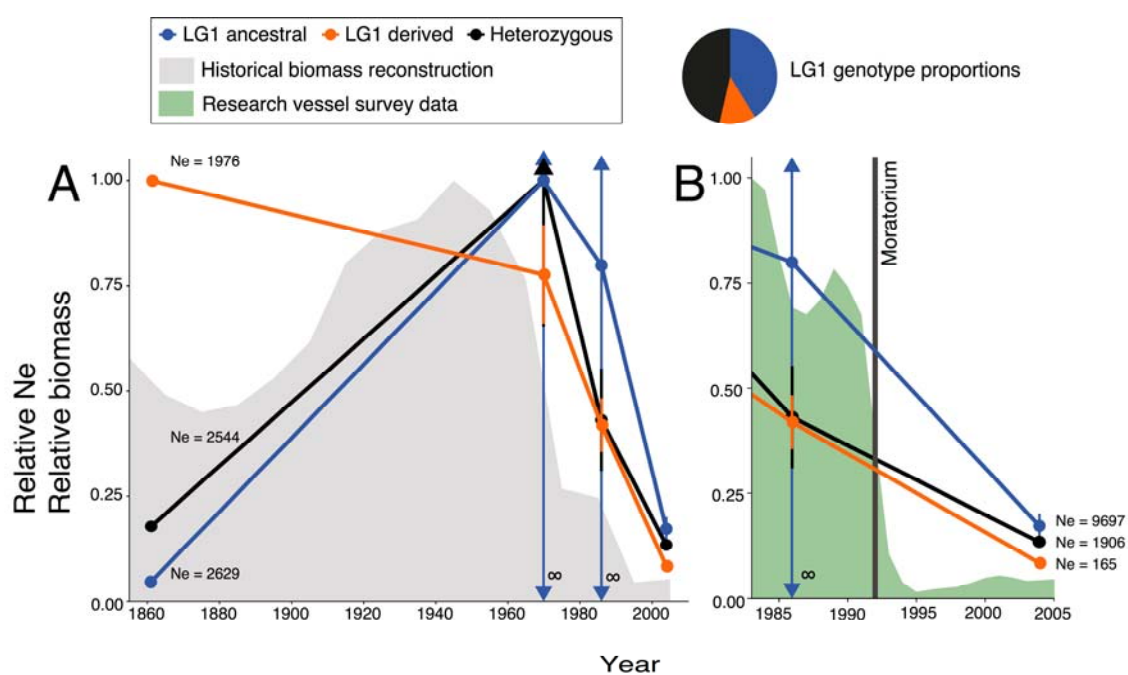


Figure 3