Developing Genomic Resources Towards Understanding the Influence of an Introgressed Inversion on Atlantic Bluefin Tuna's Recovery.

by

Abigail K. Scher

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

Bachelor of Science

Supervisor: Dr. Rebekah A. Oomen, Department of Biological Sciences

THE UNIVERSITY OF NEW BRUNSWICK

March 2025

Abstract

The influence of introgression on the recovery of depleted populations is currently an empirically understudied evolutionary process, despite its function in providing adaptive genetic variation which may favour adaptation and speciation. Atlantic bluefin tuna (*Thunnus thynnus*; hereafter, ABFT) exemplifies a unique case of introgression, where a chromosomal inversion was introduced into their population from another tuna species. Given that tuna have undergone a recent rapid demographic recovery from near extinction levels due to overexploitation, it is opportune to investigate the correlation between the inversion and ABFT's recovery. Resources were developed for the target application of investigating the correlation. Tissue, DNA and high-quality sequence banks, spanning a decade, were created and assessed temporally, using low coverage whole genome sequencing for the sequence bank. DNA content decreased as storage duration increased; however, DNA quality remained stable. Additionally, a comparative genomic approach was employed to putatively locate the inversion on chromosome 24.

Acknowledgments

I would like to thank Iris Jia and Stephanie Maheux for assistance in subsampling, extractions and useful, informative discussions. I am also thankful for Laken Devost and Laurissa Roehl for teaching DNA extraction and purification techniques. Gratitude is also given to Rebecca Krohman for her help in interpreting inversions in the beginning stages.

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature, or Abbreviations

ABFT	Atlantic bluefin tuna (*Thunnus thynnus*)
MED	Mediterranean Sea
GOM	Gulf of Mexico
SS	Slope Sea
LD	Linkage disequilibrium

Introduction

Genomic mutations can have evolutionary consequences that manifest in changes in population demographics. Mutations that increase adaptive genetic diversity can be subjected to natural selection, becoming more frequent over time. This process can be key to facilitating the recovery of nearly extinct species, occurring through mechanisms such as genetic rescue and adaptive introgression, both of which enhance the population's evolvability (Ben et al., 2020; Hamilton et al., 2016; Hedrick et al., 2010; Hedrick, 2013; Johnson et al., 2010; Whiteley et al., 2015). As a result, the viability and resilience of populations are improved. Therefore, studying how genomic mutations progress in populations or species is foundational for improving predictions of species' trajectories in changing conditions and informing effective conservation strategies.

A key class of genomic mutations are structural variants (SVs), which are large scale genomic alterations, hypothesized to be fundamental drivers of numerous evolutionary processes (Mahmoud et al., 2019; Merot et al., 2020). One of the most well studied forms of structural variation is the chromosomal inversion, a type of SV that resembles the inheritance pattern of a single locus and is hypothesized to facilitate rapid adaptation (Kirkpatrick, 2010; Oomen et al., 2020). Inversions arise from a chromosome breaking at two points, where the segment between the breakpoint is reinserted in reverse orientation (Kirkpatrick, 2010). The inverted segment will not synapse correctly during meiosis between heterozygous haplotypes, thus reducing recombination and keeping the alleles within the inverted region linked (Barth et al., 2017; Berdan et al., 2023; Huang et al., 2020; Jay et al., 2021; Kirkpatrick, 2010; Kirkpatrick et al., 2006; Kulathinal et al., 2009; Oomen et al., 2020; Wellenreuther et al., 2018).

As with other linked architectures, this linkage causes alleles within the inversion to be inherited together, resembling the inheritance of a single locus (Oomen et al., 2020). This inherent property is hypothesized to aid in rapid and local adaptation, even in the presence of high connectivity and gene flow, as rather than accumulating advantageous alleles over multiple generations, they can be inherited as a 'package' within a single generation (Barth et al., 2017; Kirkpatrick, 2010; Oomen et al., 2020).

