

**Tissue library and DNA dataset creation for the investigation
of migration patterns of Atlantic Bluefin Tuna (*Thunnus
thynnus*) in Atlantic Canada**

by

Stephanie Maheux

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Bachelor of Science in Marine Biology

in the Graduate Academic Unit of Biological Sciences

Supervisor: Rebekah Oomen, Department of Biological Sciences

THE UNIVERSITY OF NEW BRUNSWICK

April 2025

© Stephanie Maheux, 2025

Abstract

Atlantic Bluefin Tuna (*Thunnus thynnus*) are a migratory fish species that travel from the Gulf of Mexico, the Mediterranean Sea and the Slope Sea to forage in Atlantic Canada, however, the logistical specifics of these migrations are still unclear. The objective of this study was to create a tissue library and genomics database consisting of bluefin tuna samples with which their stock of origin can be identified. With those identifications, general migration timings of each stock through Canada can be determined. Muscle tissue was collected from 226 individuals across Eastern Nova Scotia. Of those samples, DNA was extracted from 51 and sequenced using Low-coverage Whole Genome Sequencing. This DNA created a high-quality genomics database with which researchers can use to genotype bluefin individuals, and, by cross referencing stock with date of capture, can be used to better understand bluefin migration timing.

Dedication

I would like to dedicate this thesis to my wonderful family and friends who showed me endless love and support throughout my undergraduate degree and during my honors project. I would particularly like to thank Amelia Fox and Ryleigh Doucette for taking care of me physically and mentally, driving me around, and letting me talk their ears off about my degree and project, even when they had no idea what I was talking about. To them, I am forever grateful.

Acknowledgements

I would like to acknowledge our collaborators at the Department of Fisheries and Oceans, Saint Andrews Biological Station including Dr. Alex Hanke, Nathan Stewart and Dheeraj Busawon, without whom the collection of all tissue samples used in this study would not have been possible. Their support and guidance was extensive and greatly appreciated.

I would also like to acknowledge, Dr. Scott Pavey and his lab, in particular Larissa Roehl, for their guidance, assistance, and the occasional use of their lab space, supplies and Nanodrop Spectrophotometer. I would also like to acknowledge Dr. Nathalie Leblanc who completed the DNA sequencing and bioinformatics portion of this project.

I would further like to acknowledge my fellow lab members, in particular Abigail Scher and Iris Jia with whom I worked closely on this project.

Table of Contents

Abstract.....	ii
Dedication.....	iii
Acknowledgements.....	iv
Table of Contents.....	v
List of Tables.....	vi
List of Figures.....	vii
List of Symbols, Nomenclature, or Abbreviations.....	viii
Introduction.....	1
Methods.....	6
Tissue Samples.....	6
DNA Extraction.....	8
Low-Coverage Whole Genome Sequencing.....	8
Tissue Library and Database Quality Analysis.....	9
Results.....	12
Tissue Library.....	12
DNA Quality.....	13
Initial DNA Sequencing Quality.....	14
Initial Read Quantity.....	14
Initial Raw Duplicate Quantity.....	15
Adapter Content.....	17
Read Depth and Coverage.....	17
Total SNP Presence.....	19
Discussion.....	21
Tissue and DNA Quality.....	21
Use of low coverage sequencing on a novel reference genome.....	23
Future Research Directions.....	23
Data availability.....	25
Significance.....	26
Contributions to Thesis.....	27
Bibliography.....	28

List of Tables

Table 1: Two-way ANOVA results for tests of differences in nucleotide concentration ($\mu\text{l}/\text{ng}$) and nucleic acid mass (ng) of Atlantic bluefin tuna (<i>Thunnus thynnus</i>) DNA between year and month of capture. No results were statistically significant.	14
Table 2: Post-hoc Tukey test results showing significant relationships ($p < 0.05$) from a two-way ANOVA testing for differences between duplicate % of individual-level DNA sequence libraries between months and years of capture.	17
Table 3: Read quantity of sequenced Atlantic bluefin tuna (<i>Thunnus thynnus</i>) DNA samples upon initial return G�nome Qu�bec organized by year and month.	36
Table 4: Duplicate percentage of sequenced Atlantic bluefin tuna (<i>Thunnus thynnus</i>) DNA samples upon initial return G�nome Qu�bec organized by year and month.	36
Table 5: Adapter content of sequenced Atlantic bluefin tuna (<i>Thunnus thynnus</i>) DNA samples upon initial return G�nome Qu�bec organized by year and month.	36
Table 6: Mean read depth coverage of Atlantic bluefin tuna (<i>Thunnus thynnus</i>) DNA samples post sequencing and alignment to the reference genome.	37

List of Figures

Figure 1: Map displaying all sample locations across eastern Nova Scotia from which Atlantic Bluefin Tuna (<i>Thunnus thynnus</i>) muscle tissue was sourced that was used in this study.....	13
Figure : Boxplot displaying the initial read quantity of sequences from Génome Québec between years and months of sample capture. A two-way ANOVA determined a significant ($p < 0.05$) difference between the read quantity of DNA samples from May and December ($p = 0.025$).	15
Figure : Boxplot displaying the initial duplicate quantity (%) of sequences from Génome Québec between years and months of sample capture. A two-way ANOVA determined significant relationships ($p < 0.05$) between duplicate content and year ($p = 0.004$), month ($p = 0.002$) and year*month ($p = 0.024$) of capture.....	16
Figure 4: Boxplot displaying the mean read depth of samples years and months of capture. A two-way ANOVA determined significant differences ($p < 0.05$) of read depth between May and September ($p = 0.04$), and May and December ($p = 0.02$)......	18
Figure 5: Boxplot displaying the mean read depth of samples years and months of capture. A two-way ANOVA determined no significant differences ($p > 0.05$)......	19
Figure 6: Number of SNPs present per number of individuals. Figure includes 48 samples selected for analysis plus three extra samples that were selected as back-ups. Figure by Dr. Nathalie Leblanc.	20

List of Symbols, Nomenclature, or Abbreviations

CFL: Curved Fork Length – The length of a fish from the front of its upper jaw the to tail fork

COSEWIC: The Committee on the Status of Endangered Wildlife in Canada

DFO: Department of Fisheries and Oceans Canada

ICCAT: International Commission for the Conservation of Atlantic Tunas

lcWGS: Low-Coverage Whole Genome Sequencing

RADseq: Restriction-site Associated DNA sequencing

SNP: Single Nucleotide Polymorphism

TAC: Total Allowable Catch – Represents the total mass of one species a fishery is legally allowed to harvest each year

Introduction

Atlantic bluefin tuna (*Thunnus thynnus*, Scombridae; hereafter 'bluefin tuna') is a large, migrating, pelagic teleost whose range spans much of the northern Atlantic (Rooker et al., 2007). Bluefin tuna are managed by the International Commission for the Conservation of Atlantic Tuna (ICCAT). In the northern hemisphere, bluefin tuna are divided into the western and eastern stocks with the 45° meridian acting as a boundary (ICCAT, 2016). In Canada, charter fishing of these tuna has taken place since the 1930s. Commercial fishing of the species expanded in the 1980s along with the rise of the Japanese sushi-sashimi market, which made the fishery far more profitable than it had been previously (Deonarine, 2019; Fromentin & Powers, 2005). The sudden spike in demand for bluefin tuna caused them to be overfished due to the overcapacity of fishing vessels in international waters, wide geographical expansion of the fishery and lack of effective governance at both national and international levels (Fromentin et al., 2014).

In 1998, a total allowable catch (TAC) system was put in place in an attempt to reduce fishery stress on bluefin tuna stocks. However, from the late 90s to 2008, the TAC limit was still above the recommendations of ICCAT's own scientists. Catches by many fisheries were significantly underreported until 2007, which only furthered the damage to the bluefin tuna population. In 2009, ICCAT officially adopted the lower, scientifically advised TAC, which led to a significant recovery of the populations (Fromentin et al., 2014). Despite their impressive recovery, western populations remain a magnitude smaller than eastern populations, and the species is classified as endangered under the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) (DFO, 2011; ICCAT, 2016; Rodríguez-Ezpeleta et al., 2019). Further investigation into the behaviour

and movements of bluefin tuna is important to make well-informed management decisions for their conservation.

