Genomics-based Mixed-stock Analysis of Brook Trout Reveals Cryptic Population Structure and Complex Lake Migrations

Badrouyk Chamlian

A Thesis

In the Department of

Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of

Master of Science (Biology)

at Concordia University

Montréal, Québec, Canada

April 2024

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Badrouyk Chamlian

Entitled: Genomics-based Mixed-stock Analysis of Brook Trout Reveals Cryptic Population Structure and Complex Lake Migrations

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____Chair

Dr. James Grant

_____External Examiner

Dr. Daniel Ruzzante

_____Examiner

Dr. James Grant

_____Examiner

Dr. Selvadurai Daya Dayanandan

_____Thesis Supervisor

Dr. Dylan Fraser

Approved by:

_____

Dr. Robert Weladji, Graduate Program Director

_____

Dr. Pascale Sicotte, Dean of Faculty

April 9, 2024

Abstract

Genomics-based Mixed-stock Analysis of Brook Trout Reveals Cryptic Population Structure and Complex Lake Migrations

Badrouyk Chamlian

Effective fishery management relies on knowing the contributions of genetically distinct populations to mixed-stock harvests. We investigated population genomic structure and harvest contributions of lake-migratory brook trout inhabiting three large Quebec lakes (Mistassini, Mistasiniishish, Waconichi). These brook trout support fisheries important to the Cree Nation of Mistissini and their tourism outfitting industry. Together with local partners we collected 1063 samples from spawning sites and feeding areas between 2020-2022. We then used a GTseq (Genotyping-in-Thousands by sequencing) panel of 393 single nucleotide polymorphisms to: i) infer population genetic structure and test for unknown populations; ii) assign individuals to their population of origin, and iii) determine harvest contributions of genetically distinct populations. Our results revealed population structure in two of three study lakes and extensive movements of brook trout, with some individuals travelling over 100km away from spawning rivers. In the largest lake (Mistassini), two of three populations contributed over 90% of the lake's harvest and exhibited distinct spatial distributions that were stable across years. In Mistasiniishish Lake, over 80% of harvested trout originated from a single, previously known population; the remaining trout originated from a cryptic, unsampled population with a strongly overlapping spatial distribution. No population structure was detected in Waconichi Lake. We also detected low levels of migration from Mistasiniishish Lake into Mistassini Lake through a waterfall historically reported to be a dispersal barrier. Our results illustrate the precision afforded by GTseq to inform insights into the ecology and genetics of lake-migratory salmonids, thereby facilitating local management for sustainable fisheries.

Acknowledgements

I extend my gratitude to the Cree Nation of Mistissini for allowing this project to take place. Special thanks to Pamela MacLeod and Hubert Petawabano for coordinating sampling efforts and for being our key contacts with the community. A huge shout-out to the fishing guides and sample collectors: Norman Neeposh, Jacob Coon-Come, Charlo Blacksmith, Andrew Coon, Andrew Iserhoff, Brett Studden, Natalya Assance, George Coon, Anthony MacLeod, David Hardie, George Coon-Come, Jonathan Linton, Dale Blacksmith, Kia Marin, Leslie Mianscum, Nathaniel Coon-Come, Richard Iserhoff, Richard Henri, Sam Coon-Come, Shannon Clarke, Stanley Mianscum, Thomas Coon, and all Nibiischii fishers. A heartfelt thanks to Kia Marin for coordinating my field work.

I would also like to thank my colleagues at the Fraser lab for supporting me: Julie Gibelli, Hari Won, Hyung-Bae Jeon, Johnathan Lemay, Thaïs Bernos, Sofia D'Angelo, Brian Gallagher, Alexandra Engler, Raphaël Bouchard, and Jukka-Pekka Verta. Extra special thanks to Sozos Michaelides for his knowledge and guidance, as well as for creating a lively and friendly work environment.

I sincerely thank my supervisor, Dr. Dylan Fraser, for introducing me to the world of conservation biology and giving me the opportunity to conduct this research project. He has been a flawless advisor, providing support and guidance wherever required, and expressing full confidence in his student's ability to lead the research. Dr. Fraser allowed me to produce a thesis that I am truly proud of.

I also thank the members of my thesis committee, Dr. James Grant and Dr. Selvadurai Daya Dayanandan, for taking the time to meet with me to discuss my research and provide feedback.

Finally, a big thank you to the following organizations and partners for funding the FISHES project and my thesis project: Genome Canada, Genome Quebec, Mitacs, Niskamoon Corporation, Cree Nation Government, and the Eeyou Marine Region Wildlife Board. Thank you to Concordia University for funding the first year of my MSc through the Top Concordian Graduate Entrance Scholarship.

Contribution of Authors

The authors who worked on this project are Badrouyk Chamlian[1], Sozos Michaelides[1], Hari Won[2], Hyung-Bae Jeon[3], Julie Gibelli[1], Pamela MacLeod[4], Hubert Petawabano[4], and Dylan J. Fraser[1].

[1]Department of Biology, Concordia University, 7141 Sherbrooke St. West, Montreal, QC, H4B 1R6, Canada

[2]Honam National Institute Biological Resources, 99 Gohadoan-gil, Mokpo 58762, Republic of Korea

[3]Department of Aquatic Life Medicine, College of Ocean Science and Technology, Kunsan National University, Gunsan 54150, Republic of Korea

[4]Cree Nation of Mistissini

Conceptualization: DJF, PM

Local coordination of sampling efforts: PM, HP

Data generation: BC, HW, HBJ

Software: BC, JG, SM, HBJ, HW

Analysis and interpretation of data: BC, SM, JG

Writing – original draft preparation: BC

Writing – review and editing: BC, DJF

Supervision: DJF

Project administration: DJF

Funding acquisition: DJF

Table of Contents

List of Figures

List of Tables

List of Supplementary Materials

APPENDIX I – Supplementary Figures and Tables

Table S1. Effective population sizes of source populations. N_e estimates were calculated using a Pcrit (lowest allele frequency) of 0.05. Confidence intervals were calculated with the JackKnife method on loci. All values were calculated with the filtered panel of 393 SNPs and rounded to the nearest integer.

Figure S1. Rubias z-score distributions of Mistasiniishish (Albanel) Lake mixed-stock assignments. The left panel (A) shows z-score distributions of assignments made with the reference dataset of 5 known sources. The right panel (B) shows z-score distributions of assignments made with the second reference dataset of 6 sources (5 known + Albanel_distinct).

Figure S2. STRUCTURE plot of the 5 known sources followed by all mixed-stock individuals across the 3 lakes. The x-axis denotes source individuals and mixed stock individuals from each of the 3 lakes: PAP = Papas, CHE = Cheno, RUP = Rupert, TEM = Temiscamie, BDR = Bordeleau. The y-axis represents the proportion of membership to each cluster (Q-values) estimated by STRUCTURE. For each K, the run with the highest log-likelihood (out of 10 runs) is presented. At K=6 and above, a distinct cluster (blue color) is visible in the Mistasiniishish (Albanel) mixed-stock individuals that does not match any of the source clusters. These plots were generated using pophelper (Francis, 2017).

Figure S3. PCA of Mistasiniishish (Albanel) mixed stock individuals with sources. The legends on the right of the plots denote source individuals and mixed-stock individuals: PAP = Papas, CHE = Cheno, RUP = Rupert, TEM = Temiscamie, Alb_distinct = Albanel_distinct, Alb = Mistasiniishish (Albanel) Lake mixed-stock. The plots shown are for (A) Mistasiniishish (Albanel) Lake mixed-stock individuals with all known sources excluding Bordeleau, (B) Same plot as in (A) but with mixed-stock individuals identified in STRUCTURE as pure and distinct (Q-values > 0.80) labeled in blue. Bordeleau was excluded from these plots because it was clustering with Rupert on the PC1 axis, which was uninformative.

Figure S4. Accuracy of individual assignments made with Rubias using the leave-one-out method. The x-axis denotes the 6 source populations (BDR = Bordeleau, CHE = Cheno, PAP = Papas, RUP = Rupert, TEM = Temiscamie, Alb_distinct = Albanel_distinct). The y-axis represents the posterior means of group membership (PofZ) for individuals assigned to their population of origin in simulated random population mixtures. The dashed line represents the threshold PofZ value (0.80). Simulations were conducted with 1000 replicates, 2000 simulated mixture individuals and the default value for mixing proportion (Dirichlet distribution, alpha = 1.5). The number of simulated individuals assigned to each population is indicated above each box. Mean assignment accuracy across all 6 sources was 94.9% (PofZ = 0.949).

Figure S5. Changes in Mistassini Lake's Catch-Per-Unit-Effort for brook trout over 30 years. The blue bars represent fishing effort, the orange bars represent catch, the grey broken line represents catch-per-unit-effort (CPUE), and the dashed red line is the linear trendline fitted to CPUE.

Figure S6. Changes in Mistasiniishish (Albanel) Lake's Catch-Per-Unit-Effort for brook trout over 30 years. The blue bars represent fishing effort, the orange bars represent catch, the

grey broken line represents catch-per-unit-effort (CPUE), and the dashed red line is the linear trendline fitted to CPUE.

Figure S7. Changes in Waconichi Lake's Catch-Per-Unit-Effort for brook trout over 30 years. The blue bars represent fishing effort, the orange bars represent catch, the grey broken line represents catch-per-unit-effort (CPUE), and the dashed red line is the linear trendline fitted to CPUE. Linear regression: $R^2 = 0.1563$, p-value = 0.0277.

Figure S8. Evanno plots for the STRUCTURE run of 6 source populations (5 known + 1 revealed through mixed-stock analysis). The log probability of each run (A) and the derivatives of the log probability (B, C) are used to calculate ΔK (D). The K value yielding the highest ΔK is the best K value according to the Evanno method. These plots were generated using pophelper (Francis, 2017).

Figure S9. Catch-per-unit-effort in Cheno, Papas, and Rupert source rivers at 3 time periods. Values for catch-per-unit-effort (CPUE) were calculated based on the number of brook trout caught per 8-hour day of fishing per angler. The colours of the boxes correspond to the source river (yellow = Cheno; orange = Papas; green = Rupert). The line inside each box indicates mean CPUE for the specified source river in the specified time period. The time periods are shown on the x-axis.

Introduction

Global biodiversity loss has been a major public concern since the turn of the century. Threats to biodiversity were, until recently, viewed mostly at the species level by quantifying extinction rates. However, it has been shown that population extirpation rates greatly exceed extinction rates of species globally (Ceballos et al., 2017). Recent studies have thus highlighted the importance of diversity at the population level, known as intraspecific variation (Des Roches et al., 2017, 2021; Leigh et al., 2019; Mimura et al., 2017).

The preservation of genetic diversity among and within populations is fundamental to species persistence and essential for fishery stability (Bradbury et al., 2018; Hilborn et al., 2003; Schindler et al., 2010). Population diversity reduces annual variability in fishery productivity and enhances the adaptability of species via portfolio effects (Des Roches et al., 2021; Schindler et al., 2010). As such, the sustainability of fisheries relies on effective monitoring and maintenance of harvested fish populations (Hilborn et al., 2003). This goal is greatly facilitated by resolving the genetic structure that underlies diversity within harvested species (Manel et al., 2005; Shaklee et al., 1999; Ward, 2000), and by monitoring the relative contribution of genetically distinct populations to mixed-stock harvests (Andvik et al., 2016; Shaklee et al., 1999).

Determining the stock composition of mixed harvests is crucial for the stability of mixed fisheries, as small populations may suffer from overexploitation if harvest contributions are unknown (Bradbury et al., 2016). Furthermore, a population that previously contributed little to fishing harvest may become a major contributor in response to environmental change (Hilborn et al., 2003). Resolving harvest contributions of source populations within commercial, recreational, and subsistence (CRS) fisheries is therefore essential for the stability of ecosystem services, food security in local communities, the economy, and the livelihoods of local peoples (Des Roches et al., 2021; Nesbitt & Moore, 2016; Schindler et al., 2010).

Advances in molecular techniques over the past decade have greatly improved the efficiency and resolution of genetic studies for conservation and management (Allendorf, 2017; Y.-H. Li & Wang, 2017; Seeb et al., 2011). New sequencing approaches have led to a shift from traditional genetic markers (e.g. microsatellites) to genome-wide data being used to answer research questions, giving rise to the current era of conservation genomics (Bernatchez et al., 2017; Meek & Larson, 2019; Supple & Shapiro, 2018). Amplicon sequencing techniques such as GTseq (Genotyping-in-Thousands by sequencing) are cost-effective and provide SNP-based genomic data that can be used for genetic stock identification (GSI) in mixed-stock fisheries (Campbell et al., 2015; P. Li et al., 2023).

Salmonids are a family of fishes with great socioeconomic and cultural value in the northern hemisphere (Kilduff et al., 2015; Lundrigan et al., 2005; Penaluna et al., 2016; Torrissen et al., 2013). Northern CRS fisheries are commonly mixed-stock and contain multiple salmonid species, each composed of several genetically distinct populations (Beacham et al., 2019; Waples et al., 1990). Most migratory salmonids breed and spawn in freshwater streams and then migrate to marine or lake feeding areas for growth and maturation (Fraser & Bernatchez, 2005a; Saunders, 1981; Schaffer & Elson, 1975). Mature individuals then return to their natal streams for reproduction, thus completing their life cycle (Fraser & Bernatchez, 2005a; Schaffer & Elson, 1975). This philopatric behaviour can give rise to numerous genetically distinct populations of the same species within a coastal area or lake, which can mix temporally as specific stages of their life cycles overlap with harvest (Beacham et al., 2019; Fraser et al.,

2006; Fraser & Bernatchez, 2008; Harris et al., 2016; Taylor, 1991). Such research has contributed significantly to our understanding of salmonid populations and movements. However, it has largely focused on anadromous populations, with much less investigation into population structuring and mixed-stock dynamics of lake-migratory salmonids (e.g. Tessier & Bernatchez 1999; Fraser & Bernatchez 2005a; Ferguson et al. 2019).

The Cree Nation of Mistissini owns and manages the Albanel-Mistassini-Waconichi (AMW) Wildlife Reserve. This is the largest reserve in Quebec, harbouring Lakes Mistassini, Mistasiniishish (Albanel), and Waconichi, which support CRS fisheries for several key species including multiple salmonids. Fraser & Bernatchez (2005a) inferred the spatial distribution and harvest contribution of distinct lake-migratory brook trout (*Salvelinus fontinalis*) populations in Mistassini Lake from samples collected in 2000 and 2001. The study used 10 microsatellite loci to assign individual brook trout caught in the lake to their population of origin (Fraser & Bernatchez, 2005a). A more recent study by Fraser et al. (2013) reported a potential demographic decline in the lake's outflow population over the course of a decade, based on samples collected in 2011. The two inflow populations were found to be largely temporally stable in their demography and life history over the same period (Fraser et al., 2013). Given that another decade has passed since this last study, a contemporary analysis of Mistassini Lake's brook trout populations is needed to monitor their temporal stability or lack thereof. Changes in these populations over time have implications for fisheries management, as harvest contributions could be altered by demographic shifts. The use of microsatellites in past studies further warrants a temporal re-assessment of Mistassini Lake, as current genomic tools provide greater resolution to demarcate source populations and detect previously unknown sources.

In this paper, a GTseq panel of 393 SNPs, developed specifically for the focal brook trout populations of study, was used to assign individuals caught in Mistassini, Mistasiniishish, and Waconichi lakes in 2020, 2021 and 2022 to their populations of origin, based on allele frequencies generated for each known source population from spawning and pre-spawning individuals sampled in 2020 and 2021. We also conducted population structure analyses to test for the presence of unknown source populations, as well as mixture analysis to determine the relative contributions of distinct genetic stocks to the overall harvest of brook trout in the three lakes. Finally, we compared spatial distributions of migration to lake feeding areas in Mistassini Lake over a 20-year period to test for stability over time. Defining mixed-stock harvest contributions in these fisheries will shed light on the relative importance and productivity of each population, allowing community-based management to decide on timing and areas of harvest and refine fishing quotas accordingly. This will ensure fishery sustainability and food security for the local Cree who fish year-round for subsistence. In addition to subsistence fishing, many Cree make their living from being fishing guides for sport fishers in their outfitting camps. Long-term sustainability of these fisheries will thus also ensure income stability in Cree communities.