Chromosomal inversions arise within a species, or they can be obtained through interspecific introgression, the exchange of genetic material through hybridization between closely related species, primarily mediated by backcrossing (Hamilton et al., 2016; Kulathinal et al., 2009; Twyford et al., 2012). Introgression can play a significant role in adaptive evolution, particularly when recipient species are not at optimal fitness levels (e.g., range expansion, invasion and changing environments; Edelman et al., 2021). Introgression of inversions provides an opportunity to transfer a 'package' of globally or locally adaptive alleles from one population to another (Edelman et al., 2021). Although introgression may homogenize regions of genomes, inversions have low recombination rates. Thus, they will remain distinct in hybrid offspring (Kulathinal et al., 2009).

For most organisms, the environment is constantly changing, especially with reference to climate change, which makes responses like adaptive introgression essential for species persistence (Aitken et al., 2013; Hedrick, 2013). Adaptive introgression occurs when adaptive variation is transferred from one species to another (Hedrick, 2013). Growing empirical evidence demonstrates that introgressive hybridization can release populations from their adaptive limits, accelerating adaptive responses to environmental change (Brauer et al., 2023; Hamilton et al., 2016). For example, adaptive introgression has allowed for rodenticide resistance in house mice (Hamilton et al., 2016), herbivore resistance in sunflowers (Suarez-Gonzalez et al., 2018), insecticide resistance in moths (Valencia-Montoya et al., 2020) and decreased genetic susceptibility to climate change in endemic rainbow fish (Brauer et al., 2023).

Unlike adaptive introgression, which occurs when a new environmental challenge arises (Hedrick, 2013), genetic rescue is a mechanism that increases genetic diversity to reduce the high genetic load in small populations caused by genetic drift or past inbreeding (Johnson et al., 2010; Hedrick et al, 2010; Hedrick, 2013). When new beneficial alleles are introduced from interspecific migrants, leading to a reduction of deleterious alleles, it is known as hybrid rescue (Edelman et al., 2021). Furthermore, selective maintenance of pre-gene flow local adaptive variation has been observed, such as in Trinidadian guppies, which counters the prediction that hybridization swamps pre-existing locally adapted alleles (Edelman et al., 2021; Fitzpatrick et al., 2020). However, genetic rescue only provides a short term “fix” for the recovery of endangered species, temporarily reducing the risk of extinction (Hedrick et al., 2010). Long term genetic well-being of endangered species is greatly dependent on reducing or eliminating the factors that caused endangerment, such as overfishing (Hedrick et al., 2010).

This study focuses on the Atlantic bluefin tuna (*Thunnus thynnus*; hereafter, ABFT), an economically, culturally and ecologically important species (Fromentin et al., 2014; Porch et al., 2019; Worm et al., 2011), contributing CAD $10 million to the Canadian economy annually (DFO, 2022). ABFT are a highly migratory species distributed across pelagic zones of the North Atlantic and adjacent seas, primarily the Mediterranean Sea (Fromentin et al., 2005a; Kerr et al., 2016; Mather et al., 1995). They undergo transatlantic migrations (Rooker et al., 2014; Rooker et al., 2007) related to feeding (Block et al., 2005; Chow et al., 1995; Galuardi et al 2010) and spawning across three identified spawning grounds (Díaz-Arce et al., 2024).

These spawning grounds are the Mediterranean Sea, Gulf of Mexico and the Slope Sea (hereafter, MED, GOM, and SS respectively) (Díaz-Arce et al., 2024; Richardson et al., 2016). The first two are ancestral stocks while the latter is a newer stock hypothesized to be formed by mixing of the two ancestral spawning components (Díaz-Arce et al., 2024; Richardson et al., 2016). The eastern stock (ABFTE; consisting of the MED spawning ground) is notably larger than the western stock (ABFTW; consisting of GOM and SS spawning grounds), and there are far more ABFTE caught in the west than ABFTW caught in the east (Rodríguez-Ezpeleta et al., 2019; Rooker et al., 2014), suggesting that there is an unequal proportion of migrants per stock.