Bluefin tuna from the eastern and western stocks display some divergent life history characteristics, which could relate to differences in their respective population recovery and migration patterns (Fromentin & Powers, 2005; Fromentin et al., 2014). The eastern stock consists of bluefin tuna that spawn in the Mediterranean Sea (Fromentin & Powers, 2005). The specific spawning locations of these Mediterranean populations are locally dispersed, but all subpopulations have a spawning period of between 19 to 31 days, and there is little to no significant genetic differentiation between them (Carlsson et al., 2004; Medina, 2020; Puncer et al., 2018). The western stock is further subdivided into two spawning sites – the principal location in the Gulf of Mexico and the more recently established location in the Slope Sea (Aalto et al., 2023; Fromentin & Powers, 2005; Richardson et al., 2016). Bluefin tuna have an estimated spawning period of 24 days throughout the Gulf of Mexico (Medina, 2020). Since the Slope Sea population is more recently established and was only officially recognized by ICCAT in 2022, there is still little information about the life history characteristics of this subpopulation. However, we know that they are a genetic mix of the Gulf of Mexico and Mediterranean populations but tend to display more Gulf of Mexico-like characteristics, such as spawning at a similar, yet slightly smaller size to Gulf of Mexico individuals (ICCAT, 2023; Medina, 2020). Since the recent release of evidence in support of more genetic mixing between populations, there have been suggestions to remove the 45° meridian line as a basis for stock delineation (ICCAT, 2023). In their most recent report, ICCAT determined that they are retaining the management demarcation line, but plan to consider

genetic mixing in future management plans (ICCAT, 2023). These complexities have increased the need for further research into population dynamics of the species.

Once bluefin tuna from both sides of the management boundary have reached maturity, many will migrate to Atlantic Canada to forage primarily for Atlantic herring (*Clupea harengus*) and Atlantic mackerel (*Scomber scombrus*) (Pleizier et al., 2012; Turcotte et al., 2023). A previous study using telemetry data suggested that bluefin tuna from the Mediterranean stock begin to arrive in Canada between April and May and leave around December (Aalto et al., 2021). The Gulf of Mexico stock has been observed to arrive in Canada between May and June and leave around the same time (Aalto et al., 2021; Dedman et al., 2023). Foraging timing for the Slope Sea population in Atlantic Canada is still not entirely clear. It appears that they display a similar pattern to the Gulf of Mexico stock, but they might arrive slightly later (Aalto et al., 2023). In past studies, the general timelines for when each stock of bluefin tuna arrive and leave the Canadian foraging grounds is fairly consistent. However, the timeline for bluefin tuna presence in Atlantic Canada is still not entirely clear because of varying methodology (for example, acoustic telemetry or microsatellite analysis) and year of publication of these studies (Aalto et al., 2021; Block et al., 2019; Carlsson et al., 2004; Dedman et al., 2023; Wilson et al., 2015). More recent data and contemporary methods may provide more insight into these timelines. It also appears as though some individuals, particularly those from the Mediterranean stock, will not return to their spawning site some years. Rather, they elect to skip one or sometimes multiple years of spawning to continue foraging (Aarestrup et al., 2022; Corriero et al., 2020; Medina, 2020).

Single nucleotide polymorphisms (SNPs) can be used to identify bluefin tuna migrating and foraging in Atlantic Canada to their spawning site of origin. SNPs are the most common type of genetic variation and occur when one nucleotide in a DNA strand switches one of its bases (Sherry et al., 1999; Wright, 2005). All variation in the genome arises from mutations, and the heritability and survival of these mutations is dependent on their effect on reproductive fitness, population dynamics, their location in the chromosome, and recombination rates (Bürger, 1999). In the past, 96 stock-identifying SNPs have been discovered, allowing researchers to assign individual bluefin tunas to their stock of origin with a higher degree of confidence through genotyping than with the use of previous methods, such as otolith microchemistry (Rodríguez-Ezpeleta et al., 2019).

Low-coverage whole genome sequencing (lcWGS) is a recently developed genotyping method, which allows for better genomic analysis on a population scale (Lou et al., 2021). This is accomplished by lowering the read depth, i.e., the number of times a locus is sequenced in each individual. While lowering the read depth can mean sacrificing individual information, this in turn increases the coverage for the population as a whole for the same amount of sequencing effort, making this method useful in population level analyses (Lou et al., 2021).

The objectives of this project are to develop: (1) a tissue library and (2) a genomic database which can be used to investigate if there are differences in migration timing of Atlantic bluefin tuna between the Mediterranean Sea, Gulf of Mexico, and Slope Sea populations in Atlantic Canada. If bluefin tuna are migrating to Atlantic Canada to forage, they not only have to ensure that they arrive at the right time to follow their

preferred prey, but also to migrate back to their spawning site in time for their stock's spawning window (Dedman et al., 2023; Rooker et al., 2007). These foraging and spawning windows are some of the primary drivers of bluefin tuna migration. For my thesis, I created a tissue library containing bluefin tuna muscle tissue collected from the Department of Fisheries and Oceans, Saint Andrews Biological Station (DFO) that future researchers can use to investigate differences in migration timing between different stocks of bluefin tuna throughout Atlantic Canada. Using lcWGS, I also created a genomic database aligned with a recently released bluefin tuna reference genome with which researchers can investigate questions relating to differences in migration timing between different stocks of bluefin tuna through Atlantic Canada.

Methods

Tissue Samples

Tissue samples were taken directly from landed individuals in Canso, Nova Scotia on December 6th, 2023, and from the Wedgeport Tuna Tournament and Festival from August 21st to 24th, 2024 (a short sampling trip to Pinkney's Point was also taken during this time). After the individuals from Canso were landed over the course of the previous month, staff on the wharf removed the heads and stored them at -20 °C. Alongside the DFO team, we cut each head in half and head steaks were removed. From the Wedgeport Tuna Festival, fish were sampled almost immediately after being pulled from the boat. Once the tuna was weighed and the head was removed from the body, the sampling process was the same for both collection locations, except that the Wedgeport Tuna Festival samples were being taken from fresh fish rather than frozen. From these head steaks, we took approximately 100 g of tissue and preserved each sample in 1.5 mL microcentrifuge tubes and 95% ethanol. Samples were kept in a cooler or refrigerator until they were transferred to a -20 °C freezer in the Oomen Laboratory at the Canadian Rivers Institute building on the University of New Brunswick Saint John campus. Each sample was given an identification number by DFO so that metadata for each individual (e.g. landing date, size, weight, capture coordinates) that was taken prior could be added to a database created by DFO and shared with the Oomen Lab. From the same head steak, a larger piece of tissue was taken back to DFO to be added to their tissue archive.

To bolster the tissue library size and obtain data from a broader period, all other samples used in this study were taken from DFO's tissue archive. These samples were originally collected by the DFO sampling team from along the eastern coast of Nova

Scotia using the same methods outlined above. From DFO's tissue archive, I subsampled muscle tissues originally collected between 2021 and 2023. I chose samples from each year by organizing them by catch date in DFO's database and splitting up the samples into 'early season', which included the first third of individuals landed each year, and 'late season', which included the last third of individuals landed each year. These two groups were further split in half, allowing for four time periods to represent when each stock could be present in the Canadian fishery. These seasonal time periods are as follows: early season, early-middle season, late-middle season, and late season. Once these samples were chosen from the database, they were taken from the archive freezers and either transferred directly into centrifuge tubes with fresh 95% ethanol, or if the samples were too large, they were cut down to approximately 100 mg and then added to the centrifuge tubes with 95% ethanol. There were some samples from DFO's tissue library that had less than 100 mg of muscle tissue available. When this was the case, we collected as much tissue from the sample as possible while ensuring sufficient material remained in the DFO archive. These samples were kept in the freezer or on ice until they were transferred to the Oomen Laboratory and to be added to the tissue library stored at -20 °C.