Materials & Methods

Study Sites

Mistassini Lake (2335 km^2; 50.72°N, 73.86°W) is the largest lake in the AMW reserve and the largest natural lake in Quebec (Figure 1). Postglacial in origin and oligotrophic, the lake has been documented to be home to at least three genetically distinct populations of lake-migratory brook trout characterized by their distinct feeding and breeding areas within the lake and its rivers (Fraser & Bernatchez, 2005b, 2008). One population breeds in the lake's outflow (Rupert), and the other two populations breed in the two main inflows (Cheno and Pepeshquasati) (Pepeshquasati is hereafter referred to as Papas) (Fraser et al., 2013; Fraser & Bernatchez, 2005b, 2008). The fall spawning of adults and rearing of juveniles in these rivers are followed by seasonal migrations to lake feeding areas where individuals mature and grow, after which they return to their natal rivers to spawn and complete their life cycle (Fraser & Bernatchez, 2005a, 2008). As the migrations occur in freshwater only (lake-river or river-lake), these populations are potamodromous (Ferguson et al., 2019). The outflow population undergoes allacustrine migration, while the inflow populations undergo lacustrine-adfluvial migration (Ferguson et al., 2019). Such a lake-migratory life history is relatively rare for the species, with analogous populations only described in a few other large lakes (e.g. Lake Superior, Lake Nipigon) (Mucha & Mackereth, 2008; Ridgway, 2008; Schreiner et al., 2008; Stott et al., 2010).

Mistasiniishish (Albanel) Lake (445 km^2; 51.05°N, 73.09°W) is another large, remote postglacial lake located just east of Mistassini Lake (Figure 1). There have been very few studies on Mistasiniishish brook trout, and so far, the lake is scientifically known to contain a single source population that breeds in the inflow river Temiscamie (Flick, 1977). Mistasiniishish drains into Mistassini Lake through a waterfall that reportedly prevents fish from Mistassini from migrating into Mistasiniishish (Flick, 1977). However, a past study found a closer genetic relationship between northeastern Mistassini brook trout and Mistasiniishish brook trout than that between inflow and outflow brook trout within Mistassini Lake (Fraser & Bernatchez, 2005b). In addition, the local Cree reportedly catch brook trout in the lake's outlet.

Waconichi Lake (82 km^2; 50.15°N, 74.00°W) is the smallest of the three main lakes in the AMW reserve, located south of Mistassini Lake (Figure 1). There have been no studies on the fish inhabiting Waconichi Lake, and it is thought to contain only one brook trout population that spawns in the inflow river Bordeleau. Like Mistassini and Mistasiniishish, Waconichi Lake is also fished for subsistence by the local Cree community.

Sample collection

Mixed stock samples were collected throughout the lakes over three field seasons (May 21-September 3, 2020-2022). Source samples were collected from spawning brook trout in source rivers during the known spawning seasons (September 10-October 20) of 2020 and 2021. Samples consisted of adipose fin tissue collected non-lethally from individual brook trout caught in Mistassini, Mistasiniishish and Waconichi Lakes, either by local Cree partners, recreational fishers, or Concordia personnel. To collect a sample, the adipose fin (located in between the dorsal and caudal fins) was cut and placed in a sample tube filled with 95% ethanol and labeled with an individual fish identification number. For each collected sample, the location where the fish was caught (GPS coordinates or proximity to an island/bay/river), its total length and fork

length, mass, and sex (for spawning individuals) were also recorded. Overall, 645 mixed stock samples were collected from Mistassini Lake, 173 from Mistasiniishish Lake, and 78 from Waconichi Lake. Fifty source samples were collected from the Papas (PAP) river, 30 from Cheno (CHE), 43 from Rupert (RUP), 36 from Temiscamie (TEM), and 8 from Bordeleau (BDR). See Figure 1 for sampling locations.

DNA extraction

DNA extraction from adipose fin tissue samples was done using Qiagen DNeasy Blood and Tissue kits for source individuals and a salt extraction protocol (based on Aljanabi, 1997 with minor modifications) for mixed stock individuals. See Appendix I for protocol details.

GT-seq library preparation

Extracted DNA concentrations were quantified with Qubit (Invitrogen broad range kit). Samples with concentrations above 30ng/uL were diluted to 15ng/uL to avoid over-amplification due to high concentration. Final concentrations before library preparation ranged from 2 to 25 ng/uL. The DNA samples were then prepared for sequencing following the GT-seq library preparation protocol described in Campbell et al. (2015), with minor modifications. The GT-seq panel adopted was designed as described in Beemelmanns et al (submitted). The panel markers were derived from distinct brook trout populations in Mistassini Lake and along the James Bay coast of northwest Quebec. The Mistassini Lake markers were selected to maximize genetic differentiation (Fst) between the Cheno and Papas populations. The final panel contained 449 SNPs.

A total of 1063 DNA samples were aliquoted into twelve 96-well plates for library preparation. Two PCR steps (PCR1 and PCR2) were performed on each plate. PCR1 consisted of multiplex PCR amplification of target loci (SNPs) based on GT-seq primers. The PCR1 products were diluted 20-fold, and 3 uL from each well of the PCR1 plate was transferred into each well of a new 96-well plate. Barcode sequences and Illumina capture sites (i5 and i7) were then added to the plate and PCR2 was performed. PCR2 amplified the barcodes and capture sites, allowing to identify individual samples based on the unique combinations of i5 and i7 primers added to each well of the plate. PCR2 products were then pooled into 1.5 mL tubes, yielding 12 pooled libraries. Aliquots from each pooled library were then purified using AMPure XP beads for double-sided size selection. Each purified library was then quantified using Qubit (Invitrogen broad range kit), normalized to 4 nM, pooled (12 purified libraries in 1 tube), and sequenced with MiSeq v3 (paired-end, 2 x 75 bp) at the Institute of Integrative Biology and Systems (IBIS, Université Laval).

Data Preparation and filtering

Individual-specific fastq files were generated from the sequencing done at IBIS. These files were processed through a simple bioinformatics pipeline developed by Fraser lab personnel. Individual fastq files were genotyped with a custom perl script (AmpliconReadCounter.pl) using allele-specific primer probe sequences listed in a text file (bt_GTseq449_PrimerProbefile.txt).

The quality of SNP data and sample data was checked using GTscore (McKinney et al., 2020). Twenty-two SNPs were filtered out for having genotype rates below 50% (i.e. SNPs with missing data in more than half of all samples), 21 with minor allele frequency < 0.015, and 13 identified as paralogous (ploidy = 4). In addition, 8 samples were filtered out due to having genotype rates

below 60% (i.e. individuals with missing data for more than 40% of SNPs in the panel), 2 were filtered out due to elevated heterozygosity (>0.40), 12 were filtered out for having no location information, and 5 were filtered out due to potential contamination. Overall, the filtered dataset contained 1036 individuals and 393 SNPs to be used for analysis.

We tested the filtered SNP dataset (393 SNPs) for departure from Hardy-Weinberg Equilibrium (HWE) in the source individuals using snpR (Hemstrom & Jones, 2023). We applied a Bonferroni correction to control for multiple testing. None of the SNPs in the filtered dataset significantly departed from HWE across all source populations, and thus all SNPs were kept for analysis.

The quality of the filtered dataset (1036 individuals; 393 SNPs) was checked with GTscore by calculating the average sample genotype rate, average SNP genotype rate, and the average read depth per SNP. The sample genotype rate is the proportion of SNPs that yielded data for a given individual. The SNP genotype rate is the proportion of individuals with data for a given SNP. The average read depth per SNP is the average number of reads across all individuals for a given SNP (e.g. a SNP with an average read depth of 30 has, on average, 30 reads for each individual in the dataset). The filtered dataset had an average sample genotype rate of 93%, average SNP genotype rate of 93%, and an average read depth per SNP of 40.8.

Population structure and unsampled population testing

Population structure among sources was assessed using snpR (Hemstrom & Jones, 2023) and STRUCTURE (Pritchard et al., 2000). Pairwise fixation index (Fst) values were calculated with snpR between each source population pair to quantify the extent of genetic differentiation. The Fst calculations were performed using the Genepop method (ROUSSET, 2008) and 9,999 bootstraps. Additionally, Principal Component Analysis (PCA) was performed with snpR to visualize population clustering among individuals. PCA plots were generated for i) all individuals, ii) source individuals only, iii) source individuals and Mistassini mixed stock individuals, iv) source individuals and Mistasiniishish (Albanel) mixed stock individuals, v) source individuals and Waconichi mixed stock individuals.

We used STRUCTURE as an additional method of visualizing population structure. STRUCTURE assigns individuals to population clusters based on their allele frequencies at each locus. The program assumes linkage equilibrium and Hardy-Weinberg equilibrium, and it infers individual membership to clusters using a Bayesian framework. If an individual's genotype contains alleles from two or more clusters, then the individual is defined as admixed and assigns jointly to several clusters. The proportion of membership to each cluster is calculated as a Q-value ranging from 0 to 1. Individuals are assigned to K clusters, where K is a number defined by the user. We ran STRUCTURE on i) the five known sources only, ii) all individuals, and iii) the five known sources plus a sixth, unsampled source that was detected within the mixed-stock individuals of Mistasiniishish Lake. The five known sources were run first to see how well they differentiated/clustered among themselves. Mixed-stock individuals were then run with source individuals to view how they related to known populations and to test for unknown populations that may form clusters in mixed-stock individuals distinct from the source clusters. A distinct cluster was detected within Mistasiniishish mixed-stock individuals from this run. A distinct genetic signal was also detected in the PCA plot of Mistasiniishish mixed-stock individuals with known sources. The individuals generating this distinct signal were confirmed to be the same across STRUCTURE and PCA, and so they were added as a sixth source

population for the third STRUCTURE run. See discussion for details on detecting the unsampled population.

For the five known source dataset, 100,000 burn-ins followed by 200,000 MCMC repeats were used to run K=1 to K=6 with 10 replicates for each K. For the all-individuals dataset, 100,000 burn-ins followed by 200,000 MCMC repeats were used to run K=2 to K=7 with 5 replicates for each K. The admixture model with correlated allele frequencies was used to estimate admixture proportions within individuals.

After adding the sixth source population (detected in Mistasiniishish Lake), a final STRUCTURE run was performed with only source individuals from all 6 sources (100,000 burn-ins followed by 200,000 MCMC repeats were used to run K=1 to K=7 with 5 replicates for each K). The admixture model with correlated allele frequencies was used to estimate admixture proportions within individuals.

Mixed-stock assignments

We used Rubias (Moran & Anderson, 2019) to assign mixed-stock individuals to their population of origin, based on the allele frequencies of each known source population generated from source samples collected near spawning grounds in 2020 and 2021. Rubias uses Bayesian inference via MCMC to assign individuals to their population of origin based on individual posterior means of group membership (PofZ). The accuracy of assignments made with source allele frequencies was tested using the function “assess_reference_loo”. This function uses a leave-one-out (LOO) method to simulate random population mixtures and then calculate the likelihood of each source individual belonging to its known population after removing its genotype from the population’s allele counts. The simulations were conducted with 1000 replicates, 2000 simulated mixture individuals and the default value for mixing proportion (Dirichlet distribution, alpha = 1.5).

The Rubias function “infer_mixture” was used to assign mixed stock individuals to populations of origin using the source allele frequencies as baseline data. The function was performed with 200,000 MCMC iterations, 40,000 burn-ins and 1,000 bootstraps. The function was first run on all mixed stock individuals combined across the 3 lakes, with all source individuals used as the baseline data. Each lake was then run separately to generate mixed stock harvest contributions per lake. For all runs, the baseline data included all source individuals from all 6 populations (5 known + 1 unsampled but detected in Mistasiniishish) because some fish were suspected to move between lakes, particularly downstream from Waconichi or Mistasiniishish to Mistassini than vice versa. A total of 853 mixed stock individuals across the 3 lakes were assigned to their population of origin. A total of 845 of the assignments had PofZ values above the threshold (>0.8), and these were kept for further analyses.

In addition to PofZ, Rubias calculates individual z-scores from the log-likelihood for each assignment. The distribution of z-scores can be examined to detect mixed-stock individuals originating from sources not present in the reference dataset (i.e. unsampled sources). For each lake, we compared mixed-stock z-score distributions to source z-score distributions. Source z-scores were calculated using the “self_assign” function. Our expectations were that mixed-stock assignments yielding z-scores lower than the source z-score minimum may reflect individuals originating from unsampled sources (Horne et al., 2023; Moran & Anderson, 2019).

Population spatial distributions and assessment of temporal stability within Mistassini Lake

Mixed stock individuals were mapped onto their corresponding lakes using the software QGIS. Coordinates (latitude, longitude) were used to map each individual to the location where it was caught. For individuals with only physical descriptions of location (i.e. proximity to a landmark), coordinates were determined manually based on the landmark's location on the map. Multiple individuals caught in the same location were grouped into pie-charts depicting relative proportions of population assignments in that location.

Mistassini mixed stock individuals were mapped onto the lake using QGIS and labeled with their population of origin based on Rubias assignment results. Three maps were first generated, one for each year of contemporary sampling (2020, 2021, 2022). Sample numbers for 2020 were too low (n = 52) compared to 2021 and 2022 (n = 223 and n = 329, respectively), so we removed this sampling year from the comparison. The two maps (2021, 2022) were then used to assess temporal stability in spatial distribution of Mistassini brook trout over two decades by comparing with (Fraser & Bernatchez, 2005a). Only individuals assigned to RUP, CHE, and PAP were included in the temporal comparison, as Fraser & Bernatchez (2005a) did not include other possible sources. Fraser & Bernatchez had divided the lake into 9 sectors to conduct their assessment, and we evaluated the same 9 sectors for the best comparison. The lake sectors were overlaid on the map of contemporary assignments in QGIS. For each sampling year, population proportions were calculated for each sector by dividing the number of individuals in a given sector assigned to a population by the total number of individuals assigned to that population for that year. CHE and PAP assignments were combined into a single population called "inflows" before calculating the proportions per sector (see Discussion).

The spatiotemporal distribution of populations was then tested by performing a MANOVA in R using the function "manova". The dependent variables were the population proportions (RUP, inflows), and the independent variables were lake sector (sector 1 to sector 9) and sampling year (2000, 2001, 2021, 2022). We conducted 3 tests: i) temporal stability between contemporary years only (2021 vs 2022), ii) temporal stability across all sampling years (2000 vs 2001 vs 2021 vs 2022), iii) temporal stability across time periods (historical vs contemporary). For the last test (time periods) we pooled contemporary years into one time period (2021-2022) and historical years into one time period (2000-2001), so the resulting comparison was for a single change in time period (20 years). The pooling of years within each time period was valid because our first test showed that there was no significant variance over time between 2021 and 2022. As for pooling 2000 and 2001, Fraser & Bernatchez (2005b) had reported no significant variance over time between those two years.

Results

Population structure of five known sources and unsampled population testing

All 5 known source populations (BDR, CHE, PAP, RUP, TEM) across the 3 lakes showed statistically significant genetic differentiation ($P<0.05$) based on pairwise Fst values (Figure 2A). The lowest Fst value (0.016) was between CHE and PAP, suggesting low differentiation between these two populations. Despite this low differentiation, PCA clustering supported a distinction between CHE and PAP as evidenced by individual (though partly overlapping) clusters for these two populations (Figure 2B). STRUCTURE analysis of source populations further supported a differentiation between CHE and PAP, as they were sorted into distinct clusters at K = 5 (Figure 2D). We found closer genetic relationships between TEM, CHE, and PAP than between CHE, PAP, and RUP despite the latter being found in the same lake. This was evidenced by pairwise Fst values (Figure 2A), PCA clustering (Figure 2B), and STRUCTURE separating RUP from CHE/PAP/TEM as early as K = 2 (Figure 2D).

We then ran STRUCTURE on the 5 known sources (BDR, CHE, PAP, RUP, TEM) combined with all mixed stock individuals from the three lakes to test for unknown population clustering which may suggest an unsampled source. At K6 and above, we found the formation of a distinct genetic cluster within Mistasiniishish (Albanel) mixed stock individuals (Figure S2). The cluster consisted of 16 pure individuals (Q-values > 0.80) generating a genetic signal distinct from all the clusters present in the 5 source population individuals. It should be noted that, while CHE and PAP were not fully divided into 2 clusters in these runs, there was the formation of 2 distinct genetic signals within this population pair. The pure cluster formed within the Mistasiniishish mixed stock individuals was completely distinct from these 2 clusters in the CHE/PAP pair.

The PCA plot of Mistasiniishish mixed stock individuals with 5 sources also showed several individuals forming an independent cluster that was distinct from all source clusters (Figure S3A). To see if these were the same individuals as those forming a distinct cluster in STRUCTURE, we labelled those 16 pure distinct (Q-values > 0.80) individuals as “Alb_distinct” and re-ran the PCA. These were confirmed to be the same individuals forming a distinct cluster in the PCA (Figure S3B).

Mistasiniishish Lake’s mixed-stock z-score distribution of Rubias assignments revealed drastic deviations from the source distribution (Figure S1A). All 16 of the mixed-stock individuals identified in STRUCTURE and PCA had z-scores that were lower than the minimum value in the source z-score distribution (leftmost tail of mixed-stock distribution, Figure S1A).

We ruled out the possibility of missing data causing these individuals to appear as outliers. With a higher filtering threshold (>80% sample genotype rate, i.e. <20% missing data per sample), all individuals generated the same distinct signal on STRUCTURE and PCA, as well as having Rubias z-scores lower than the minimum value in the source z-score distribution.

Given the validation across all three analyses, we created a sixth source population “Alb_distinct” (n=16) and re-ran population structure analyses.