ABFT fisheries have been around for millenia, but their economic value skyrocketed in the 1980s, with the development of the Japanese sushi-sashimi market, disrupting sustainability in international bluefin fisheries (Fromentin et al., 2005a; Fromentin et al., 2005b; Fromentin et al., 2014; Mather et al., 1995). With their newly gained international popularity, ABFT faced overexploitation and dramatic declines (Fromentin et al., 2005a; Fromentin et al., 2005b; Mather et al., 1995). Fortunately, they have undergone a remarkable recovery, especially the eastern stock, since 2013, aided by the implementation of the International Commission for the Conservation of Atlantic Tunas (ICCAT) recovery plan in 2008 (Fromentin et al., 2014; ICCAT, 2021, 2024). However, there is uncertainty regarding their genetic composition and without understanding these uncertainties, there is risk of oversimplifying fisheries management, compromising long-term conservation (Fromentin et al., 2014).

Recently, an albacore (*Thunnus alalunga*) introgressed chromosomal inversion has been found in all three ABFT spawning grounds at varying frequencies (Díaz-Arce et al., 2024), with the highest in the MED, followed by the SS then the GOM. These species are known to hybridize, with mitochondrial introgression being previously reported in the MED and SS (Bremer et al., 2005; Díaz-Arce et al., 2024; Viñas et al., 2009). Now, introgressed mtDNA has been found in the GOM, along with the first evidence of nuclear introgression in all spawning grounds (Díaz-Arce et al., 2024). The higher frequency in the MED, where the spawning location and timing of these tunas overlap (Albaina et al., 2013; Alemany et al., 2010; Huang et al., 2020; Pujolar et al., 2003; Reglero et al., 2017), suggests that the introgression event occurred here, or has been more intense in the MED prior to diffusing to SS and GOM by gene flow (Díaz-Arce et al., 2024). Notably, the differentiation of allele frequencies in this inversion is stronger than the observed genome-wide differentiation between the GOM and the MED spawning components, suggesting that the inversion may be contributing to ABFT's adaptation capabilities (Díaz-Arce et al., 2024) and thus their species recovery.

ABFT and albacore are closely related, sharing similar morphology (Davies et al., 2011) and both exhibiting highly migratory behaviours (Viñas et al., 2009). While both species are temperate (Davies et al., 2011), they have different temperature preferences. Albacore prefer warmer temperatures (Alemany et al., 2010; Arrizabalaga et al., 2015; Chen et al., 2005; Fromentin et al., 2005a; Muhling et al., 2011), With their preferred spawning temperature exceeding 27ºC, while ABFT prefers a range of 21.5-26.5ºC in the

MED (Alemany et al., 2010) and 24-27ºC in the GOM (Muhling et al., 2011). ABFT spawning conditions and larval occurrences have been found to be correlated with temperature, making them likely vulnerable to impacts from climate change (Muhling et al., 2011). If the inversion in question is associated with thermal tolerance, introgression could have provided ABFT with adaptive variation to better cope with changing climates. Alternatively, and/or additionally, the introgression event could have contributed to ABFT's recovery through hybrid rescue by enhancing genetic diversity. In this case, the reduction of genetic load may have allowed ABFT to increase in population size initially, while effective management further facilitated their recovery.

The region of linkage disequilibrium (LD) associated with the inversion was found to be 2.63 Mb long (Díaz-Arce et al., 2024), however the location of the inversion within the genome has yet to be determined. Knowing where an inversion is in the genome is extremely important as it allows elucidation of the genes involved, providing insight to the inversion's function and thereby evolutionary significance. With the recently available annotated reference genome for, determining the location of the inversion will allow for association with specific genes. This can aid in further investigations regarding the inversion's influence on recovery and contribute to species trajectories.