After all sampling was completed, 4 samples from each of the 4 outlined time periods from 2021 to 2023 were used for genetic analysis, resulting in with a total sample size of 48 for the DNA library, with three extra samples added as backups for those with lower nucleotide concentration values, as measured with a NanoDrop 2000 Spectrophotometer (Thermofisher Scientific). If there were more than four available

within each seasonal time period, samples that were used for DNA extraction were haphazardly selected.

DNA Extraction

DNA from tissue samples was extracted using the Omega Bio-Tek's E.Z.N.A.[®] Tissue DNA Kit, following the manufacturer's instructions. Before extraction began, tissue from each sample was cut down to 30 mg. If there was less than 30 mg of tissue available, all tissue was added, and the mass of the sample was recorded. Between the lysis step and DNA purification step of the protocol, samples were left overnight in a shaking incubator at 55 °C with a speed of 120 rpm. DNA was eluted in two rounds at 25 µl per round and was centrifuged for one minute in between. Extracted DNA was stored in 1.5 mL microcentrifuge tubes in the freezer at -20 °C.

After extraction, each DNA sample was tested for its nucleotide concentration on a NanoDrop 2000 Spectrophotometer (ThermoFisher Scientific). If samples had too low of a nucleotide concentration (below 70 ng/µl) and there was remaining tissue that could be used, DNA extraction was attempted a second time. If nucleotide concentration was over 150 ng/µl, samples were diluted and retested until concentration was at or below 150 ng/µl per the requirements of the sequencing protocol.

Low-Coverage Whole Genome Sequencing

For genome sequencing of all individuals, DNA samples were sent to the G enome Qu ebec sequencing centre for Low-Coverage Whole Genome Sequencing (lcWGS). This

genotyping method includes library preparation and an initial short read sequence (Lou et al., 2021). Then, shotgun sequencing is used whereby many, randomly placed reads are copied from the DNA sequence. The genome is then reconstructed from the overlapping reads aligned to the reference genome (Gan, 2023; Motahari et al., 2013).

Low quality read filtering and adapter trimming was conducted on the 3' end of DNA strands, as they interfere with downstream alignment (Kircher et al., 2011; Lou et al., 2021). This was done using *Trimmomatic 0.39* (Bolger et al., 2014). Data was then mapped to the reference genome created from an individual of Mediterranean stock (NCBI accession number: PRJEB72088) (Lou et al., 2021; NCBI, 2024).

Tissue Library and Database Quality Analysis

Nucleotide concentration and nucleic acid mass of the samples selected for the DNA database were determined using a NanoDrop 2000 Spectrophotometer (ThermoFisher Scientific). These values were used as a proxy to determine the overall quality of library samples. Because the oldest samples in the tissue library (from 2021) were used in the DNA library, all other samples not being used (i.e. remaining in the tissue library) should be of similar quality or higher, as there has been less time for the tissue and DNA to degrade (Bhoyar et al., 2024). The ideal nucleic acid concentration for the purposes of this project is 70 $\mu\text{l}/\text{ng}$ per the sequencing instructions. Nucleic acid mass (ng) is desired to be as high as possible to maximize the utility of the DNA library resource. A two-way ANOVA was run to determine if there is a difference in tissue quality between years and months of capture.

To assess the quality of the DNA library, initial read quantity, duplicate quantity, and initial adapter quantity was examined. Ideally, duplicate quantity will be 20% or lower in order to limit false confidence in SNP calls as much as possible (Tin et al., 2015). As adapters are added to DNA fragments shorter than the target read length, higher adapter content would indicate shorter, more degraded and possibly lower quality reads, ideally, initial adapter quantity will be below 5% of sequences or lower (Mohideen et al., 2020). Three, two-way ANOVAs were run to determine if there was a difference in initial read quantity, duplicate quantity, and adapter content between years and months of capture. Months of capture were used for analysis instead of the previous seasonal time periods laid out, as the months in which each time period covers varies each year based on sample availability. Using month in summary statistics instead allowed for a more consistent overview of sample quality throughout each year.

Quality was further assessed by read depth and coverage once aligned to the reference genome by following the pipeline provided by the developers of lcWGS: (1) building reference index files for sequence alignment using *bowtie2*, (2) aligning short reads with each Fastq file using *bowtie2*, (3) filtering out reads with alignment quality lower than 20, (4) removing PCR duplicates and overlapping read pairs with *Picard Tools MarkDuplicates* and *BAMUtilclipOverlap* then finally, (5) estimating read depth using *samtools depth* and coverage using *samtools coverage* (GATK, 2024; Jun et al., 2015; Langmead & Salzberg, 2012; Li et al., 2009; Therkildsen et al., 2024). The goal for read depth was an average of 4 reads, per allele per sample. For read coverage, samples will ideally have as close to 100% coverage as possible. Two, two-way ANOVAs were

run to determine if there are significant differences between read depth and coverage of samples between year and month of capture.

Results

Tissue Library

After collection for the tissue library was completed and DNA was extracted from selected samples, the library contained a total of 229 tissue samples with 214 samples either remaining untouched or still containing enough tissue for future use. From Canso, Nova Scotia, 43 tissue samples were taken for the tissue library, none of which were used for the DNA database and therefore remain intact. From the Wedgeport Tuna Festival, in Wedgeport, Nova Scotia, a combined total of 59 samples were taken for the tissue library from both fresh and frozen bluefin tuna (frozen samples were collected on a side trip to Pinkney's Point during the festival), all of which were not used for the DNA database and remain complete. From DFO's tissue archive, I collected 127 subsamples, including those originally collected from wharfs across eastern Nova Scotia including: Wedgeport, Canso, Eastern Passage, Sambro, Liverpool, Woods Harbour, Pinkney's Point and L'Ardoise (Figure 1). From those 127 samples, I used 48 for genetic analysis. Of those 48 samples, 45 still contain enough tissue for a second round of DNA extraction. From these samples, I created a metadata file containing information collected in tandem with DFO on the individual's day and location of capture, curved fork length (CFL), and weight to inform future studies.



Figure 1: Map displaying all sample locations across eastern Nova Scotia from which Atlantic bluefin tuna (*Thunnus thynnus*) muscle tissue was sourced for this study.

DNA Quality

For the 48 samples from which DNA was collected and submitted to Génome Québec for lcWGS, the average nucleotide concentration was 108.6 ng/μl with a standard deviation of 33 ng/μl. The average nucleic acid mass per sample of 3991.7 ng with a standard deviation of 1422 ng. Two, two-way ANOVAs were run to determine if there was a difference in nucleotide concentration (ng/μl) and nucleic acid mass (ng) between collection year and collection month of samples. Both two-way ANOVAs revealed that there was no significant relationship between nucleotide concentration or nucleic acid mass and year or month of capture ($p > 0.05$)(Table 1).

Table 1: Two-way ANOVA results for tests of differences in nucleotide concentration ($\mu\text{l}/\text{ng}$) and nucleic acid mass (ng) of Atlantic bluefin tuna (*Thunnus thynnus*) DNA between year and month of capture. No results were statistically significant.

	Degrees of Freedom	F statistic	P value
Nucleotide Concentration*Year	1	0.891	0.350
Nucleotide Concentration*Month	1	0.001	0.979
Nucleic Acid Mass*Year	1	0.511	0.478
Nucleic Acid Mass*Month	1	0.133	0.718

Initial DNA Sequencing Quality

Initial Read Quantity

The average read quantity of samples upon their return from G enome Qu ebec was 25e+6 base pairs with a standard deviation of 3e+6 base pairs (Table 3, Appendix). There was no significant relationship between read quantity, year and month*year. However, there was a significant relationship between read quantity and month ($p = 0.02$). A post-hoc Tukey test was conducted to determine which months drove the significant relationship with read quantity. The Tukey test determined that the difference in read quantity between May and December was significant ($p = 0.025$) (Figure 2). It is worth noting that 2023 was the only year in which samples from May were available. There were also no samples available from August in 2023 and no samples available from September in 2022.