Population structure of 6 sources

The unsampled population (Alb_distinct) showed statistically significant genetic differentiation ($P<0.05$) from all other sources based on pairwise Fst values (Figure 2A). Its lowest Fst value (0.064) was with TEM within Mistasiniishish Lake. PCA of 6 sources showed a distinct cluster for this population that did not overlap with any of the other 5 sources (Figure 2C). STRUCTURE analysis of 6 source populations further validated this distinction as Alb_distinct individuals formed a distinct cluster at K = 6 (the best K suggested by the Evanno method) (Figure 2E) (See Figure S8 for Evanno method results of STRUCTURE analysis). Overall, these results suggested the presence of 6 source populations across the 3 lakes.

Mixed-stock assignments and harvest proportions

We used Rubias with the reference dataset of 6 sources to assign mixed-stock individuals from all 3 lakes to their population of origin. Simulated mixtures using the leave-one-out method yielded an average assignment accuracy of 94.9% across all six source populations (0.8% improvement from the 5-source reference dataset) (Figure S4). Importantly, the z-score distribution of Mistasiniishish mixed-stock assignments was greatly improved and overlapped with the source z-score distribution when applying the 6-source dataset vs. the 5-source one (Figure S1B). The final dataset with 6 sources was then used for all subsequent analyses.

In Mistassini Lake, 640 mixed stock individuals were assigned to their population of origin by Rubias. 634 of these assignments were above the confidence threshold (PofZ > 0.80), with 630 of them exceeding the threshold of PofZ > 0.90. The six individuals with assignments below the confidence threshold were assigned to CHE (PofZ: 0.51, 0.67, 0.72), PAP (PofZ: 0.51, 0.79) and Alb_distinct (PofZ: 0.59). Overall, 331 individuals (52.2%) assigned confidently to PAP, 269 (42.4%) to RUP, 15 (2.4%) to CHE, 15 (2.4%) to TEM, and 4 (0.6%) to Alb_distinct (Figure 3A).

In Mistasiniishish Lake, 133 of 135 mixed stock individuals were assigned to their population of origin by Rubias with an assignment confidence PofZ > 0.90; only two individuals did not meet the minimum PofZ threshold of 0.80 (TEM, PofZ: 0.56; PAP, PofZ: 0.66). Overall, 121 individuals (91%) assigned confidently to TEM, 11 (8.3%) to Alb_distinct, and 1 (0.7%) to PAP (Figure 3B). However, the 16 individuals used as the 6^{th} reference population for Rubias assignments were originally mixed stock individuals. Therefore, the true number of Alb_distinct individuals in the mixed-stock harvest of Mistasiniishish Lake is 27 out of 149 (18.1%). With 149 total mixed-stock assignments, the respective proportions of TEM and PAP assignments are 81.2% and 0.7%.

In Waconichi Lake, all 78 mixed stock individuals were assigned to the BDR source population with the highest possible confidence (PofZ = 1) (Figure 3C).

Population spatial distributions and assessment of temporal stability within Mistassini Lake

In Mistassini Lake, PAP individuals were present in all areas of the lake, being predominately located along the island chain in the lake's center as well as in the south of the lake more than 100km away from the mouth of the Papas River (Figure 3). RUP individuals were also present in most areas, although they were predominately found closer to the mouth of the Rupert River and to a lesser extent along the island chain. CHE individuals were rare, being located mostly along the island chain and the southern shoreline, like PAP.

Results from MANOVAs supported a heterogeneous and temporally stable spatial distribution to lake feeding areas among Mistassini Lake populations (Figure 5). Outflow (RUP) brook trout predominated along the northern shoreline (sector 2: 0.65-0.76) and were found to a lesser extent along the central island chain (sector 5: 0.16-0.19). Inflow (PAP, CHE) brook trout predominated along the central island chain (sectors 5 & 6: 0.47-0.55) and were found to a lesser extent along the southern shoreline (sectors 7-9: 0.05-0.13). Results from MANOVA tests showed that this differential spatial distribution was temporally stable across 3 different time scales: i) 2021 vs 2022, ii) 2000 vs 2001 vs 2021 vs 2022, iii) historical (2000-2001) vs contemporary (2021-2022). For each MANOVA, variance in the spatial distribution of populations was very significant (Wilk's $\lambda < 0.05$; $P << 0.001$) while variance in time was not significant (Wilk's $\lambda > 0.78$; $P > 0.4$). The variance explained by space (i.e. lake sector) was over 20-fold greater than the variance explained by time (*F* values: sector = 9.68; time = 0.47). See Table 1 for MANOVA results.

In Mistasiniishish Lake, TEM individuals were present in all sampling areas (Figure 3). Alb_distinct individuals were concentrated in the southern end of the lake. For Waconichi Lake, many individuals lacked specific location information. The resulting spatial map for Waconichi individuals did not cover much of the lake, with all individuals being located in the south (Figure 3). The presence of multiple TEM individuals in Mistassini and a single PAP individual in Mistasiniishish suggests asymmetric migration of brook trout from Mistasiniishish to Mistassini.

Discussion

Using a GT-seq panel designed for discriminating brook trout populations in the study region, we found substantial evidence for the presence of six genetically distinct populations across the three lakes studied. The large scale of these lakes (50-130km in length spanning 1.3 degrees of latitude) suggests that populations may have evolved through adaptive migratory divergence to differential spawning and feeding habitats. Multiple individuals in Mistassini Lake were caught more than 100km away from their natal river mouth, highlighting their ability to migrate long distances within the lake. Several brook trout originating from Temiscamie River in Mistasiniishish Lake were also caught in Mistassini Lake (>80km away from their natal river mouth), revealing inter-lake movement through a waterfall previously thought to be a dispersal barrier.

Sympatric populations were observed in Mistassini and Mistasiniishish Lakes; Waconichi Lake contained only one population. The most pronounced divergence between sympatric populations was in Mistassini Lake, the largest and deepest of the three lakes. This lake's populations exhibited distinct spatial distributions to feeding areas, in contrast to Mistasiniishish Lake's populations which showed strongly overlapping distributions. The heterogeneous spatial distribution of Mistassini Lake's populations was temporally stable over 20 years. The small number of rivers used for spawning by brook trout in these lakes emphasizes their critical role for the species' persistence in large lakes. While there may be additional small streams or creeks entering these lakes used occasionally as spawning areas by lake-migratory brook trout, at the scale of entire lakes, their productivity is likely negligible compared to the source populations surveyed using genomics in this study. For example, a small feeder creek in the northern end of Mistassini Lake has been observed to contain a few pairs of mature brook trout during the spawning season (Dylan J. Fraser, personal communication). However, these pairs likely originate from the Papas or Rupert rivers based on the proximity of the feeder creek to the river mouths (Dylan J. Fraser, personal communication). The overall genomic patterns between mixed-stock individuals and source populations also imply that any small streams used for occasional spawning are not genetically distinct.

Brook trout with lacustrine-adfluvial or allacustrine migratory life histories are exceedingly rare, known to occur in only a few other North American lakes (e.g. Lake Superior, Lake Matamek, and Lake Nipigon) (O'Connor & Power, 1973; Robillard et al., 2011; Stott et al., 2010), and more generally, a literature gap exists in the study of potamodromous migrations in other salmonids (Ferguson et al. 2019). Our study adds two new lakes to the literature for lake-migratory brook trout, namely Mistasiniishish (Albanel) Lake and Waconichi Lake. We found evidence for strong natal philopatry in Mistassini Lake brook trout, based on strong genetic divergence and low admixture between allacustrine (Rupert) and lacustrine-adfluvial (Papas, Cheno) populations. The two lacustrine-adfluvial populations (Papas, Cheno) exhibited lower genetic differentiation and higher admixture between source individuals, suggesting consistent gene flow between these tributaries. A similar relationship has been found among coaster (i.e. lake-migratory) brook trout in Nipigon Bay, Lake Superior, with coasters being produced by tributary populations and also acting as vectors for gene flow between tributaries (D'Amelio et al., 2008). A radio telemetry study on coaster brook trout in Nipigon Bay found the maximum distance travelled by individual brook trout to be 46 km (Mucha & Mackereth, 2008). Although our study did not use telemetry data, mixed-stock individuals in Mistassini Lake assigned to their population of origin were caught more than 100 km away from their source river. Together, our results shed light on the

scale of possible population structure and spatial distribution for the unique life history of lake-migratory brook trout. Below, we report our findings for each lake in terms of their population structure, harvest contributions of populations, and spatial distributions.

Temporally stable spatial segregation of sympatric populations, unequal harvest contributions, and population trends in Mistassini Lake

The pronounced genetic structure we observed between outflow-spawning (allacustrine) and inflow-spawning (lacustrine-adfluvial) populations in Mistassini Lake was consistent with past studies and likely reflects adaptive migratory divergence. For example, Fraser & Bernatchez (2005a) reported that inflow brook trout migrated longer distances and had more streamlined morphology than outflow brook trout which were short-distance migrants. Inflow brook trout also exhibited earlier spawning timing than outflow brook trout, likely due to faster cooling of the inflows in the fall (Fraser et al., 2004). Similarly, Tessier et al. (1997) studied population structure of sympatrically-occurring land-locked Atlantic salmon in Lake St-Jean, Quebec, and found genetic divergence between populations spawning in different rivers. They suggested that partial reproductive isolation maintaining this divergence may be due to differential timing of spawning migrations (Tessier et al., 1997). Migratory divergence between inflow- and outflow-spawning populations has also been reported in other salmonids (Ferguson & Taggart, 1991; Northcote, 1962).

We also found that inflow and outflow brook trout groups showed significant variance in their spatial distribution to lake feeding areas (Figure 5), consistent with past studies. Outflow individuals predominated along the northern shoreline of Mistassini Lake close to the Rupert River mouth, while inflow individuals predominated along the island chain in the lake's center and along the southern shoreline. This selective migration to feeding areas may be linked to the different coastline habitat types in the littoral zone of the lake (northern shoreline: boulder beaches; island chain/southern shoreline: dolomite cliffs) (Figure 1 in Fraser & Bernatchez 2005a). Interestingly, there is evidence that inflow and outflow brook trout originate from different ancestral groups, and that the different littoral zone habitats each uses are associated with the colonization directions of their respective ancestral groups (Fraser & Bernatchez, 2005b).

Our findings provide further evidence that differential habitat use of sympatric brook trout populations is maintained over time using a MANOVA. Although the 2000-2001 distributions were based on individual assignments using microsatellite loci, they were highly efficient in distinguishing Rupert from inflow (Cheno, Papas) individuals (98.0% mean assignment success) (Fraser & Bernatchez, 2005a). A comparison for Rupert vs. inflow populations combined was appropriate given these high assignment rates. Combining Cheno and Papas as inflows was necessary for the comparison given the high error rate (20%) in historical assignments between these two populations using microsatellites. Temiscamie was excluded from the temporal comparison, as this population was not included in the historical (2000-2001) study and corresponds to a negligible proportion (2.4%) of the contemporary harvest. Temporal stability in genetic structure between sympatric salmonid populations has also been reported for land-locked Atlantic salmon in Lake St-Jean, Quebec (Tessier & Bernatchez, 1999).

Some spatial sectors in Mistassini Lake captured fewer fish in contemporary years despite being specifically targeted for sampling, resulting in them having lower contemporary sample numbers when compared to 2000-2001 (Fraser & Bernatchez 2005; Table 2). These low

contemporary sample numbers could reflect an overall decline in brook trout populations across all of Mistassini Lake. For example, using provincial data (Ministère des Forêts, de la Faune et des Parcs) on Mistassini Lake fishing effort and catch over 30 years, a linear trendline fitted to catch-per-unit-effort (CPUE) shows a clear downward trend, with the CPUE in 2018 approximately a third of what it was in 1987 (Figure S5). CPUE data is not a perfect indicator of abundance, as the proportionality between the two metrics can be highly variable (Harley et al., 2001). Many factors are known to affect catch rates, such as fishing efficiency, targeting, and population dynamics (Maunder et al., 2006). Nonetheless, in the absence of other data, declines in CPUE can provide a warning sign of declines in abundance (Fraser et al., 2013). Despite potential declines, the proportions in spatial distributions of divergent populations were temporally stable over the past 20 years, suggesting that brook trout populations currently occupying the lake are still proportionally migrating to the same areas. If there is a decline in overall abundance, then areas that previously harboured small proportions of larger populations would currently contain fewer fish from declining populations.

The CPUE in Mistassini Lake's source rivers is also declining. Historically (2000-2002), the CPUE (trout captured per 8 hour fishing day) in Cheno, Papas and Rupert was 2.0-3.2, 10.4-12.2 and 3.3-4.0, respectively (Fraser et al., 2004). A decade later, the CPUE for Cheno and Papas was found to be stable since 2002, but the Rupert CPUE was found to be declining (Fraser et al., 2013). The contemporary (2020-2021) CPUE values for Cheno and Papas are less than half of the historical values (Cheno: 1.5 vs 3.2, Papas: 5.6 vs 12.2). We can only speculate on the causes of these changes, however, CPUE declines may not be caused by overfishing, as total fishing effort in Mistassini and Mistasiniishish Lakes has decreased over time (Figures S5, S6). A more likely explanation may be ongoing climate change. For example, in 2011, Cree fishers expressed major concerns about changes in water temperature and water levels of the Rupert River (Fraser et al., 2013). The low contemporary CPUE in all three source rivers thus may be attributed to the continued effects of climate change over another decade since 2011.

The major Mistassini sources are Papas and Rupert, contributing 94.6% of the lake's harvest (52.2% and 42.4% respectively). These harvest proportions highlight two important points: i) Papas and Rupert are vital stocks for Mistassini Lake's brook trout harvest, and ii) Cheno is a very minor harvest contributor. From a total of 634 mixed stock individuals confidently assigned to population of origin, only 15 individuals (2.4%) assigned to Cheno. This is nearly a 9-fold decrease from what has been historically reported. Cheno contributed 17.7% and 20.1% of Mistassini Lake's brook trout harvest in 2000 and 2001, respectively (Fraser & Bernatchez, 2005a). The drastic difference between historical and contemporary proportions may in part be attributed to a higher error rate in historical assignments between Cheno and Papas (20% using 10 microsatellite loci). Another reason for the discrepancy between Papas and Cheno is a metapopulation dynamic inferred from differences in their abundance and CPUE, combined with asymmetric gene flow, with Papas serving as the source and Cheno as a satellite. For example, from a past study, Papas was reported to be more than twice as large as Cheno in effective population size (Papas N_e = 994, Cheno N_e = 435) (Fraser et al., 2004). The same study also found significantly asymmetric gene flow from Papas to Cheno (PAP-to-CHE: m = 0.021, CHE-to-PAP: m = 0.008) (Fraser et al., 2004). With our panel of 393 SNPs, Papas is estimated to be over 7 times larger than Cheno in effective population size (Papas Ne = 667, Cheno Ne = 89). In addition, CPUE in Cheno is much lower than in Papas over time. From the source population sampling conducted for this study in the spawning seasons of 2020 and 2021, the CPUE in

Cheno was nearly 4 times lower than that in Papas (1.5 vs. 5.6) (Figure S9). Fraser et al. (2004) reported a similar difference in CPUE (Cheno: 2.0-3.2, Papas: 10.4-12.2).

CPUE declines in the source rivers of Mistassini Lake, and consequently in the mixed-stock harvest produced by these sources, highlight the dynamics between spawning and feeding habitats in this lake system. Metapopulation dynamics may be common for brook trout inhabiting continuous lake-river systems (D'Amelio & Wilson, 2008). These dynamics have implications for conservation, as fisheries management should prioritize protection of the most productive spawning habitat to prevent population collapse.

Mixed-stock harvest and spatial population structure in Mistasiniishish (Albanel) and Waconichi Lakes

Prior to this study, the Temiscamie River was considered the only source of brook trout in Mistasiniishish (Albanel) Lake (Flick, 1977). Our results support these historical observations in showing that 81.2% of the lake's brook trout harvest originates from Temiscamie. However, we also found that 18.1% of Mistasiniishish mixed stock individuals assigned to a cryptic, unsampled source, denoted as "Alb_distinct" in our study. Our coverage did not extend north of the lake's center due to no fishing effort for brook trout in those areas, as they do not contain suitable brook trout habitat. Based on the locations of capture of these individuals, we suggest three possibilities for the cryptic source location: i) the outlet of Mistasiniishish draining into Mistassini; ii) Richmond River in the south of Mistasiniishish Lake; iii) a tributary of Temiscamie River. The outlet and Richmond River have been mentioned as brook trout spawning locations by the Cree with family traplines in the area (personal communications). An outlet-spawning population would also be supported by PCA clustering of Alb_distinct individuals being halfway between TEM and RUP on the PC1 axis (Figure 4C). Common ancestry of the unsampled source with Rupert (allacustrine) brook trout could explain an adaptation for spawning in a lake outlet. Finally, the source could also be one of the many tributaries of Temiscamie River. For instance, genetically distinct populations of walleye in Mistasiniishish Lake are known to spawn in Temiscamie as well as in the Metawashish river, a Temiscamie tributary (Gibelli, 2023). Given that Metawashish contains similar habitat to Temiscamie, it may be a suitable spawning ground for brook trout. Regardless of the source, the unsampled population is significantly differentiated from the known Temiscamie population (Fst = 0.064). Future studies should sample the aforementioned potential spawning sites in order to resolve the location of this unknown source.