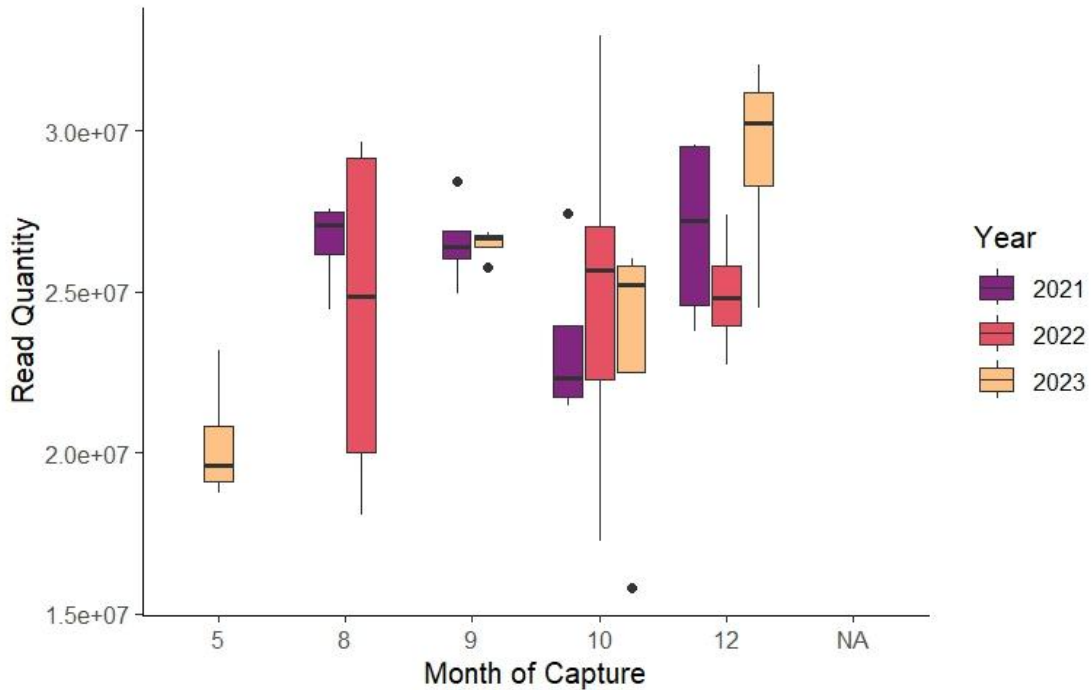


Figure 22: Boxplot displaying the initial read quantity of sequences from Génome Québec between years and months of sample capture. A two-way ANOVA determined a significant ($p < 0.05$) difference between the read quantity of DNA samples from May and December ($p = 0.025$).

Initial Raw Duplicate Quantity

The average duplicate content of samples upon their return from Génome Québec was 19% with a standard deviation of 2.2% (Table 4, Appendix). The two-way ANOVA determined that there was a significant relationship ($p < 0.05$) between duplicate content and year of capture ($p = 0.004$), month of capture ($p = 0.002$), and month*year of capture ($p = 0.024$) (Figure 3). Note that 2023 was the only year in which samples from May were available. There were no samples available from August in 2023, and no samples available from September in 2022.

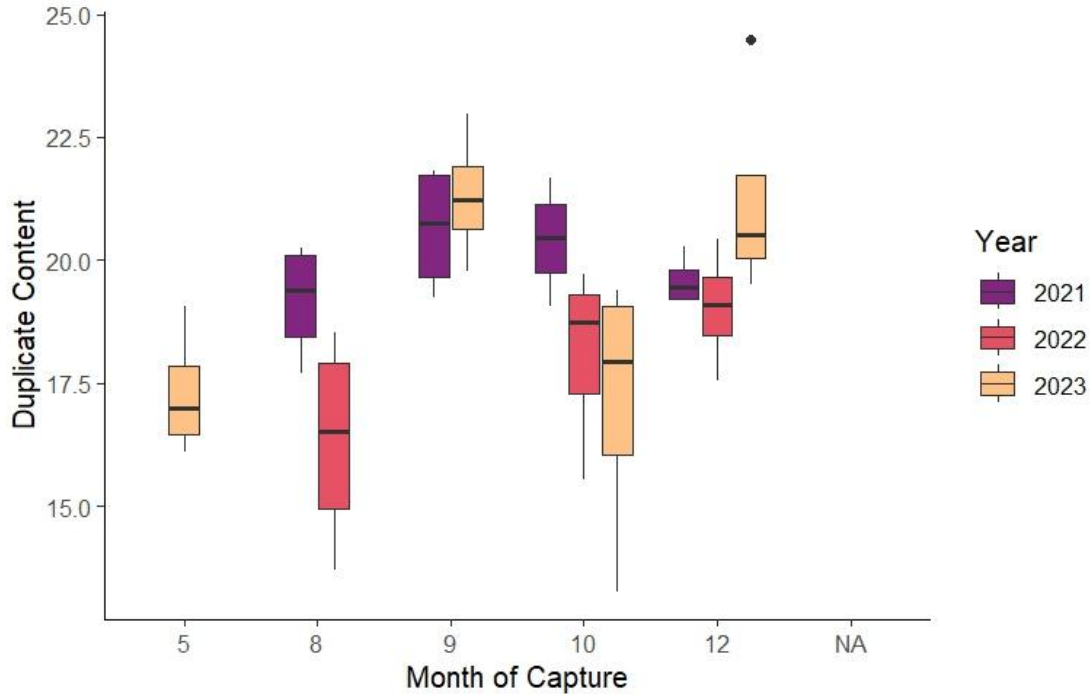


Figure 33: Boxplot displaying the initial duplicate quantity (%) of sequences from G enome Qu ebec between years and months of sample capture. A two-way ANOVA determined significant relationships ($p < 0.05$) between duplicate content and year ($p = 0.004$), month ($p = 0.002$) and year*month ($p = 0.024$) of capture.

A further post-hoc Tukey test was conducted to determine which months and years drove the significant relationship with duplicate content. The years that displayed the most significant pairwise differences in duplicate content were 2021 compared to 2022 (Table 2). The months that displayed significant differences in duplicate content were May compared to September and December, and August compared to September and December (Table 2). Between year*month, the combinations that displayed significant differences were August 2022 compared to September 2021/2023 and October 2023, and December 2023 compared to September 2023 (Table 2).

Table 2: Post-hoc Tukey test results showing significant relationships ($p < 0.05$) from a two-way ANOVA testing for differences between duplicate % of individual-level DNA sequence libraries between months and years of capture.

Formulae	Time period	P-value
Duplicate % ~ Month	May- September	0.015
	May- December	0.030
	August - September	0.024
	August - December	0.045
Duplicate % ~ Year	2022-2023	0.003
Duplicate % ~ Month*Year	September, 2021 - August, 2022	0.034
	September, 2023 - August, 2022	0.007
	October, 2023 - August, 2022	0.048
	December, 2023 - September, 2023	0.008

Adapter Content

The average adapter content of sequences from Génome Québec was 0.17% with a standard deviation of 0.09% (Table 5, Appendix). A Two-way ANOVA was run to determine whether there was a significant relationship between initial adapter content compared to year and month of capture. The ANOVA showed that there were no significant differences in adapter content between month and year of capture.

After adapters were trimmed using *Trimmomatic 0.39*, adapter content for all samples were removed and are no longer a concern for downstream analysis.

Read Depth and Coverage

The average read depth of samples overall was 4.03x with a standard deviation of 2.5x (Table 5, Appendix). In total, 62.5% of samples achieved the set read depth goal of 4x reads or higher, 29% of samples reached a read depth of between 3-4x, and the remaining 8.5% of samples reached a read depth of between 2-3x. A two-way ANOVA was conducted to determine if there were any significant differences in read depth

between year and month of capture ($p < 0.05$). The ANOVA determined that there was a significant difference between read depths and month of capture ($p = 0.014$). A post-hoc Tukey test was further conducted to determine which months showed significant differences. The Tukey test showed that there were significant differences in read depth between May and September ($p = 0.04$) and May and December ($p = 0.02$) (Figure 4).

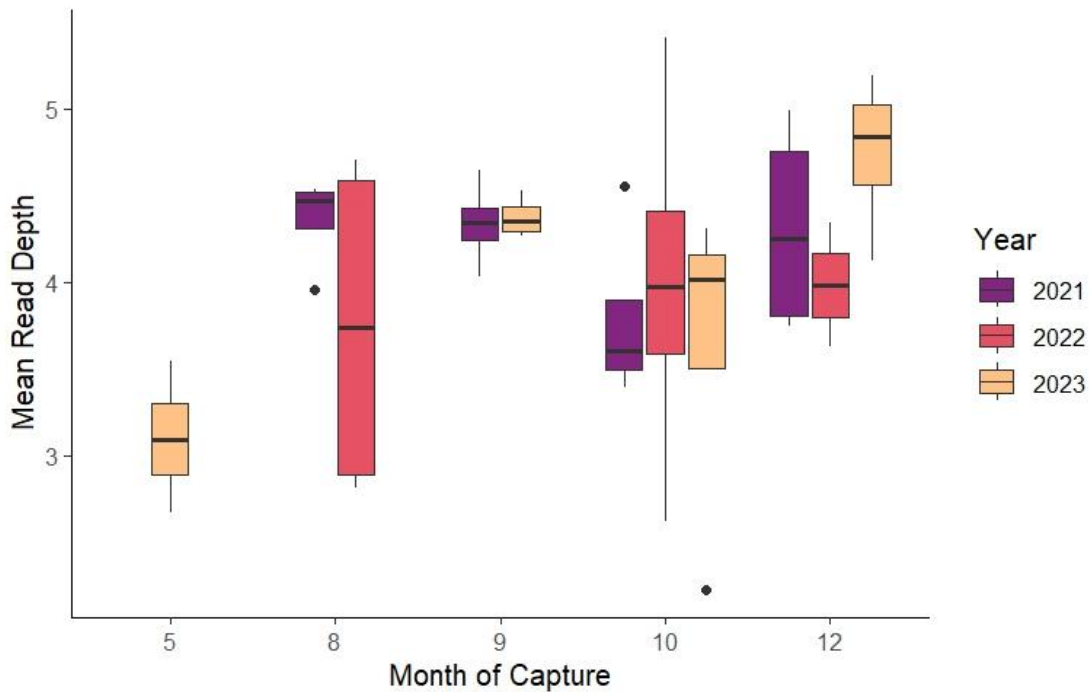


Figure 4: Boxplot displaying the mean read depth of samples years and months of capture. A two-way ANOVA determined significant differences ($p < 0.05$) of read depth between May and September ($p = 0.04$), and May and December ($p = 0.02$).

The average read coverage of all analyzed samples was 90.5% (Table 6, Appendix). A two-way ANOVA was conducted to determine if there were any significant differences in coverage between year and month of capture. The ANOVA revealed no significant differences in coverage between months and year of capture (Figure 5).

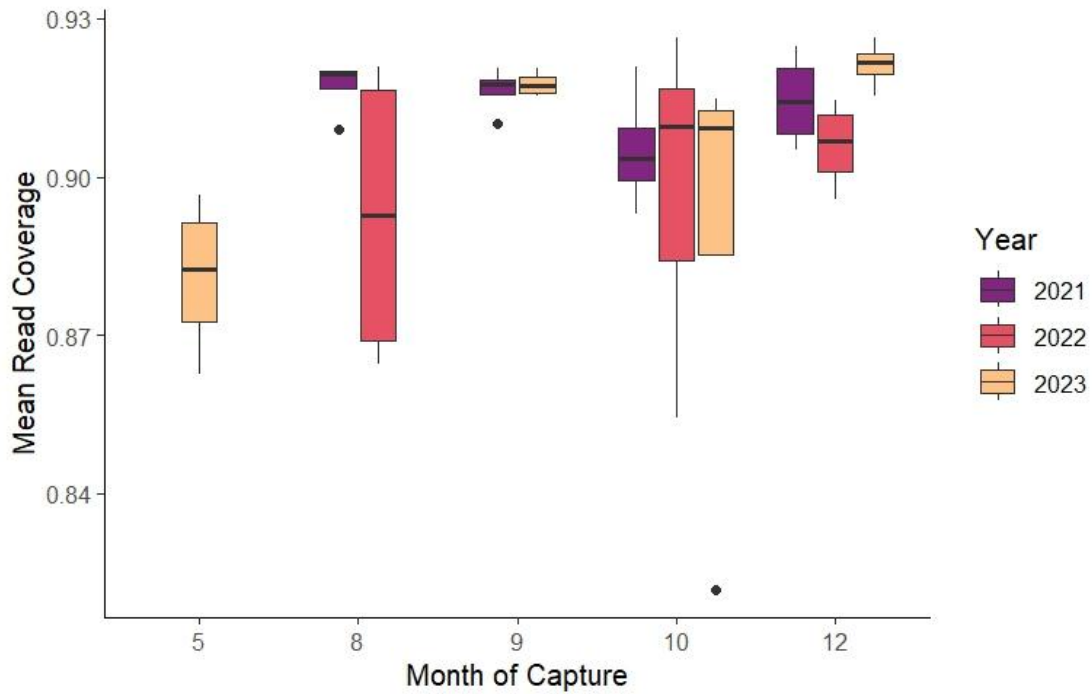


Figure 5: Boxplot displaying the mean read depth of samples years and months of capture. A two-way ANOVA determined no significant differences ($p > 0.05$).

Total SNP Presence

After alignment of the samples with the reference genome and SNPs were identified, it was determined that the vast majority (approximately $2.5e+8$) of SNPs identified in the reference genome were also identifiable in all or most individuals that were sampled (Figure 6).

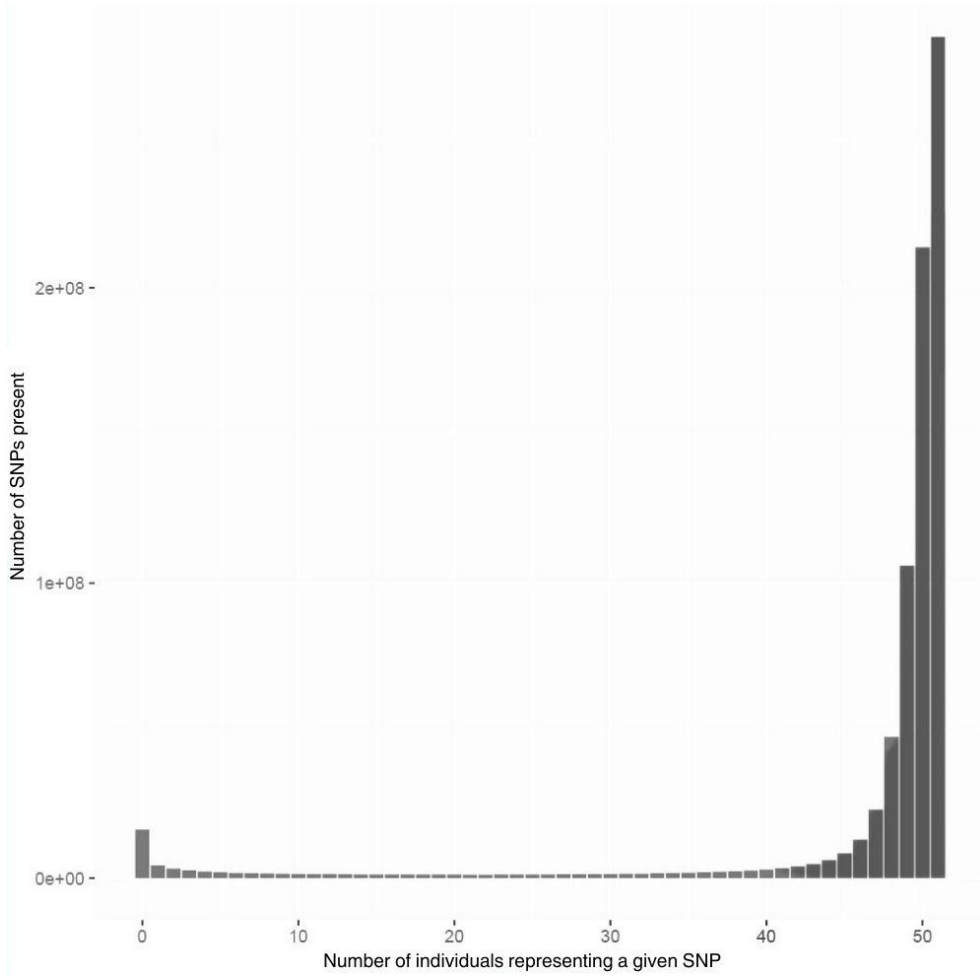


Figure 6: Number of SNPs present per number of individuals. Figure includes 48 samples selected for analysis plus three extra samples that were selected as back-ups. Figure by Dr. Nathalie Leblanc.