The presence of more than one stock in Mistasiniishish Lake has conservation implications, as its harvested brook trout benefit from portfolio effects. The unknown stock currently contributing little to Mistasiniishish harvest may become a major contributor in response to climate change, as those individuals may have alleles allowing them to adapt to a changing environment.

No past work had been done on Waconichi Lake prior to our study. It was considered to contain just one population that spawns in the Bordeleau River. Given that several hundred brook trout are harvested from this lake every year, mixed-stock analysis was needed to resolve population structure (if any). We found that Waconichi Lake is a closed system for brook trout, with Bordeleau River being the only source of brook trout in this lake. No mixed-stock individuals in the other two lakes assigned to the Bordeleau source population, and no mixed-stock individuals in Waconichi Lake assigned to any other source populations. The presence of a single population in Waconichi Lake also has conservation implications, as this lake's brook trout do not benefit from portfolio effects. Special consideration should be given to conserving

the Bordeleau stock, as disturbances affecting this source could lead to the decline and collapse of brook trout in Waconichi Lake.

Inter-lake movement of lake-migratory brook trout

Brook trout populations inhabiting Mistassini and Mistasiniishish Lakes travel long distances (>100 km) towards feeding areas. Our study found Mistassini brook trout over 100 km away from their source river, as well as Mistasiniishish brook trout caught in Mistassini Lake (more than 80 km away from their source). Our mixed-stock assignments revealed inter-lake movement of brook trout between Mistasiniishish and Mistassini Lakes. Nineteen Mistasiniishish brook trout were caught in Mistassini Lake, and 1 Mistassini brook trout was caught in Mistasiniishish Lake, suggesting that inter-lake movement is largely unidirectional from Mistasiniishish to Mistassini.

Out of 149 confident assignments, one individual in Mistasiniishish assigned to Papas. This individual was caught in the top of the outlet of Mistasiniishish that leads downstream to a waterfall draining into Mistassini. In contrast, 19 Mistasiniishish individuals (15 TEM, 4 Alb_distinct) were caught in Mistassini Lake. These individuals had a wide spatial distribution, with some being caught at the foot of the waterfall and others more than 60 km away from it. Despite the substantial number of migrant brook trout from Temiscamie, we found no evidence of dispersal based on source individual analysis. With the exception of CHE and PAP, we found no genetic overlap between source populations (Figure 2C, E). This raises the question whether migrant brook trout between these lakes are engaging in dispersal. Future studies with larger sample sizes for source populations should provide a definite answer to this question.

We found no evidence for inter-lake movement between Mistassini and Waconichi Lakes. There is a waterfall connecting these two lakes, which may serve as a barrier to migration. The habitat at the southern end of Mistassini Lake is also not suitable for brook trout and harbours warm water species with dominant populations, such as walleye (Bowles et al., 2021; Gibelli, 2023).

Conclusion

We resolved the population structure of lake-migratory brook trout inhabiting three large lakes in northern Quebec. Despite the widespread evidence that brook trout exhibit some of most pronounced population genetic structure at small scales known in vertebrate species (Kazyak et al., 2022; Wood et al., 2014), we provide evidence for a low number of distinct, sympatrically-occurring lake migratory populations occupying large scales in each lake. Based on our observed number of populations present in each lake, there is likely an effect of lake size on the evolution of sympatric brook trout populations. In Mistassini Lake, 2 of the 3 sources produce over 90% of the brook trout harvest. In Mistasiniishish Lake, the Temiscamie river produces over 80% of the harvest, and an unknown source produces the remainder. In Waconichi Lake, the Bordeleau river is the only source of brook trout. These brook trout populations travel up to hundreds of kilometers across their respective lakes in search of feeding areas. We found evidence for inter-lake migration between Mistassini and Mistasiniishish Lakes, without subsequent dispersal to spawning habitats. The presence of few source populations in such large lake systems highlights the importance of conserving each population and their habitats to ensure the persistence of brook trout. The potential demographic declines observed in Mistassini and Mistasiniishish Lakes reflect the sensitivity of these populations to environmental change. The future of this species (and the harvest/food security it provides) is dependent on the preservation of the population structure present in the Albanel-Mistassini-Waconichi reserve. It will be up to community-based management to define the appropriate timing, quotas, and areas of harvest within these lakes to prevent further declines in brook trout populations. In doing so, these fisheries will be effectively sustained for the Cree who have relied on them for generations.

REFERENCES

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

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology*, *26*(2), 420–430. https://doi.org/10.1111/mec.13948

Andvik, R. T., Sloss, B. L., VanDeHey, J. A., Claramunt, R. M., Hansen, S. P., & Isermann, D. A. (2016). Mixed stock analysis of Lake Michigan's Lake Whitefish Coregonus clupeaformis commercial fishery. *Journal of Great Lakes Research*, *42*(3), 660–667. https://doi.org/10.1016/j.jglr.2016.02.005

Beacham, T. D., Wallace, C., Jonsen, K., McIntosh, B., Candy, J. R., Willis, D., Lynch, C., Moore, J.-S., Bernatchez, L., & Withler, R. E. (2019). Comparison of coded-wire tagging with parentage-based tagging and genetic stock identification in a large-scale coho salmon fisheries application in British Columbia, Canada. *Evolutionary Applications*, *12*(2), 230–254. https://doi.org/10.1111/eva.12711

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., Maes, G. E., Martinsohn, J. T., Miller, K. M., Naish, K. A., Ovenden, J. R., Primmer, C. R., Young Suk, H., Therkildsen, N. O., & Withler, R. E. (2017). Harnessing the Power of Genomics to Secure the Future of Seafood. *Trends in Ecology & Evolution*, *32*(9), 665–680. https://doi.org/10.1016/j.tree.2017.06.010

Bowles, E., Marin, K., MacLeod, P., & Fraser, D. J. (2021). A three-pronged approach that leans on Indigenous knowledge for northern fish monitoring and conservation. *Evolutionary Applications*, *14*(3), 653–657. https://doi.org/10.1111/eva.13146

Bradbury, I. R., Hamilton, L. C., Chaput, G., Robertson, M. J., Goraguer, H., Walsh, A., Morris, V., Reddin, D., Dempson, J. B., Sheehan, T. F., King, T., & Bernatchez, L. (2016). Genetic mixed stock analysis of an interceptory Atlantic salmon fishery in the Northwest Atlantic. *Fisheries Research*, *174*, 234–244. https://doi.org/10.1016/j.fishres.2015.10.009

Bradbury, I. R., Wringe, B. F., Watson, B., Paterson, I., Horne, J., Beiko, R., Lehnert, S. J., Clément, M., Anderson, E. C., Jeffery, N. W., Duffy, S., Sylvester, E., Robertson, M., & Bentzen, P. (2018). Genotyping-by-sequencing of genome-wide microsatellite loci reveals fine-scale harvest composition in a coastal Atlantic salmon fishery. *Evolutionary Applications*, *11*(6), 918–930. https://doi.org/10.1111/eva.12606

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, *15*(4), 855–867. https://doi.org/10.1111/1755-0998.12357

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, *114*(30). https://doi.org/10.1073/pnas.1704949114

D'Amelio, S., Mucha, J., Mackereth, R., & Wilson, C. C. (2008). Tracking Coaster Brook Trout to Their Sources: Combining Telemetry and Genetic Profiles to Determine Source Populations. *North American Journal of Fisheries Management*, *28*(4), 1343–1349. https://doi.org/10.1577/M05-193.1

D'Amelio, S., & Wilson, C. (2008). Genetic Population Structure among Source Populations for Coaster Brook Trout in Nipigon Bay, Lake Superior. *Transactions of the American Fisheries Society*, *137*(4), 1213–1228. https://doi.org/10.1577/T05-275.1

Des Roches, S., Pendleton, L. H., Shapiro, B., & Palkovacs, E. P. (2021). Conserving intraspecific variation for nature's contributions to people. *Nature Ecology & Evolution*, *5*(5), 574–582. https://doi.org/10.1038/s41559-021-01403-5

Des Roches, S., Post, D. M., Turley, N. E., Bailey, J. K., Hendry, A. P., Kinnison, M. T., Schweitzer, J. A., & Palkovacs, E. P. (2017). The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, *2*(1), 57–64. https://doi.org/10.1038/s41559-017-0402-5

Ferguson, A., Reed, T. E., Cross, T. F., McGinnity, P., & Prodöhl, P. A. (2019). Anadromy, potamodromy and residency in brown trout *Salmo trutta*: The role of genes and the environment. *Journal of Fish Biology*, *95*(3), 692–718. https://doi.org/10.1111/jfb.14005

Ferguson, A., & Taggart, J. B. (1991). Genetic differentiation among the sympatric brown trout (Salmo trutta) populations of Lough Melvin, Ireland. *Biological Journal of the Linnean Society*, *43*(3), 221–237. https://doi.org/10.1111/j.1095-8312.1991.tb00595.x

Flick, W. A. (1977). Some observations, age, growth, food-habits and vulnerability of large brook trout (Salvelinus-fontinalis) from 4 Canadian lakes. *Naturaliste Canadien*, *104*(4), 353–359.

Francis, R. M. (2017). POPHELPER: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Fraser, D. J., & Bernatchez, L. (2005a). Adaptive migratory divergence among sympatric brook charr populations. *Evolution*, *59*(3), 611. https://doi.org/10.1554/04-346

Fraser, D. J., & Bernatchez, L. (2005b). Allopatric origins of sympatric brook charr populations: Colonization history and admixture. *Molecular Ecology*, *14*(5), 1497–1509. https://doi.org/10.1111/j.1365-294X.2005.02523.x

Fraser, D. J., & Bernatchez, L. (2008). Ecology, Evolution, and Conservation of Lake-Migratory Brook Trout: A Perspective from Pristine Populations. *Transactions of the American Fisheries Society*, *137*(4), 1192–1202. https://doi.org/10.1577/T05-251.1

Fraser, D. J., Calvert, A. M., Bernatchez, L., & Coon, A. (2013). Multidisciplinary population monitoring when demographic data are sparse: A case study of remote trout populations. *Ecology and Evolution*, *3*(15), 4954–4969. https://doi.org/10.1002/ece3.871

Fraser, D. J., Coon, T., Prince, M. R., Dion, R., & Bernatchez, L. (2006). Integrating Traditional and Evolutionary Knowledge in Biodiversity Conservation: A Population Level Case Study. *Ecology and Society*, *11*(2), art4. https://doi.org/10.5751/ES-01754-110204

Fraser, D. J., Lippé, C., & Bernatchez, L. (2004). Consequences of unequal population size, asymmetric gene flow and sex-biased dispersal on population structure in brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, *13*(1), 67–80. https://doi.org/10.1046/j.1365-294X.2003.02038.x

Gibelli, J. (2023). *Genomics-based mixed-stock analysis reveals potential unsampled populations and population differences in intra-lake migration in walleye.*

Harley, S. J., Myers, R. A., & Dunn, A. (2001). Is catch-per-unit-effort proportional to abundance? *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(9), 1760–1772. https://doi.org/10.1139/f01-112

Harris, L. N., Bogsuski, D. A., Gallagher, C. P., & Howland, K. L. (2016). Genetic Stock Identification and Relative Contribution of Arctic Char (*Salvelinus alpinus*) from the Hornaday and Brock Rivers to Subsistence Fisheries in Darnley Bay, Northwest Territories + Supplementary Appendix Tables S1 to S4 (See Article Tools). *ARCTIC*, *69*(3), 231. https://doi.org/10.14430/arctic4578

Hemstrom, W., & Jones, M. (2023). snpR: User friendly population genomics for SNP data sets with categorical metadata. *Molecular Ecology Resources*, *23*(4), 962–973. https://doi.org/10.1111/1755-0998.13721

Hilborn, R., Quinn, T. P., Schindler, D. E., & Rogers, D. E. (2003). Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences*, *100*(11), 6564–6568. https://doi.org/10.1073/pnas.1037274100

Horne, J. B., Roden, S. E., LaCasella, E. L., Frey, A., Martin, S. L., Jones, T. T., Murakawa, S., Brunson, S., Balazs, G. H., & Dutton, P. H. (2023). Origins of green turtle fishery bycatch in the central Pacific revealed by mixed genetic markers. *Frontiers in Marine Science*, *10*, 1112842. https://doi.org/10.3389/fmars.2023.1112842

Kazyak, D. C., Lubinski, B. A., Kulp, M. A., Pregler, K. C., Whiteley, A. R., Hallerman, E., Coombs, J. A., Kanno, Y., Rash, J. M., Morgan, R. P., Habera, J., Henegar, J., Weathers, T. C., Sell, M. T., Rabern, A., Rankin, D., & King, T. L. (2022). Population Genetics of Brook Trout in the Southern Appalachian Mountains. *Transactions of the American Fisheries Society*, *151*(2), 127–149. https://doi.org/10.1002/tafs.10337

Kilduff, D. P., Di Lorenzo, E., Botsford, L. W., & Teo, S. L. H. (2015). Changing central Pacific El Niños reduce stability of North American salmon survival rates. *Proceedings of the National Academy of Sciences*, *112*(35), 10962–10966. https://doi.org/10.1073/pnas.1503190112

Leigh, D. M., Hendry, A. P., Vázquez-Domínguez, E., & Friesen, V. L. (2019). Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications*, *12*(8), 1505–1512. https://doi.org/10.1111/eva.12810

Li, P., Van Coeverden De Groot, P., Sun, Z., & Lougheed, S. C. (2023). A new genomics tool for monitoring Arctic char (Salvelinus alpinus) populations in the Lower Northwest Passage, Nunavut. *Fisheries Research*, *258*, 106523. https://doi.org/10.1016/j.fishres.2022.106523

Li, Y.-H., & Wang, H.-P. (2017). Advances of genotyping-by-sequencing in fisheries and aquaculture. *Reviews in Fish Biology and Fisheries*, *27*(3), 535–559. https://doi.org/10.1007/s11160-017-9473-2

Lundrigan, T. A., Reist, J. D., & Ferguson, M. M. (2005). Microsatellite genetic variation within and among Arctic charr (Salvelinus alpinus) from aquaculture and natural populations in North America. *Aquaculture*, *244*(1–4), 63–75. https://doi.org/10.1016/j.aquaculture.2004.11.027

Manel, S., Gaggiotti, O., & Waples, R. (2005). Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, *20*(3), 136–142. https://doi.org/10.1016/j.tree.2004.12.004

Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P., & Harley, S. J. (2006). Interpreting catch per unit effort data to assess the status of individual stocks and communities. *ICES Journal of Marine Science*, *63*(8), 1373–1385. https://doi.org/10.1016/j.icesjms.2006.05.008

McKinney, G. J., Pascal, C. E., Templin, W. D., Gilk-Baumer, S. E., Dann, T. H., Seeb, L. W., & Seeb, J. E. (2020). Dense SNP panels resolve closely related Chinook salmon populations. *Canadian Journal of Fisheries and Aquatic Sciences*, *77*(3), 451–461. https://doi.org/10.1139/cjfas-2019-0067

Meek, M. H., & Larson, W. A. (2019). The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular Ecology Resources*, *19*(4), 795–803. https://doi.org/10.1111/1755-0998.12998

Mimura, M., Yahara, T., Faith, D. P., Vázquez-Domínguez, E., Colautti, R. I., Araki, H., Javadi, F., Núñez-Farfán, J., Mori, A. S., Zhou, S., Hollingsworth, P. M., Neaves, L. E., Fukano, Y., Smith, G. F., Sato, Y., Tachida, H., & Hendry, A. P. (2017). Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, *10*(2), 121–139. https://doi.org/10.1111/eva.12436

Moran, B. M., & Anderson, E. C. (2019). Bayesian inference from the conditional genetic stock identification model. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*(4), 551–560. https://doi.org/10.1139/cjfas-2018-0016

Mucha, J. M., & Mackereth, R. W. (2008). Habitat Use and Movement Patterns of Brook Trout in Nipigon Bay, Lake Superior. *Transactions of the American Fisheries Society*, *137*(4), 1203–1212. https://doi.org/10.1577/T05-273.1

Nesbitt, H. K., & Moore, J. W. (2016). Species and population diversity in Pacific salmon fisheries underpin indigenous food security. *Journal of Applied Ecology*, *53*(5), 1489–1499. https://doi.org/10.1111/1365-2664.12717

Northcote, T. G. (1962). *Migratory Behaviour of Juvenile Rainbow Trout, Salrno gairdneri, in Outlet and Inlet Streams of Loon Lake, British Columbia*. *19*(2).

O'Connor, J. F., & Power, G. (1973). Homing of Brook Trout (*Salvelinus fontinalis*) in Matamek Lake, Quebec. *Journal of the Fisheries Research Board of Canada*, *30*(7), 1012–1014. https://doi.org/10.1139/f73-166

Penaluna, B. E., Abadía-Cardoso, A., Dunham, J. B., García-Dé León, F. J., Gresswell, R. E., Luna, A. R., Taylor, E. B., Shepard, B. B., Al-Chokhachy, R., Muhlfeld, C. C., Bestgen, K. R., Rogers, K., Escalante, M. A., Keeley, E. R., Temple, G. M., Williams, J. E., Matthews, K. R., Pierce, R., Mayden, R. L., … Fausch, K. D. (2016). Conservation of Native Pacific Trout Diversity in Western North America. *Fisheries*, *41*(6), 286–300. https://doi.org/10.1080/03632415.2016.1175888

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Ridgway, M. S. (2008). A Roadmap for Coasters: Landscapes, Life Histories, and the Conservation of Brook Trout. *Transactions of the American Fisheries Society*, *137*(4), 1179–1191. https://doi.org/10.1577/T05-268.1

Robillard, M. M., Casselman, J. M., McLaughlin, R. L., & Mackereth, R. W. (2011). Alternative growth histories in populations of Lake Superior brook trout: Critical support for partial migration. *Biological Conservation*, *144*(7), 1931–1939. https://doi.org/10.1016/j.biocon.2011.03.022

ROUSSET, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Saunders, R. L. (1981). Atlantic Salmon (*Salmo salar*) Stocks and Management Implications in the Canadian Atlantic Provinces and New England, USA. *Canadian Journal of Fisheries and Aquatic Sciences*, *38*(12), 1612–1625. https://doi.org/10.1139/f81-211

Schaffer, W. M., & Elson, P. F. (1975). The Adaptive Significance of Variations in Life History among Local Populations of Atlantic Salmon in North America. *Ecology*, *56*(3), 577–590. https://doi.org/10.2307/1935492

Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, *465*(7298), 609–612. https://doi.org/10.1038/nature09060

Schreiner, D. R., Cullis, K. I., Donofrio, M. C., Fischer, G. J., Hewitt, L., Mumford, K. G., Pratt, D. M., Quinlan, H. R., & Scott, S. J. (2008). Management Perspectives on Coaster Brook Trout Rehabilitation in the Lake Superior Basin. *North American Journal of Fisheries Management*, *28*(4), 1350–1364. https://doi.org/10.1577/M05-173.1

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., & Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, *11*(s1), 1–8. https://doi.org/10.1111/j.1755-0998.2010.02979.x

Shaklee, J. B., Beacham, T. D., Seeb, L., & White, B. A. (1999). Managing ®sheries using genetic data: Case studies from four species of Paci®c salmon. *Fisheries Research*.

Stott, W., Quinlan, H. R., Gorman, O. T., & King, T. L. (2010). Genetic Structure and Diversity among Brook Trout from Isle Royale, Lake Nipigon, and Three Minnesota Tributaries of Lake Superior. *North American Journal of Fisheries Management*, *30*(2), 400–411. https://doi.org/10.1577/M08-213.1

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*(1), 131. https://doi.org/10.1186/s13059-018-1520-3

Taylor, E. B. (1991). A review of local adaptation in Salmonidac, with particular reference to Pacific and Atlantic salmon. *Aquaculture*, *98*(1), 185–207. https://doi.org/10.1016/0044-8486(91)90383-I

Tessier, N., & Bernatchez, L. (1999). Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, *8*(2), 169–179. https://doi.org/10.1046/j.1365-294X.1999.00547.x

Tessier, N., Bernatchez, L., & Wright, J. M. (1997). Population structure and impact of supportive breeding inferred from mitochondrial and microsatellite DNA analyses in land-locked Atlantic salmon *Salmo salar* L. *Molecular Ecology*, *6*(8), 735–750. https://doi.org/10.1046/j.1365-294X.1997.00244.x

Torrissen, O., Jones, S., Asche, F., Guttormsen, A., Skilbrei, O. T., Nilsen, F., Horsberg, T. E., & Jackson, D. (2013). Salmon lice – impact on wild salmonids and salmon aquaculture. *Journal of Fish Diseases*, *36*(3), 171–194. https://doi.org/10.1111/jfd.12061

Waples, R. S., & Do, C. (2008). LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, *8*(4), 753–756. https://doi.org/10.1111/j.1755-0998.2007.02061.x

Waples, R. S., Winans, G. A., Utter, F. M., & Mahnken, C. (1990). Genetic Approaches to the Management of Pacific Salmon. *Fisheries*, *15*(5), 19–25. https://doi.org/10.1577/1548-8446(1990)015<0019:GATTMO>2.0.CO;2

Ward, R. D. (2000). Genetics in fisheries management. In A. M. Solé-Cava, C. A. M. Russo, & J. P. Thorpe (Eds.), *Marine Genetics* (pp. 191–201). Springer Netherlands. https://doi.org/10.1007/978-94-017-2184-4_18

Wood, J. L. A., Belmar-Lucero, S., Hutchings, J. A., & Fraser, D. J. (2014). Relationship of habitat variability to population size in a stream fish. *Ecological Applications*, *24*(5), 1085–1100. https://doi.org/10.1890/13-1647.1

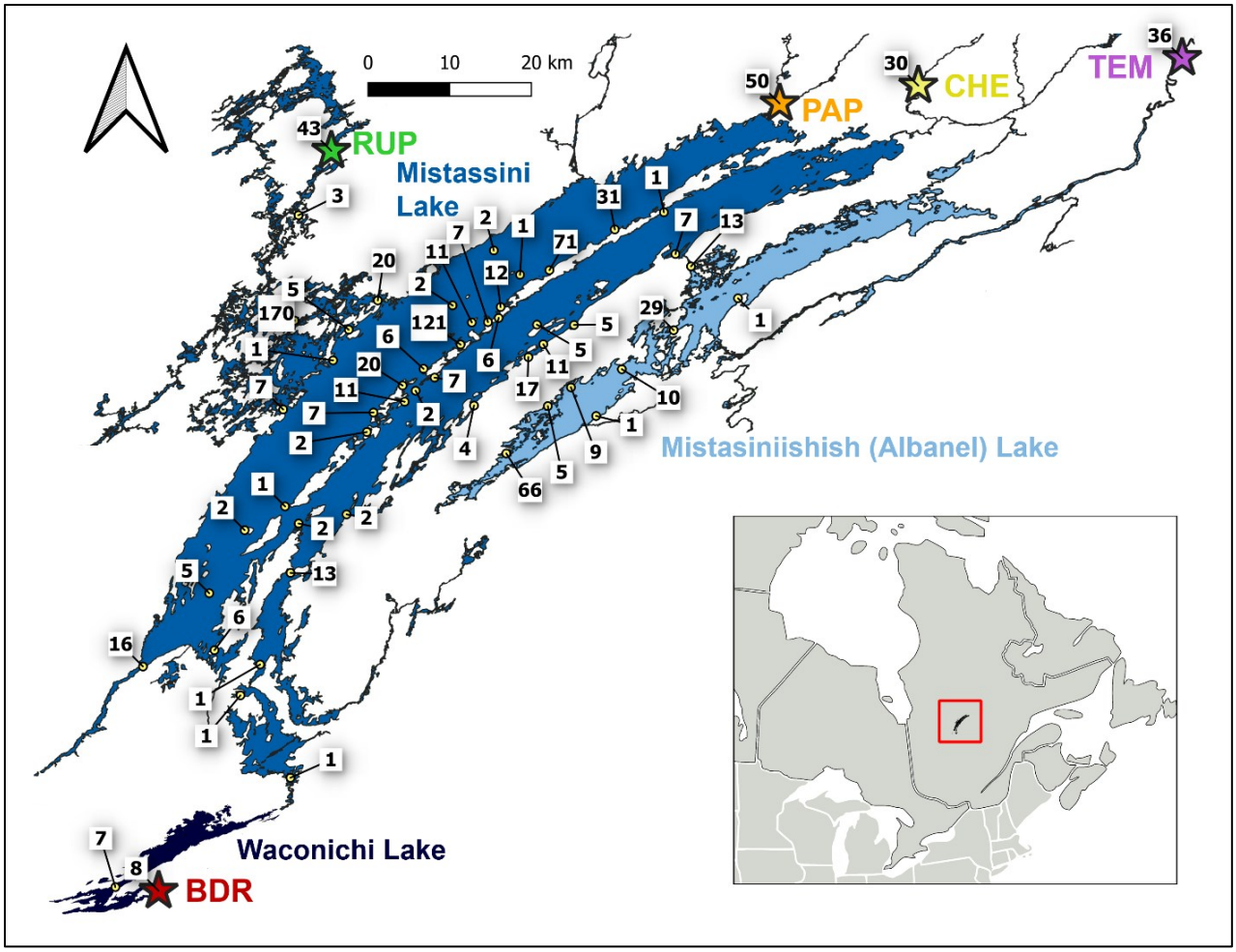

Figure 1. Study lakes and sampling locations. The numbers shown represent the number of individuals sampled at each location. The stars represent known brook trout spawning areas (BDR = Bordeleau River, RUP = Rupert River, PAP = Papas River, CHE = Cheno River, TEM = Temiscamie River), and the attached numbers represent the number of source samples collected at each spawning area. Mixed-stock sample numbers (points in the lakes) include only samples with precise location information.

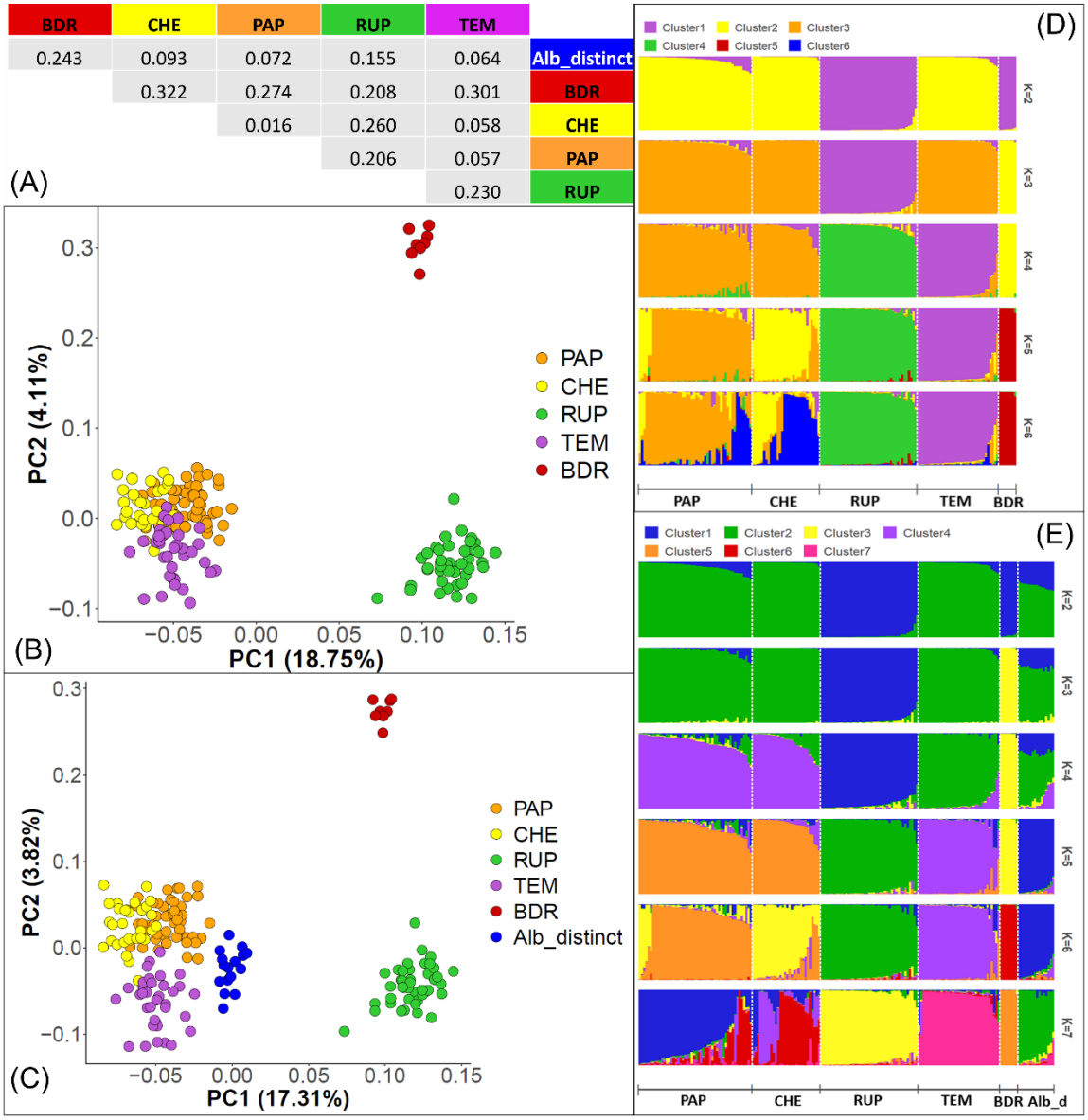

Figure 2. Genetic population structure of brook trout source populations. Top panel (A) shows pairwise fixation index (Fst) values between Bordeleau (BDR), Cheno (CHE), Papas (PAP), Rupert (RUP), Temiscamie (TEM), and Albanel_distinct (Alb_distinct) estimated using the genepop method. Middle-left and bottom-left panels show PCA results for (B) 5 source populations; (C) 6 source populations. PC1 supports strong divergence (18.75%, 17.31%) between (B) RUP, CHE, PAP, and TEM; (C) RUP, CHE, PAP, TEM, and Alb_distinct. The y-axis (PC2) separates BDR source individuals from all other sources. Right panels show results from STRUCTURE analyses of source-individuals from (D) the 5 known sources for K2 to K6; (E) 6 sources for K2 to K7. The individuals are ordered by their river of origin. The y-axis represents the proportion of membership to each cluster (Q-values) estimated by STRUCTURE. For each K, the run with the highest log-likelihood (out of 10 runs) is presented. The PCA plots were generated using snpR (Hemstrom & Jones, 2023). The STRUCTURE plots were generated using pophelper (Francis, 2017).

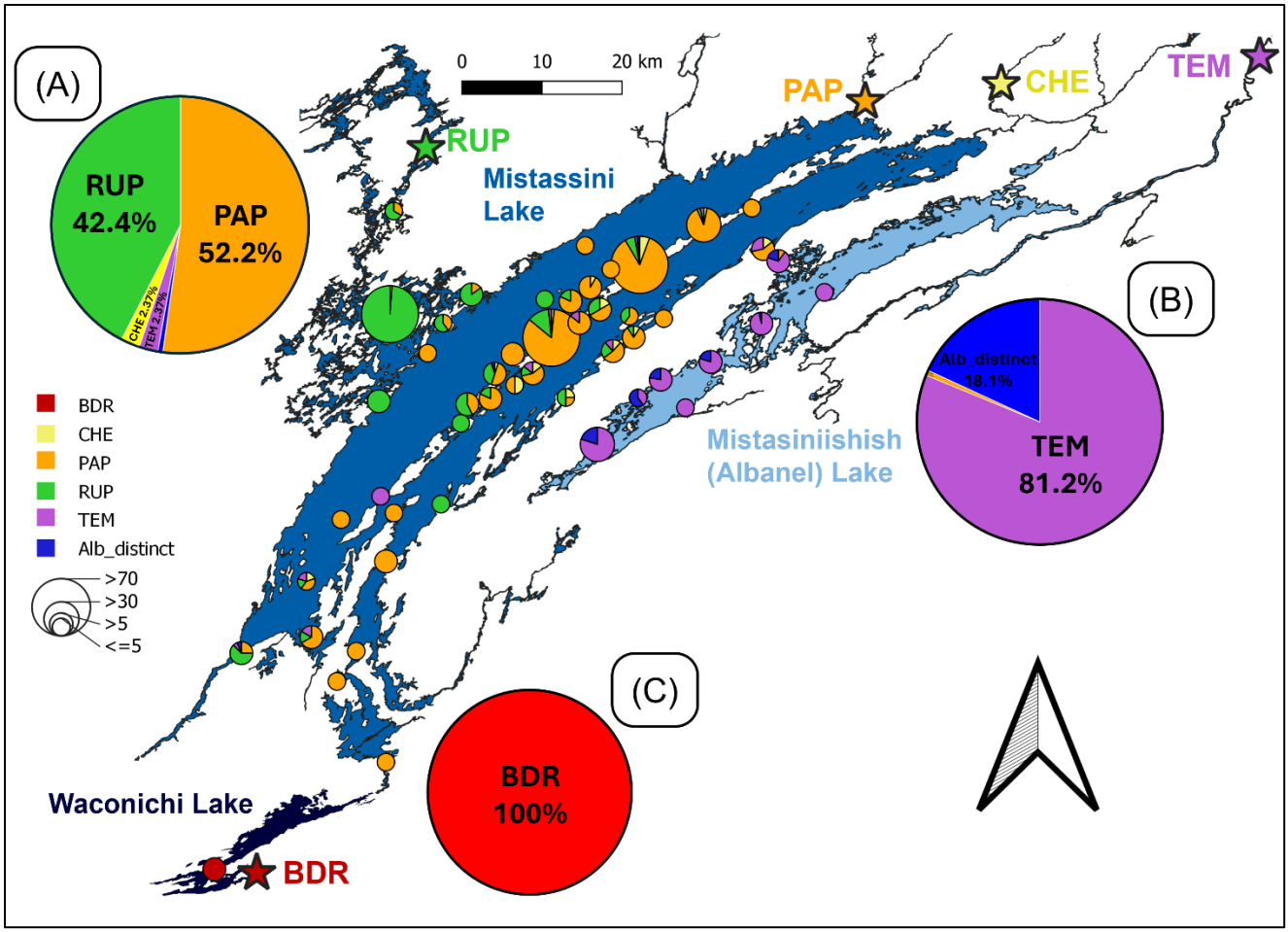

Figure 3. Proportions of mixed-stock brook trout assigned to source populations in Mistassini, Mistasiniishish (Albanel) and Waconichi Lakes. The size of the pie-charts within each lake is relative to the sample-size in those locations and the color represents the proportional assignment to the source population. The stars show the locations of the known source rivers. The 3 large pie-charts represent the proportions of each source population contributing to the overall harvest of (A) Mistassini Lake (n = 634), (B) Mistasiniishish (Albanel) Lake (n = 149), (C) Waconichi Lake (n = 78). (BDR = Bordeleau, CHE = Cheno, PAP = Papas, RUP = Rupert, TEM = Temiscamie, Alb_distinct = Albanel_distinct).