Discussion

Tissue and DNA Quality

Over the course of this study, I achieved my objectives of creating a high-quality tissue library and genomic database which can be used to investigate differences in migration timing of Atlantic bluefin tuna between the Mediterranean Sea, Gulf of Mexico, and Slope Sea populations in Atlantic Canada.

From the DNA extracted from bluefin tuna tissue during this study, all but three samples met the 70 ng/ μ l nucleotide concentration threshold. All samples passed the quality control inspection conducted by G enome Qu ebec prior to sequencing. Because some of the oldest and therefore likely the most degraded samples I collected were among the samples sent out for sequencing, we can assume that all other samples within the tissue library will be either of similar or higher quality.

According to quality control analyses conducted after post-sequencing processing of the raw reads, all DNA samples were of high enough quality to align to the reference genome at or near the target read depth of 4x and with sufficient coverage. Further, approximately 2.5×10^8 SNPs were able to be called for 51 individuals, which far exceeds the past numbers of SNPs required for population genomic inference from other studies (Alexander et al., 2009; Patterson et al., 2006). For example, in a similar study that successfully used restriction-site associated DNA sequencing (RADseq) to identify bluefin tuna to their stock of origin, 25,000 SNPs were identified; this is still substantial, but my results yielded a magnitude more (D iaz-Arce et al., 2024). This remains substantial even considering that many are likely to be filtered out downstream, as only the most informative SNPs will be retained for better computational efficiency (Roshyara

et al., 2014). Based on the numbers of SNPs available for all individuals, we can expect that these samples will also be of high enough quality to genotype individuals to their stock of origin.

There was little evidence of samples from one specific month of year of capture being of lower quality than those collected from other time periods. However, it could be argued that samples collected from May (exclusively collected in 2023) could be of lower quality than some of the others. Samples from May had significantly lower read quantity and read depth than samples from other months in the dataset. On the other hand, samples from May outperformed samples from other months when it came to their lower duplicate content. While this could be indicative of samples from May being of lower quality, there being so few samples from May from only one year could also have skewed results.

We can conclude that both the tissue library and genomics dataset created are of high enough quality to genotype individuals to either the Mediterranean Sea, the Gulf of Mexico, or the Slope Sea to investigate whether there is a difference in migration timing between the three populations. It is worth noting that as the Slope Sea population is a genetic mix of mostly Gulf of Mexico-associated alleles, but also a smaller proportion of Mediterranean-associated alleles. Therefore, it may be more difficult to distinguish a genotype signature for individuals from the Slope Sea. However, with the high amount of SNPs that were able to be called and the recent release of two reference genomes for Atlantic bluefin tuna (one from the Mediterranean, to which my data is aligned, and one from the Gulf of Mexico, which could be used to ascertain additional SNPs in the future) we expect that distinguishing the Slope Sea population will be possible in future studies.

Use of low coverage sequencing on a novel reference genome

Unlike this study, all previous studies examining bluefin tuna SNPs have been conducted using RADseq instead of whole genome methods. Low-coverage whole genome sequencing provides a broader range of coverage across the genome compared to RADseq and provides more general information for population level analyses as opposed to individual analyses (Lou et al., 2021).

For example, since RADseq only sequences short fragmented segments across the genome, whole genome is better at identifying runs of homozygosity (Martchenko & Shafer, 2023). Runs of homozygosity refers to regions in the genome where the haplotypes that are inherited from each parent are identical copies derived for shared ancestry, which can include arrays of SNPs (Ceballos, Hazelhurst, et al., 2018). They are informative about levels of inbreeding and population bottlenecks, therefore, populations with more admixture tend to have less runs of heterozygosity (Ceballos, Joshi, et al., 2018). Since the Slope Sea population is thought to experience higher levels of admixture than the other two stocks, the presence or absence of certain SNPs tied to runs of heterozygosity could be helpful in their identification. This study capitalized on the recent publication and opportunity for first-time use of the first chromosome-level reference genome assembly for Atlantic bluefin tuna. Our whole genome approach ensured that as many SNPs were identified as possible for use in future research.

Future Research Directions

As previously mentioned, I created this genomic dataset for the purpose of future researchers being able to use it to investigate if there is a difference in migration timing of Atlantic bluefin tuna between the Mediterranean Sea, Gulf of Mexico and Slope Sea

populations while in Atlantic Canada through the genotyping of landed individuals. I hypothesize that future researchers will find there will be a difference in migration timing between spawning populations. I formed this hypothesis because, if bluefin tuna are migrating to Atlantic Canada to forage, they not only have to ensure they arrive at the right time to follow their preferred prey, but they also have to ensure they migrate back to their spawning site in time for their stock's spawning window if they are in fact spawning that year (Dedman et al., 2023; Rooker et al., 2007). These spawning windows vary between populations, which could be a driver of any variation in migration timing (Richardson et al., 2016).

Based on this hypothesis, I predict that Mediterranean individuals will migrate through Atlantic Canada earlier and remain to forage for longer than other populations (Aalto et al., 2023; Dedman et al., 2023). While this may seem counterintuitive, this may be because many of these Mediterranean individuals are not spawning that year, and therefore do not have a 'deadline' for when they have to be back in the Mediterranean (Corriero et al., 2020). Gulf of Mexico individuals will arrive later in the season and stay for a shorter period of time (Aalto et al., 2023; Dedman et al., 2023). Slope Sea individuals will arrive last and leave at a similar time to the Gulf of Mexico population (Aalto et al., 2023). These predictions are based on previous studies that have also investigated bluefin migration using acoustic telemetry data (Aalto et al., 2023; Dedman et al., 2023). However, these studies do not come to a consensus on the specific migration timing of bluefin populations, nor is the specific presence of bluefin tuna in Atlantic Canada the priority of these studies. Furthermore, since the Slope Sea is newly designated as an official bluefin spawning site, there are limited studies taking those

individuals into consideration in their investigations and therefore large gaps in our ecological understanding of them.

Data availability

The genomic dataset I created over the course of this project will initially, only be available for members of the Atlantic Bluefin Tuna Consortium. Members of the consortium include fellow bluefin tuna researchers from Canada, Norway, Sweden, Italy, Malta, and the United Kingdom. I will be limiting initial access to this group of researchers as it was the reference genome developed by this group that was used to align with my DNA sequences. Furthermore, this limited initial release to a smaller group of experienced and knowledgeable bluefin tuna researchers will enable me to get feedback on accessibility and usability before making it openly accessible to the general public.

In the future, the data will be openly available through either the National Center for Biotechnology Information (NCBI) or the European Nucleotide Archive (ENA) for all international researchers to use for the purpose of further investigation into migration differences of stocks of bluefin tuna through Atlantic Canada and possibly other questions of interest. For example, open release of this data would occur alongside an open-access publication to ensure those that wish to use it would have adequate information about its intended purpose, source, and the methods through which it was developed. Release of this data set would be conducted in accordance with the FAIR principles (Findable, Accessible, Interoperable and Reusable) for data management (Wilkinson et al., 2016).