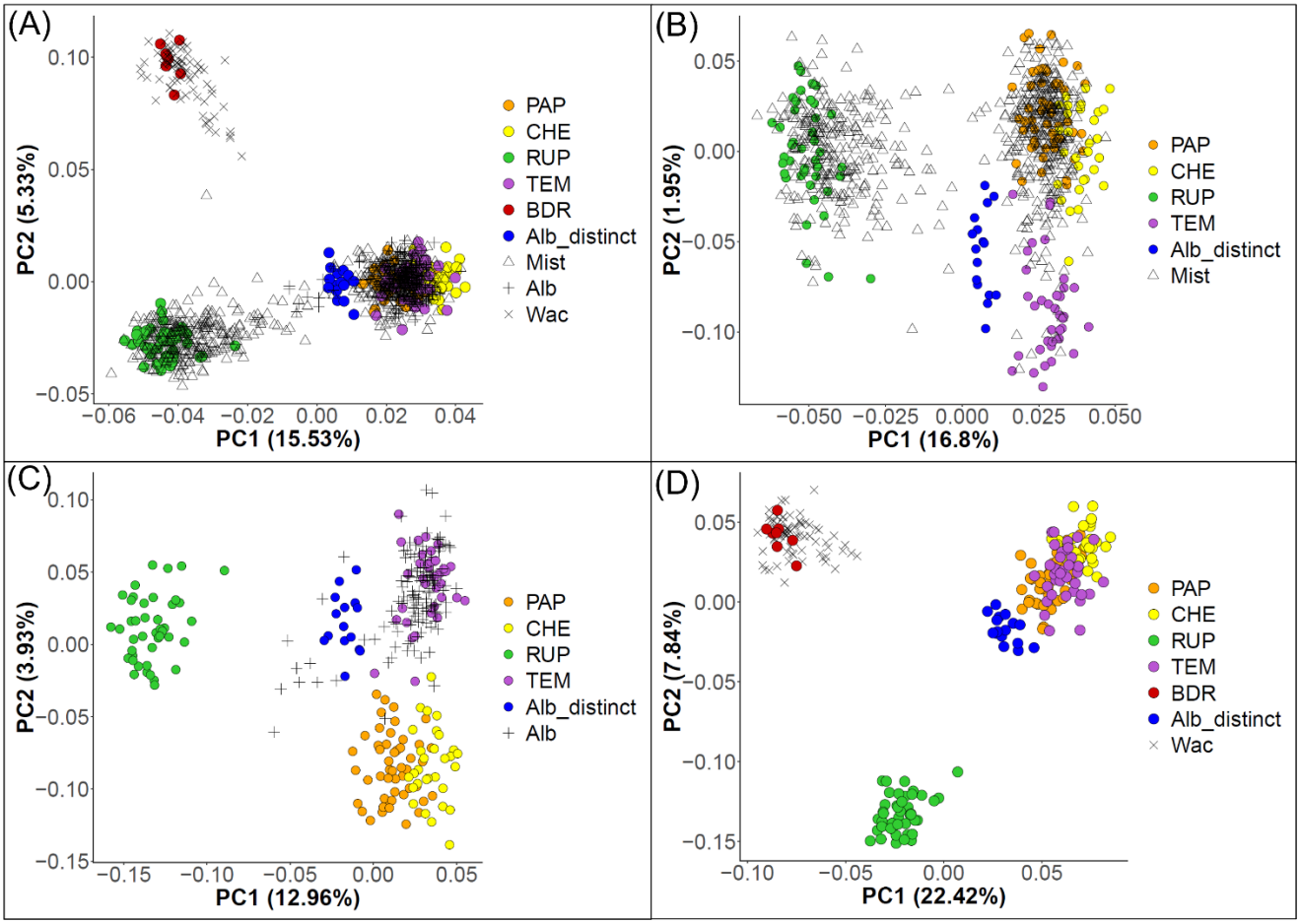

Figure 4. PCA of mixed-stock individuals combined with all known sources. The legends on the right of the plots denote source individuals and mixed-stock individuals from each of the three study lakes: PAP = Papas, CHE = Cheno, RUP = Rupert, TEM = Temiscamie, BDR = Bordeleau, Alb_distinct = Albanel_distinct, Mist = Mistassini Lake mixed-stock, Alb = Mistasiniishish (Albanel) Lake mixed-stock, Wac = Waconichi Lake mixed-stock. The plots shown are for (A) all mixed-stock individuals with all sources, (B) Mistassini Lake mixed-stock with all sources excluding Bordeleau, (C) Mistasiniishish (Albanel) Lake mixed-stock with all sources excluding Bordeleau, (D) Waconichi Lake mixed-stock with all sources. Bordeleau was excluded from the plots in (B) and (C) because it was clustering with Rupert on the PC1 axis, which was uninformative.

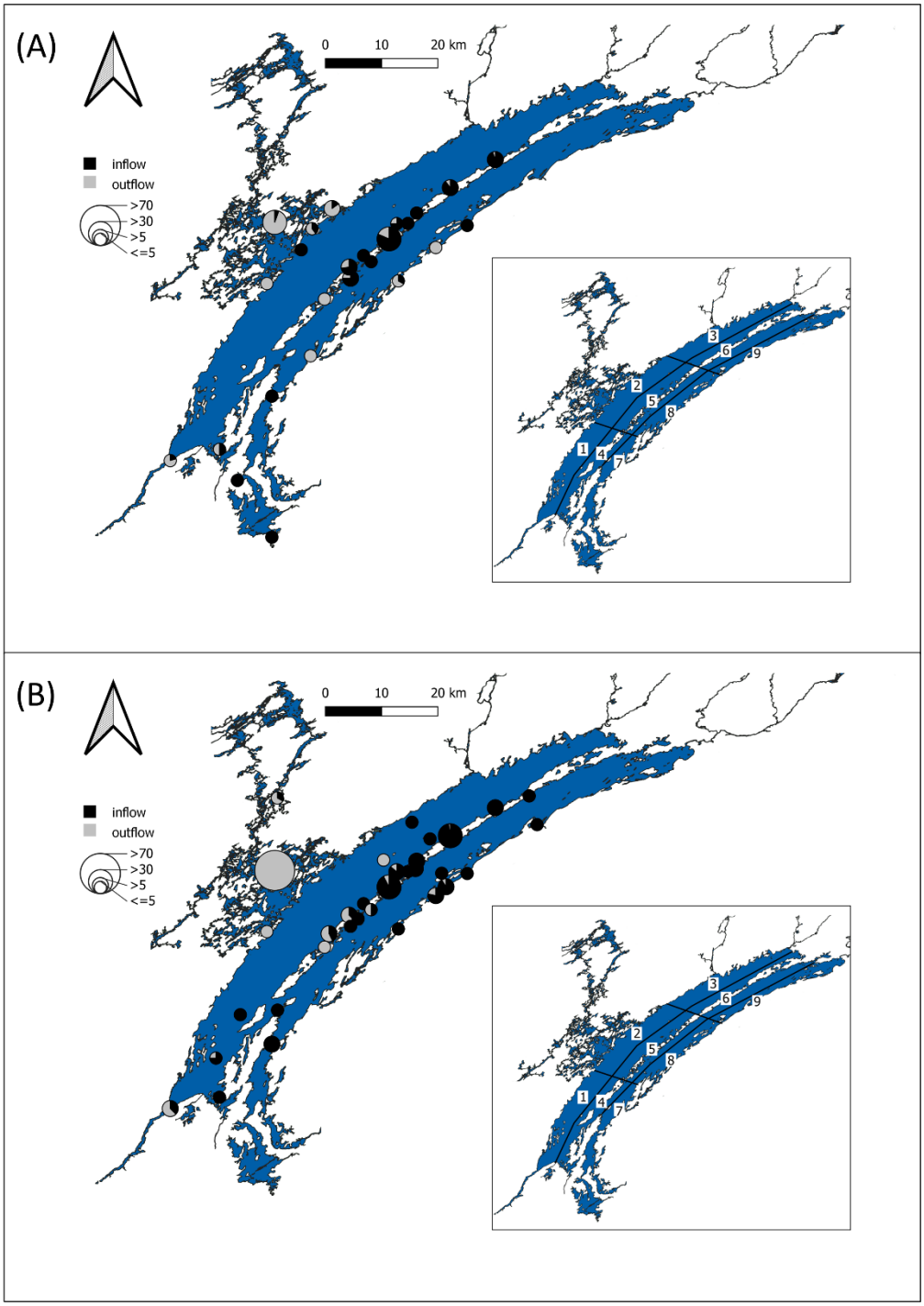

Figure 5. Contemporary spatial distributions of Mistassini Lake brook trout populations. Maps showing the spatial distributions of inflow and outflow brook trout in the sampling years of (A) 2021 and (B) 2022. The size of the pie-charts is relative to the sample-size in those locations and the color represents the proportional assignment to the population (black = inflow, grey = outflow). The inset map in the bottom left of each panel shows the 9 numbered sectors.

Table 1. MANOVA results for spatiotemporal analysis of Mistassini Lake brook trout populations. The dependent variables were population proportions (outflow, inflow). The independent variables were lake sector (sector 1 to sector 9) and time (sampling year). For the last test (historical vs. contemporary), time corresponds to pooled sampling period (2000-2001 vs. 2021-2022).

Test	Independent variable	Wilk's λ	*F*	*P*-value
2021 vs. 2022	Sector	0.001	26.80	8.92×10^{-8}
	Time	0.787	0.95	0.43
2000 vs. 2001 vs. 2021 vs. 2022	Sector	0.019	17.89	1.35×10^{-14}
	Time	0.833	0.74	0.62
Historical (2000-2001) vs. Contemporary (2021-2022)	Sector	0.007	9.68	5.45×10^{-5}
	Time	0.882	0.47	0.64

Table 2. Sample numbers per sector in Mistassini Lake. Historical sample numbers were retrieved from Fraser & Bernatchez 2005a. The nine sectors are shown in the inset map of Figure 5.

region	sector	contemporary (2020+2021+2022)	historical (2000+2001)
north shore	1	20	106
north shore	2	206	277
north shore	3	2	90
island chain	4	7	71
island chain	5	207	186
island chain	6	100	223
south shore	7	18	74
south shore	8	39	90
south shore	9	5	49
total		604	1166

APPENDIX I

DNA extraction details

In all cases, a small piece of tissue was cut using scissors and pliers sterilized by dipping in 95-99% ethanol and then passing through a flame in between each individual sample. For mixed stock individuals, the small piece of tissue is incubated overnight in cell lysis solution (ddH2O + EDTA + NACl + Tris + 10% SDS + proteinase K + RNase A) to allow for complete digestion. The solution is then centrifuged, and the supernatant is transferred into a tube with NaCl. A series of centrifugations is then performed to yield a DNA pellet. The pellet is washed with ethanol, dried for 1 hour and finally mixed with AE buffer to resuspend the DNA in solution. For source individuals, the small piece of tissue is incubated overnight in ATL + Proteinase K solution to allow for complete digestion. AL buffer is then added to the solution, followed by 99% ethanol, and the entire mixture is then transferred into a DNeasy Mini spin column. A series of centrifugations is then performed, first with AW1 buffer and then with AW2 buffer, discarding the flow-through in between each centrifugation. The DNA held in the membrane of the spin column is then eluted by adding AE buffer and centrifuging the mixture, yielding pure DNA in solution.

Catch-per-unit-effort

Estimates of catch-per-unit-effort (CPUE) were calculated for the CHE and PAP source rivers to investigate interpopulation differences as well as to test for changes in CPUE over time by comparing with historical data. For each river sampled during spawning seasons, the number of sampling days and number of anglers per day was recorded. For each day of sampling, the catch (# of brook trout caught) was divided by the effort (# of anglers) to estimate CPUE per day. The data was then plotted as a box plot in the statistics software R (Figure S9). The mean CPUE for Cheno (2020-2021) was 1.5 and the mean CPUE for Papas (2020-2021) was 5.6.

We also analyzed changes in the annual CPUE for brook trout in Mistassini, Mistasiniishish, and Waconichi Lakes over a period of 30 years. We retrieved yearly fishing effort and catch data for brook trout for each lake from MFFP (Ministère des Forêts, de la Faune et des Parcs) and plotted it in Microsoft Excel. For each plot, we added a linear trendline to observe changes over time in CPUE (Figures S5, S6, S7). For each lake, we plotted a linear regression model in the statistics software R to calculate the correlation coefficient and p-value of the trendline. As can be seen in Figures S5 and S6, there is a noticeable downward trend in catch-per-unit-effort for both Mistassini and Mistasiniishish Lakes over 30 years.

Effective population size (N_e) of source populations

We generated effective population size (N_e) estimates for each source population using LDNE (Waples & Do, 2008). As can be seen in Table S1, the effective population size of Papas is over 7 times larger than that of Cheno.

Table S1. Effective population sizes of source populations. N_e estimates were calculated using a Pcrit (lowest allele frequency) of 0.05. Confidence intervals were calculated with the JackKnife method on loci. All values were calculated with the filtered panel of 393 SNPs and rounded to the nearest integer.