As part of the data release, I will also release a corresponding, master Excel sheet that will contain metadata information about the individual each sample was taken from

(day and location of capture, curved fork length, weight, location of sample and whether DNA has been extracted and/or used in the database), NanoDrop values of DNA samples, and quality control information from FastQC reports.

Access to the tissue library will be available to fellow lab members and collaborators investigating bluefin tuna migration in Canada using genomic methods.

Significance

Bluefin tuna have a recent, concerning past of overfishing and exploitation that has led them to still be classified as ‘endangered’ under COSEWIC. Despite this, the bluefin tuna fishing industry in Canada still brings in approximately \$10 million CAD per year (DFO, 2019).

It has only been in recent years that ICCAT has taken migration and mixing of western and eastern stocks into account in management discussion and regulations (ICCAT, 2016). Since Canada holds 22.34% of the Western total allowable catch for bluefin tuna, it is necessary to be well informed about the demographics and fishing pressures of the stocks we are fishing from in Atlantic Canada at different times of the season. This knowledge can prevent past mistakes from being repeated and to prevent new mistakes being made due to a lack of ecological and management data (DFO, 2018).

Contributions to Thesis

Collection of many of the samples used in this project was completed by the sampling team at the DFO, Saint Andrews lead by Dheeraj Busawon. The fish from which these samples were collected from were landed by commercial fishers licenced to fish in ICCAT management area BF50 and BF51.

Extracted DNA samples were sent to Génome Québec for whole genome sequencing.

The lcWGS pipeline and method used in this project was developed by Dr. Nina Therkildsen and her lab at Cornell University. Dr. Nathalie Leblanc was contracted to execute this pipeline on my dataset.

Bibliography

- Aalto, E. A., Dedman, S., Stokesbury, M. J. W., Schallert, R. J., Castleton, M., & Block, B. A. (2023). Evidence of bluefin tuna (*Thunnus thynnus*) spawning in the Slope Sea region of the Northwest Atlantic from electronic tags. *ICES Journal of Marine Science*, 80(4), 861–877. <https://doi.org/10.1093/icesjms/fsad015>
- Aalto, E. A., Ferretti, F., Lauretta, M. V., Walter, J. F., Stokesbury, M. J. W., Schallert, R. J., & Block, B. A. (2021). Stock-of-origin catch estimation of Atlantic bluefin tuna (*Thunnus thynnus*) based on observed spatial distributions. *Canadian Journal of Fisheries and Aquatic Sciences*, 78(8), 1193–1204. <https://doi.org/10.1139/cjfas-2019-0445>
- Aarestrup, K., Baktoft, H., Birnie-Gauvin, K., Sundelöf, A., Cardinale, M., Quilez-Badia, G., Onandia, I., Casini, M., Nielsen, E. E., Koed, A., Alemany, F., & MacKenzie, B. R. (2022). First tagging data on large Atlantic bluefin tuna returning to Nordic waters suggest repeated behaviour and skipped spawning. *Scientific Reports*, 12(1), 11772. <https://doi.org/10.1038/s41598-022-15819-x>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Bhoyar, L., Mehar, P., & Chavali, K. (2024). An overview of DNA degradation and its implications in forensic caseworks. *Egyptian Journal of Forensic Sciences*, 14(1), 15. <https://doi.org/10.1186/s41935-024-00389-y>

- Block, B. A., Whitlock, R., Schallert, R. J., Wilson, S., Stokesbury, M. J. W., Castleton, M., & Boustany, A. (2019). Estimating Natural Mortality of Atlantic Bluefin Tuna Using Acoustic Telemetry. *Scientific Reports*, 9(1), 4918. <https://doi.org/10.1038/s41598-019-40065-z>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Carlsson, J., McDOWELL, J. R., Díaz-Jaimes, P., Carlsson, J. E. L., Boles, S. B., Gold, J. R., & Graves, J. E. (2004). Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Molecular Ecology*, 13(11), 3345–3356. <https://doi.org/10.1111/j.1365-294X.2004.02336.x>
- Ceballos, F. C., Hazelhurst, S., & Ramsay, M. (2018). Assessing runs of Homozygosity: A comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics*, 19(1), 106. <https://doi.org/10.1186/s12864-018-4489-0>
- Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics*, 19(4), 220–234. <https://doi.org/10.1038/nrg.2017.109>
- Corriero, A., Heinisch, G., Rosenfeld, H., Katavić, I., Passantino, L., Zupa, R., Grubišić, L., & Lutcavage, M. E. (2020). Review of Sexual Maturity in Atlantic Bluefin Tuna, *Thunnus thynnus* (Linnaeus, 1758). *Reviews in Fisheries Science & Aquaculture*, 28(2), 182–192. <https://doi.org/10.1080/23308249.2019.1685456>

- Dedman, S., Aalto, E. A., Stokesbury, M. J. W., Schallert, R. J., Castleton, M. R., & Block, B. A. (2023). Assignment of tracks from tagged Atlantic bluefin tuna *Thunnus thynnus* to potential stocks using behavioural differences and habitat partitioning. *Frontiers in Marine Science*, *10*, 1165910.
<https://doi.org/10.3389/fmars.2023.1165910>
- Deonaraine, D. (2019). *The history of fisheries management of western Atlantic Bluefin tuna in the Canadian Maritimes Gulf region*. Department of Fisheries and Oceans.
- DFO. (2011). *RECOVERY POTENTIAL ASSESSMENT FOR WESTERN ATLANTIC BLUEFIN TUNA (THUNNUS THYNNUS) IN CANADIAN WATERS*.
- DFO. (2018). *Bluefin Tuna*. <https://www.dfo-mpo.gc.ca/species-especes/profiles-profil/bluefin-tuna-thon-rouge-eng.html>
- DFO. (2019). *Tuna*. <https://www.dfo-mpo.gc.ca/international/tuna-thon/index-eng.htm>
- Fromentin, J., & Powers, J. E. (2005). Atlantic bluefin tuna: Population dynamics, ecology, fisheries and management. *Fish and Fisheries*, *6*(4), 281–306.
<https://doi.org/10.1111/j.1467-2979.2005.00197.x>
- Fromentin, J.-M., Bonhommeau, S., Arrizabalaga, H., & Kell, L. T. (2014). The spectre of uncertainty in management of exploited fish stocks: The illustrative case of Atlantic bluefin tuna. *Marine Policy*, *47*, 8–14.
<https://doi.org/10.1016/j.marpol.2014.01.018>

- Gan, J. H. (2023). Unlocking genetic mysteries: The power of shotgun sequencing. *Journal of Bacteriology and Infectious Diseases*, 7(5), 161.
<https://doi.org/10.35841/jgdd-7.5.161>
- GATK. (2024). *Picard MarkDuplicates* [Computer software].
<https://gatk.broadinstitute.org/hc/en-us/articles/360037052812-MarkDuplicates-Picard>
- ICCAT. (2016). *ICCAT Manual*. International Commission for the Conservation of Atlantic TUNA.
- ICCAT. (2023). *R E P O R T for biennial period, 2022-23 PART II (2023)—Vol. 2.*
chrome-
extension://efaidnbmnnnibpcajpcgclefindmkaj/https://www.iccat.int/Documents/BienRep/REP_EN_22-23-II-2.pdf
- Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Research*, 25(6), 918–925.
<https://doi.org/10.1101/gr.176552.114>
- Kircher, M., Heyn, P., & Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*, 12(1), 382.
<https://doi.org/10.1186/1471-2164-12-382>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner’s guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, 30(23), 5966–5993. <https://doi.org/10.1111/mec.16077>
- Martchenko, D., & Shafer, A. B. A. (2023). Contrasting whole-genome and reduced representation sequencing for population demographic and adaptive inference: An alpine mammal case study. *Heredity*, 131(4), 273–281. <https://doi.org/10.1038/s41437-023-00643-4>
- Medina, A. (2020). Reproduction of Atlantic bluefin tuna. *Fish and Fisheries*, 21(6), 1109–1119. <https://doi.org/10.1111/faf.12489>
- Mohideen, A. M. S. H., Johansen, S. D., & Babiak, I. (2020). High-Throughput Identification of Adapters in Single-Read Sequencing Data. *Biomolecules*, 10(6), 878. <https://doi.org/10.3390/biom10060878>
- Motahari, A. S., Bresler, G., & Tse, D. N. C. (2013). Information Theory of DNA Shotgun Sequencing. *IEEE Transactions on Information Theory*, 59(10), 6273–6289. <https://doi.org/10.1109/TIT.2013.2270273>
- NCBI. (2024). *Genome assembly fThuThy2.1* [Dataset]. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_963924715.1/

- Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>
- Pleizier, N. K., Campana, S. E., Schallert, R. J., Wilson, S. G., & Block, B. A. (2012). Atlantic Bluefin Tuna (*Thunnus thynnus*) Diet in the Gulf of St. Lawrence and on the Eastern Scotian Shelf. *Journal of Northwest Atlantic Fishery Science*, 44, 67–76. <https://doi.org/10.2960/J.v44.m685>
- Puncher, G. N., Cariani, A., Maes, G. E., Van Houdt, J., Herten, K., Cannas, R., Rodriguez-Ezpeleta, N., Albaina, A., Estonba, A., Lutcavage, M., Hanke, A., Rooker, J., Franks, J. S., Quattro, J. M., Basilone, G., Fraile, I., Laconcha, U., Goñi, N., Kimoto, A., ... Tinti, F. (2018). Spatial dynamics and mixing of bluefin tuna in the Atlantic Ocean and Mediterranean Sea revealed using next-generation sequencing. *Molecular Ecology Resources*, 18(3), 620–638. <https://doi.org/10.1111/1755-0998.12764>
- Richardson, D. E., Marancik, K. E., Guyon, J. R., Lutcavage, M. E., Galuardi, B., Lam, C. H., Walsh, H. J., Wildes, S., Yates, D. A., & Hare, J. A. (2016). Discovery of a spawning ground reveals diverse migration strategies in Atlantic bluefin tuna (*Thunnus thynnus*). *Proceedings of the National Academy of Sciences*, 113(12), 3299–3304. <https://doi.org/10.1073/pnas.1525636113>
- Rodríguez-Ezpeleta, N., Díaz-Arce, N., Walter, J. F., Richardson, D. E., Rooker, J. R., Nøttestad, L., Hanke, A. R., Franks, J. S., Deguara, S., Lauretta, M. V., Addis, P., Varela, J. L., Fraile, I., Goñi, N., Abid, N., Alemany, F., Oray, I. K., Quattro, J. M., Sow, F. N., ... Arrizabalaga, H. (2019). Determining natal origin for

improved management of Atlantic bluefin tuna. *Frontiers in Ecology and the Environment*, 17(8), 439–444. <https://doi.org/10.1002/fee.2090>

Rooker, J. R., Alvarado Bremer, J. R., Block, B. A., Dewar, H., De Metrio, G., Corriero, A., Kraus, R. T., Prince, E. D., Rodríguez-Marín, E., & Secor, D. H. (2007). Life History and Stock Structure of Atlantic Bluefin Tuna (*Thunnus thynnus*). *Reviews in Fisheries Science*, 15(4), 265–310. <https://doi.org/10.1080/10641260701484135>

Roshyara, N. R., Kirsten, H., Horn, K., Ahnert, P., & Scholz, M. (2014). Impact of pre-imputation SNP-filtering on genotype imputation results. *BMC Genetics*, 15(1), 88. <https://doi.org/10.1186/s12863-014-0088-5>

Sherry, S. T., Ward, M., & Sirotkin, K. (1999). dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Research*, 9(8), 677–679. <https://doi.org/10.1101/gr.9.8.677>

Therkildsen, N. O., Linderøth, T., Jacobs, A., & Lou, N. (2024). *Physalia course on Population genomic inference from low-coverage whole-genome sequencing data*. <https://github.com/nt246/physalia-lcwgs/tree/main>

Tin, M. M. Y., Rheindt, F. E., Cros, E., & Mikheyev, A. S. (2015). Degenerate adaptor sequences for detecting PCR duplicates in reduced representation sequencing data improve genotype calling accuracy. *Molecular Ecology Resources*, 15(2), 329–336. <https://doi.org/10.1111/1755-0998.12314>

- Turcotte, F., Hanke, A., & McDermid, J. L. (2023). Atlantic bluefin tuna diet variability in the southern Gulf of St. Lawrence, Canada. *Marine Environmental Research*, 187, 105949. <https://doi.org/10.1016/j.marenvres.2023.105949>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wilson, S. G., Jonsen, I. D., Schallert, R. J., Ganong, J. E., Castleton, M. R., Spares, A. D., Boustany, A. M., Stokesbury, M. J. W., & Block, B. A. (2015). Tracking the fidelity of Atlantic bluefin tuna released in Canadian waters to the Gulf of Mexico spawning grounds. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(11), 1700–1717. <https://doi.org/10.1139/cjfas-2015-0110>
- Wright, A. F. (2005). Genetic Variation: Polymorphisms and Mutations. In Wiley, *Encyclopedia of Life Sciences* (1st ed.). Wiley. <https://doi.org/10.1038/npg.els.0005005>

Appendix

Table 3: Read quantity of sequenced Atlantic bluefin tuna (*Thunnus thynnus*) DNA samples upon initial return Génome Québec organized by year and month.

Time Period	Average # of reads	Minimum # of reads	Maximum # of reads	Standard deviation
2021	25,843,735.6	21,476,735	29,560,021	2,530,994.9
2022	24,871,889.9	17,242,281	32,975,317	4,247,822.3
2023	24,770,771.1	15,778,021	32,053,260	4,342,775.3
May	20,286,057	18,738,908	23,182,044	1,733,620.5
August	25,439,084	18,065,087	29,637,439	3,851,948.2
September	26,505,733.6	24,921,204	28,444,591	936,282
October	24,164,939.4	15,778,021	32,975,317	4,062,938.7
December	27,036,712.3	23,761,482	32,053,260	3,022,726
Total	25,162,132.2	15,778,021	32,975,317	3,802,954.9

Table 4: Duplicate percentage of sequenced Atlantic bluefin tuna (*Thunnus thynnus*) DNA samples upon initial return Génome Québec organized by year and month.

Time Period	Average Duplicate %	Minimum Duplicate %	Maximum Duplicate %	Standard deviation
2021	19.90	17.7	21.8	1.1
2022	17.90	13.7	20.4	1.80
2023	19.20	13.2	24.5	2.70
May	17.3	16.1	19.1	1.10
August	17.7	13.7	20.1	2.1
September	21.0	19.2	23.0	1.2
October	18.5	13.2	21.7	2.0
December	20.0	17.6	24.5	1.6
Total	19.00	13.2	24.5	2.20

Table 5: Adapter content of sequenced Atlantic bluefin tuna (*Thunnus thynnus*) DNA samples upon initial return Génome Québec organized by year and month.

Time Period	Average Adapter %	Minimum Adapter %	Maximum Adapter %	Standard deviation
2021	0.15	0.06	0.31	0.08
2022	0.20	0.06	0.34	0.09
2023	0.16	0.02	0.379	0.09
May	0.26	0.10	0.38	0.10

August	0.20	0.07	0.35	0.10
September	0.11	0.02	0.24	0.08
October	0.16	0.06	0.34	0.07
December	0.17	0.08	0.31	0.07
Total	0.17	0.056	0.379	0.09

Table 6: Mean read depth coverage of Atlantic bluefin tuna (*Thunnus thynnus*) DNA samples post sequencing and alignment to the reference genome.

Time Period	Mean Read Depth	Mean Read Depth SD	Reference genome covered (%)
2021	4.197543974	2.530597866	0.913280626
2022	3.928542795	2.485031214	0.899897586
2023	3.967084173	2.440796678	0.902166221
May	3.100884216	2.112869675	0.881092401
August	4.051187339	2.489221146	0.904888514
September	4.356761641	2.576638655	0.91698404
October	3.854246156	2.430712249	0.89867573
December	4.346305655	2.619421585	0.913945769
Total	4.031056981	2.485475252	0.905114811