Population	N_e	95% confidence intervals
Papas	667	377-2620
Cheno	89	77-104
Rupert	220	173-300
Temiscamie	101	88-119
Bordeleau	-56	-108-∞
Alb_distinct	-58	-72-∞

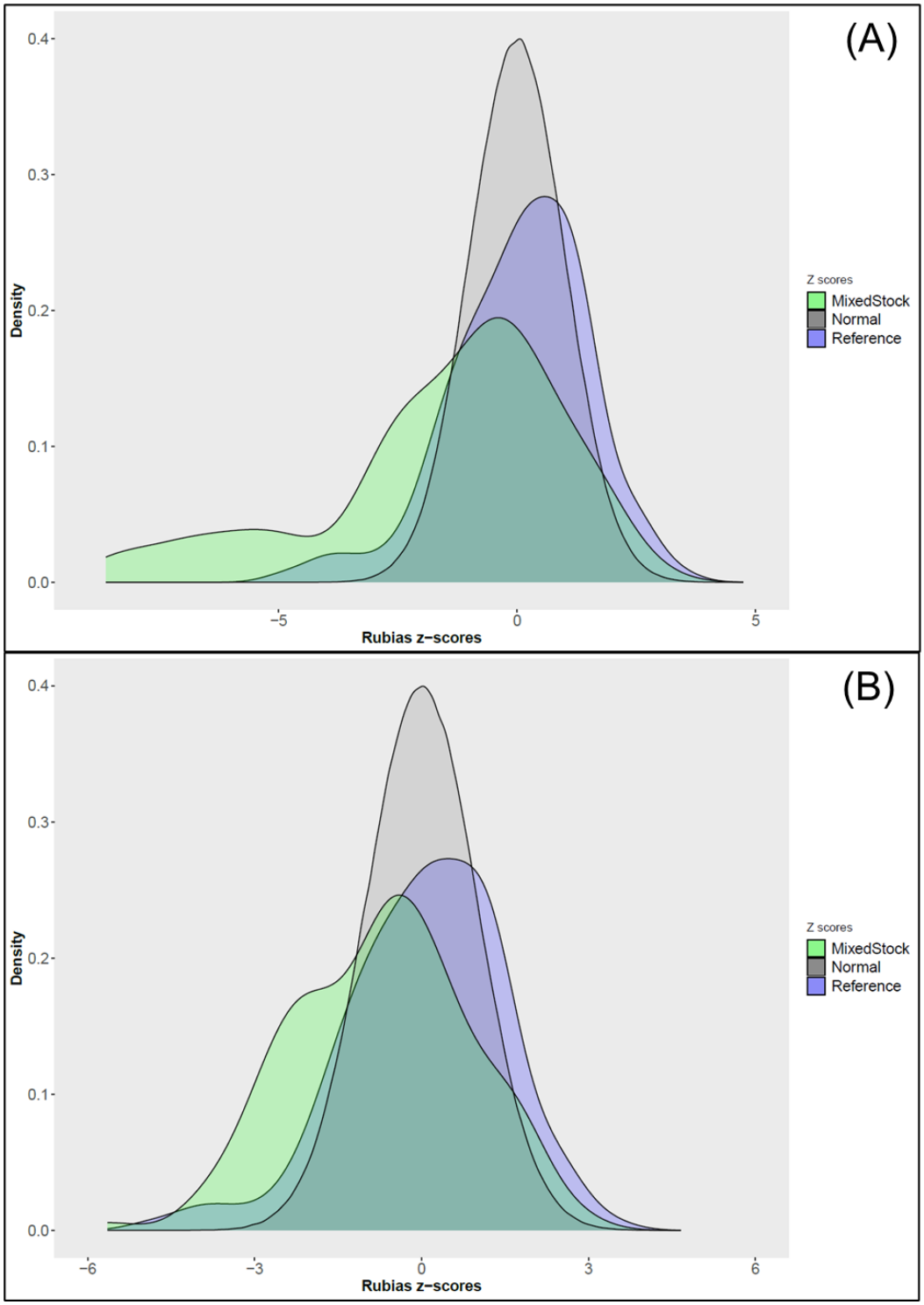

Figure S1. Rubias z-score distributions of Mistasiniishish (Albanel) Lake mixed-stock assignments. The top panel (A) shows z-score distributions of assignments made with the reference dataset of 5 known sources. The bottom panel (B) shows z-score distributions of assignments made with the second reference dataset of 6 sources (5 known + Albanel_distinct).

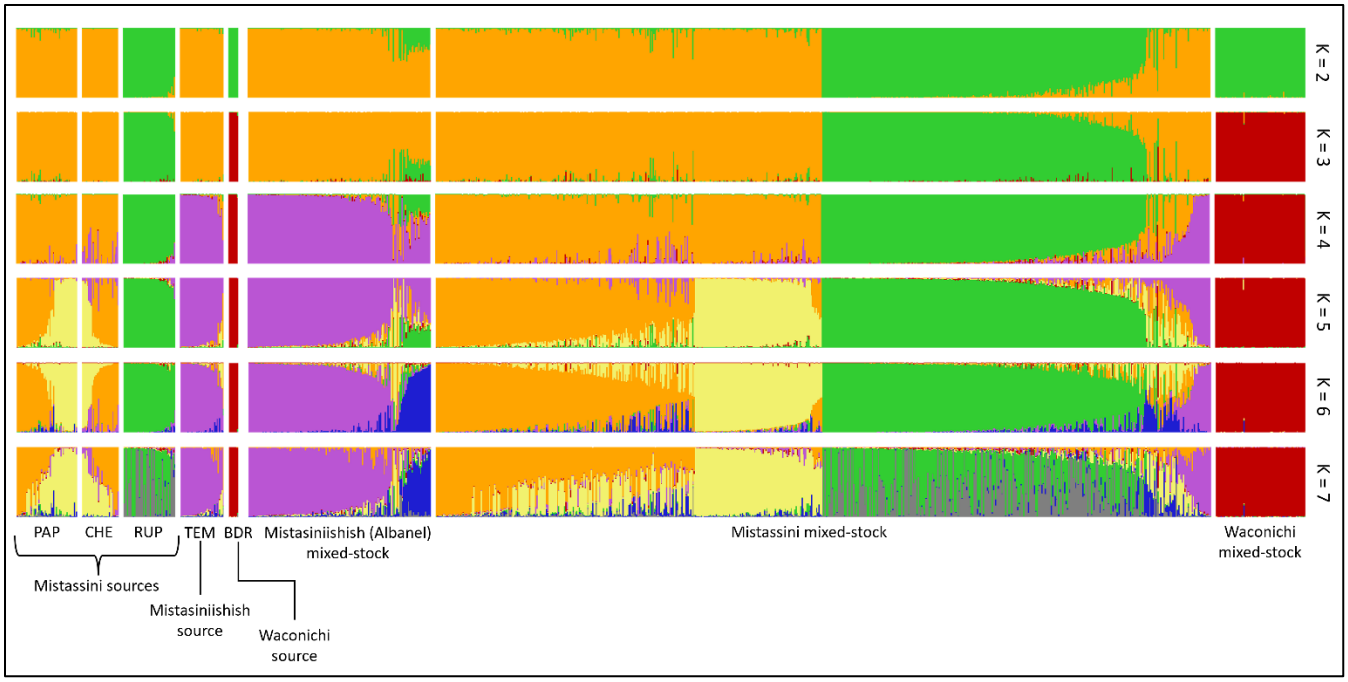

Figure S2. STRUCTURE plot of the 5 known sources followed by all mixed-stock individuals across the 3 lakes. The x-axis denotes source individuals and mixed stock individuals from each of the 3 lakes: PAP = Papas, CHE = Cheno, RUP = Rupert, TEM = Temiscamie, BDR = Bordeleau. The y-axis represents the proportion of membership to each cluster (Q-values) estimated by STRUCTURE. For each K, the run with the highest log-likelihood (out of 10 runs) is presented. At K=6 and above, a distinct cluster (blue color) is visible in the Mistasiniishish (Albanel) mixed-stock individuals that does not match any of the source clusters. These plots were generated using pophelper (Francis, 2017).

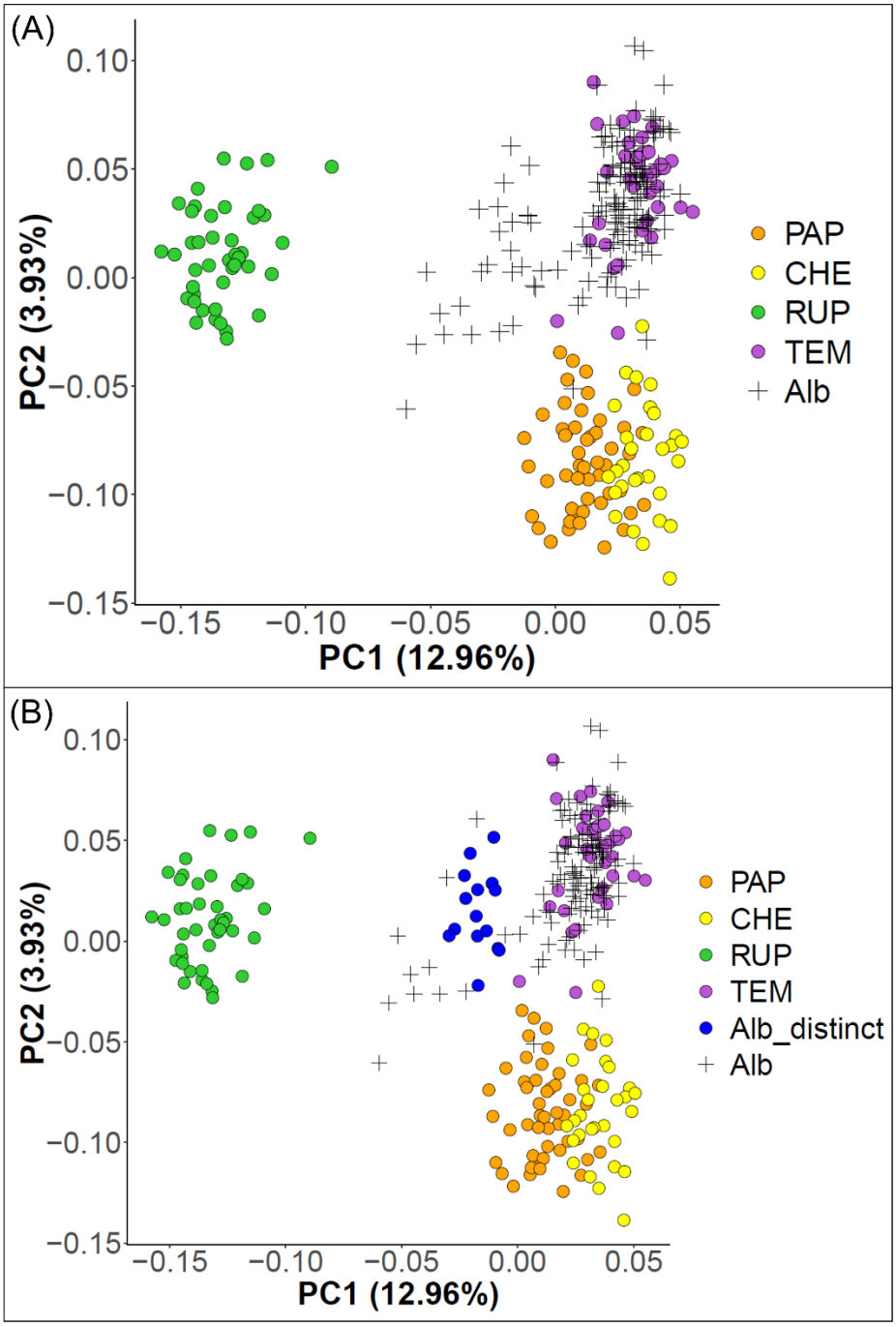

Figure S3. PCA of Mistasiniishish (Albanel) mixed stock individuals with sources. The legends on the right of the plots denote source individuals and mixed-stock individuals: PAP = Papas, CHE = Cheno, RUP = Rupert, TEM = Temiscamie, Alb_distinct = Albanel_distinct, Alb = Mistasiniishish (Albanel) Lake mixed-stock. The plots shown are for (A) Mistasiniishish (Albanel) Lake mixed-stock individuals with all known sources excluding Bordeleau, (B) Same plot as in (A) but with mixed-stock individuals identified in STRUCTURE as pure and distinct (Q-values > 0.80) labeled in blue. Bordeleau was excluded from these plots because it was clustering with Rupert on the PC1 axis, which was uninformative.

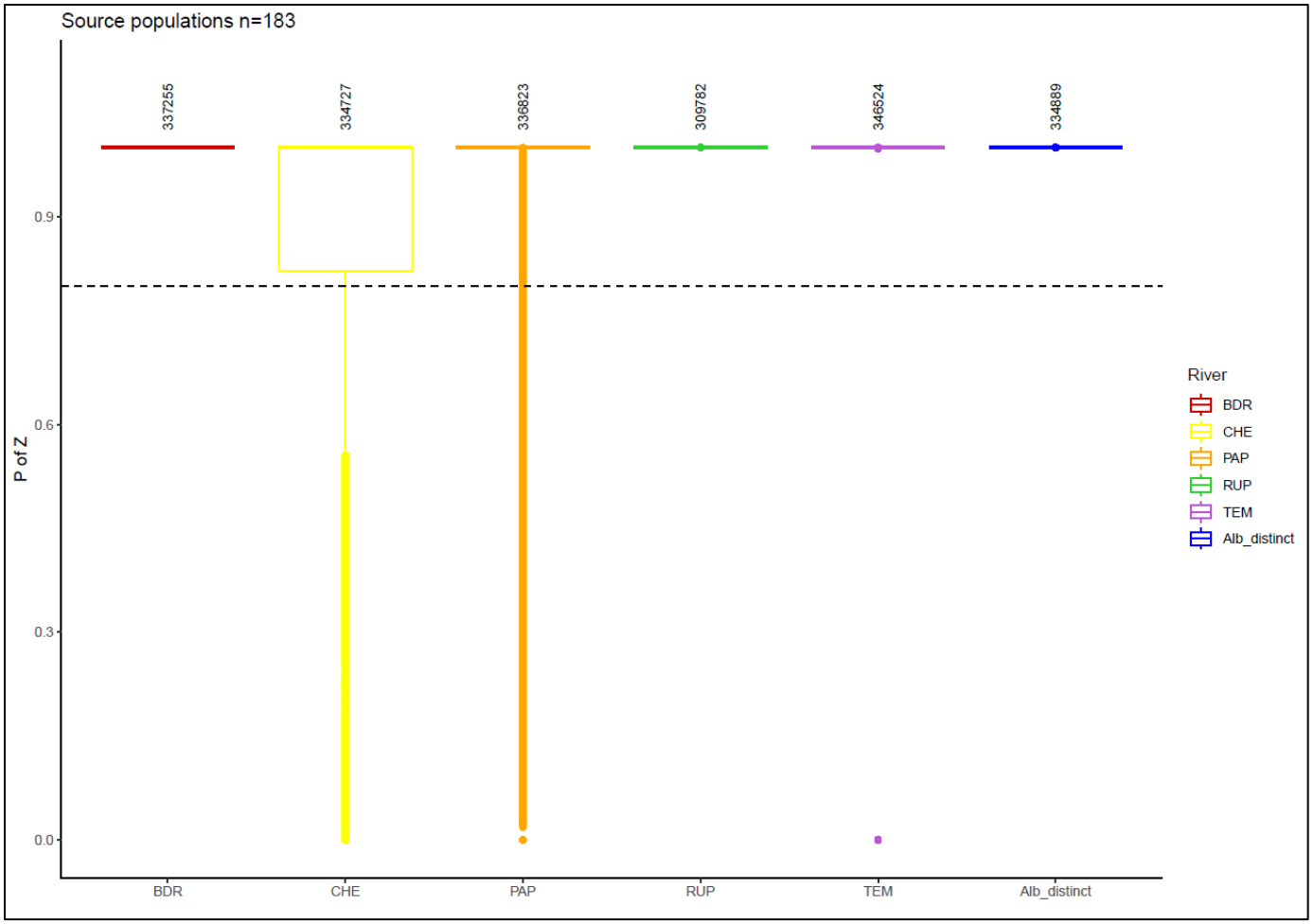

Figure S4. Accuracy of individual assignments made with Rubias. The x-axis denotes the 6 source populations (BDR = Bordeleau, CHE = Cheno, PAP = Papas, RUP = Rupert, TEM = Temiscamie, Alb_distinct = Albanel_distinct). The y-axis represents the posterior means of group membership (PofZ) for individuals assigned to their population of origin in simulated random population mixtures. The dashed line represents the threshold PofZ value (0.80). Simulations were conducted with 1000 replicates, 2000 simulated mixture individuals and the default value for mixing proportion (Dirichlet distribution, alpha = 1.5). The number of simulated individuals assigned to each population is indicated above each box.

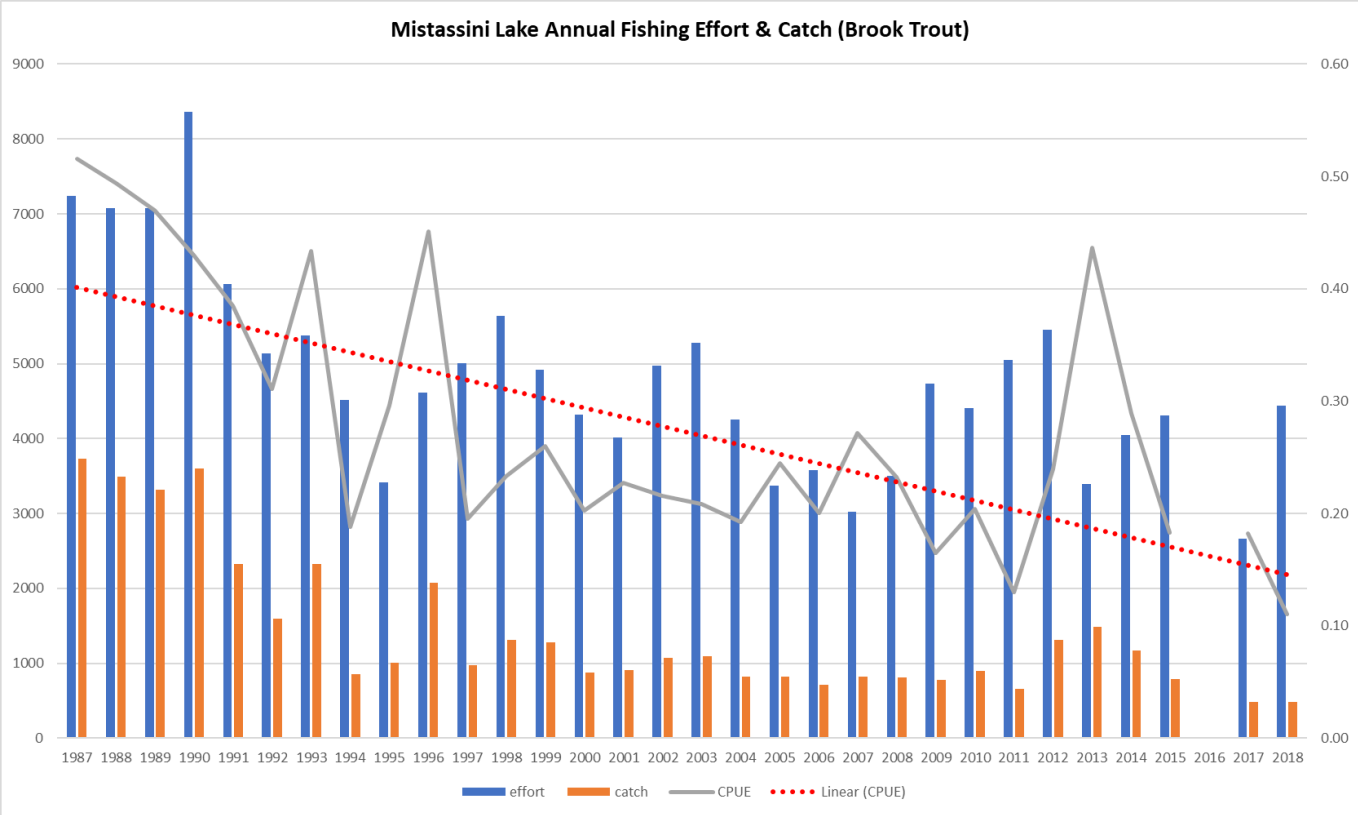

Figure S5. Changes in Mistassini Lake's Catch-Per-Unit-Effort for brook trout over 30 years. The blue bars represent fishing effort, the orange bars represent catch, the grey broken line represents catch-per-unit-effort (CPUE), and the dashed red line is the linear trendline fitted to CPUE. Linear regression: $R^2 = 0.4383$, p-value = 4.99×10^{-5}.

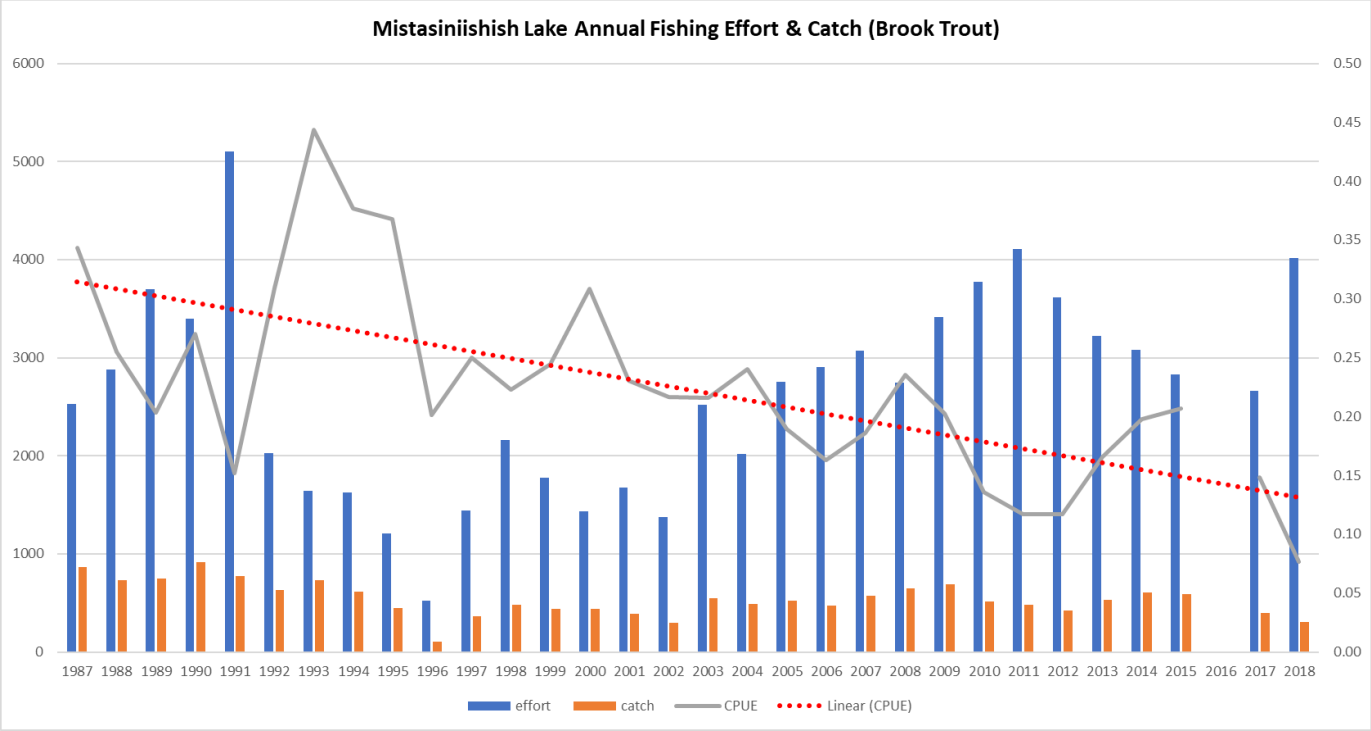

Figure S6. Changes in Mistasiniishish (Albanel) Lake's Catch-Per-Unit-Effort for brook trout over 30 years. The blue bars represent fishing effort, the orange bars represent catch, the grey broken line represents catch-per-unit-effort (CPUE), and the dashed red line is the linear trendline fitted to CPUE. Linear regression: $R^2 = 0.4405$, p-value = 4.70×10^{-5}.

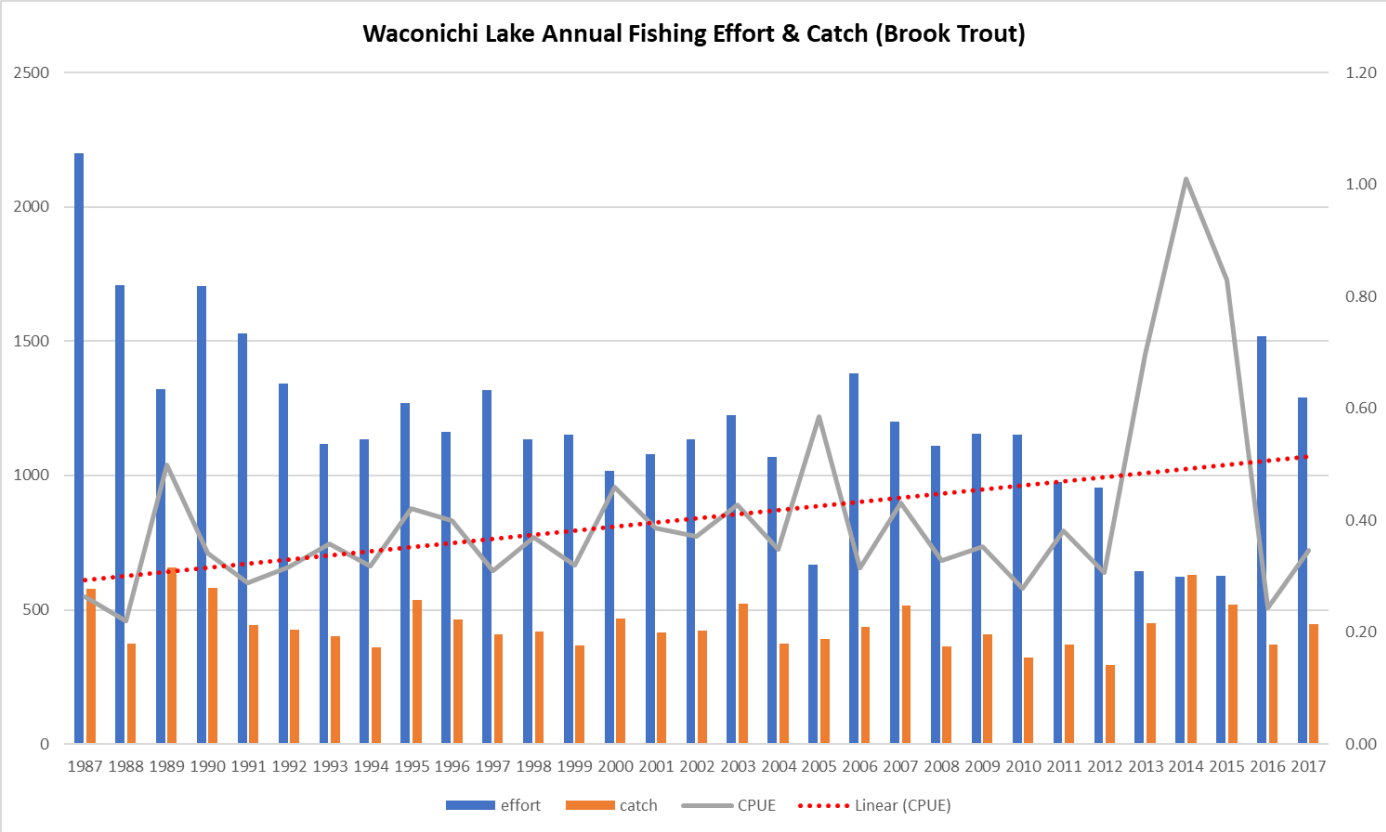

Figure S7. Changes in Waconichi Lake's Catch-Per-Unit-Effort for brook trout over 30 years. The blue bars represent fishing effort, the orange bars represent catch, the grey broken line represents catch-per-unit-effort (CPUE), and the dashed red line is the linear trendline fitted to CPUE. Linear regression: $R^2 = 0.1563$, p-value = 0.0277.

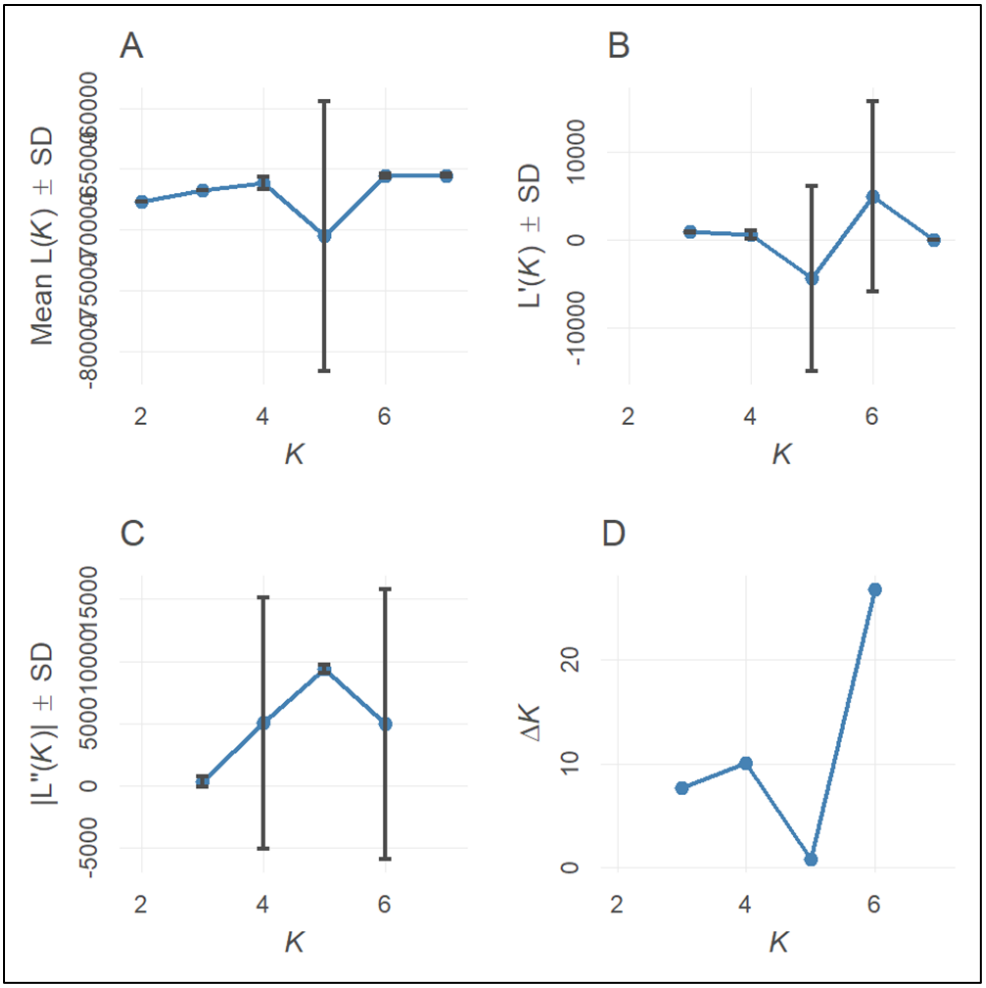

Figure S8. Evanno plots for the STRUCTURE run of 6 source populations (5 known + 1 revealed through mixed-stock analysis). The log probability of each run (A) and the derivatives of the log probability (B, C) are used to calculate ΔK (D). The K value yielding the highest ΔK is the best K value according to the Evanno method. These plots were generated using pophelper (Francis, 2017).

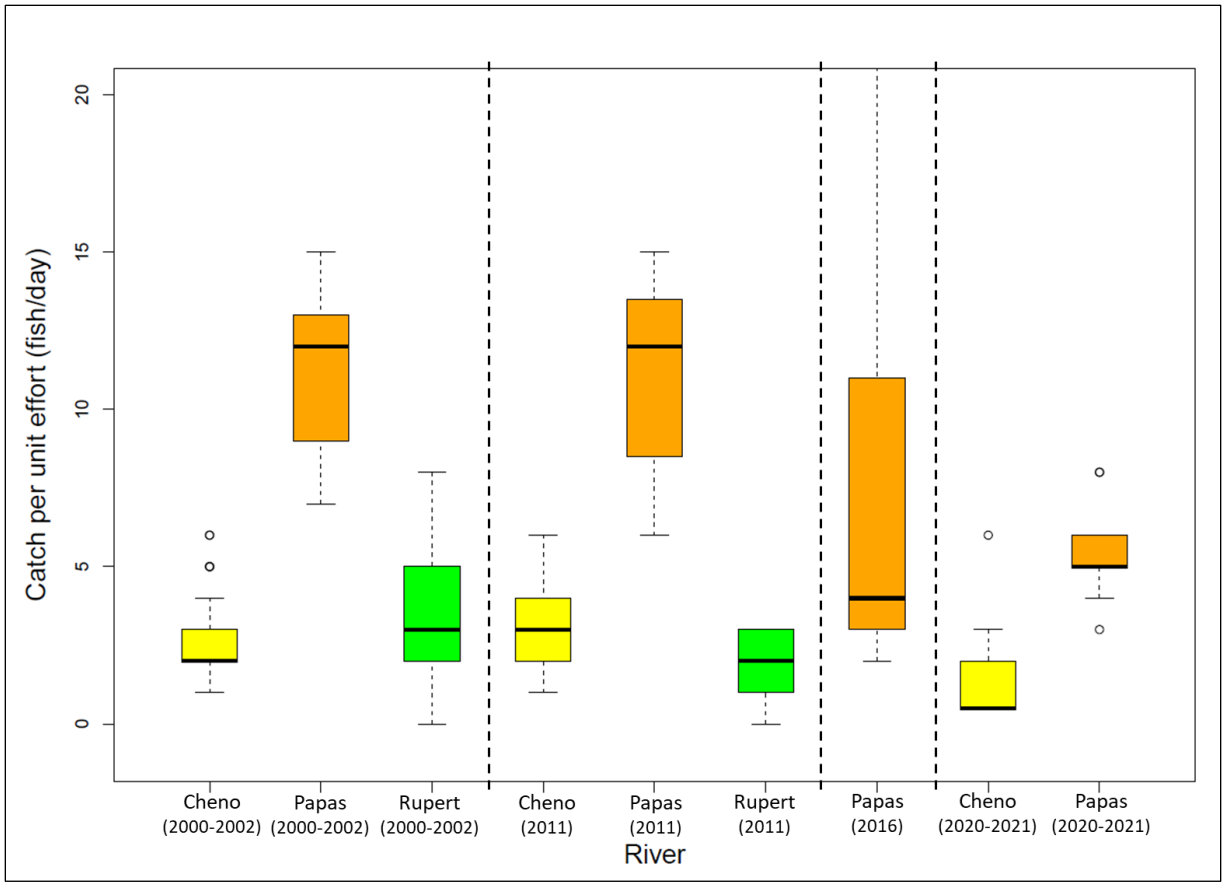

Figure S9. Catch-per-unit-effort in Mistassini Lake source rivers at 4 time periods. Values for catch-per-unit-effort (CPUE) were calculated based on the number of brook trout caught per 8-hour day of sampling per angler. The colors of the boxes correspond to the source river (yellow = Cheno; orange = Papas; green = Rupert). The time periods are shown on the x-axis. For Papas in 2016, the upper whisker extends beyond the plot's limit due to an outlier value of 36.