The overarching aim of this study is to develop genomic resources towards the target application of investigating the temporal correlation between the albacore introgressed inversion and Atlantic bluefin tuna species recovery. A correlation suggest that the inversion is under selection, potentially due to increased genetic diversity, which may be adaptive. Prior to beginning this investigation, there is need for a database of genome sequences starting with samples of ABFT individuals from 2013 (the beginning of their recovery; ICCAT, 2021, 2024). However, time-dependent sampling for DNA analysis, particularly of muscle tissue subject to cold storage, has been associated with DNA degradation linearly over time, though studies on this topic- the effects of long-term freezing- are limited (Soniat et al., 2021). Further, with the inherent improvement of tissue collection and preservation techniques for DNA analysis over time, it can be inferred that these advancements may have contributed to greater DNA integrity in more recent samples.

My specific objectives are threefold: (1) create a tissue and DNA bank from archived ABFT muscle tissue spanning a decade and temporally evaluate the extraction yield as a proxy for DNA degradation, under the hypothesis that within stored muscle tissue DNA progressively degrades with time; (2) create a whole genome sequence database from the samples in the tissue bank; (3) identify the putative location of the albacore-introgressed inversion within the genome by comparing publicly available Mediterranean and Gulf of Mexico whole genomes.

Methods

Tissue and DNA bank

Tissue sample collection

Samples of ABFT muscle tissue collected from head steaks were subsampled from the Department of Fisheries and Oceans Canada (DFO) tuna archive at the Saint Andrews Biological Station (SABS) under a set of criteria. The archive was filtered temporally and spatially for two reasons: (1) to capture variation among years while ensuring the same birth cohort was not sampled in different years of capture and (2) to ensure ABFT from all spawning regions were present. Subsamples were taken five years apart, starting in 2013, the beginning of their recovery, within a three birth year range per year captured to ensure the same cohort is not sampled in different years of capture (Table 1). The geographical area was filtered to only include those caught in the Gulf of Saint Lawrence (GSL) to minimize spatial variation. This included landing ports in Port Hood, Murray Harbour and Ballantynes cove, collectively grouped as Location 1; North Lake and Morell, grouped as Location 2, and Tignish classified as Location 3 (Table 1). Only those caught between August and October were selected, as during these months, ABFT from all spawning locations are expected to be present in the GSL (Dedman et al., 2023).

DFO scientists used otoliths to age individuals up to 2018. For 2023, age was estimated using a growth model. To decide the most effective method to estimate the ages, two growth models, a von Bertalanffy (VB) growth model (Stewart et al., 2022) and ICCAT's currently accepted Richards growth curve (Ailloud et al., 2017) were compared. Both models were compared to observed data from individuals caught in 2013 and 2018 with known ages. The raw residuals representing the difference between the observed data and the expected data according to each growth model was then compared. The typical VB model was rearranged to estimate age from straight fork length (equation 1). The model was run using parameters derived by Stewart et al. (2022); the values of these parameters are 0.119, -0.47, and 300.4 for K, $t0$ and L∞, respectively (equation 1). The Richards growth curve is already in the format to estimate age from size, so no adjustments were made, and the model was run using parameters derived from Ailloud et al. (2017). Both growth models were run in R (version 4.4.2) using RStudio (version 2024.12.0 build 467).

The VB model was found to produce a more favorable mean raw residual value than the Richards growth curve for archived individuals caught in 2013 and 2018 (-1.04 ± 3.61 compared to -3.32 ± 5.68) (Figure A1). Therefore, ages for individuals caught in 2023 were estimated using the VB model. However, the model required straight fork lengths, which were not available for 2023. Prior to running the model, straight fork lengths were estimated from curved fork lengths using ABFTW ICCAT conversion factors (Rodriguez-Marin et al., 2015) with code from Stewart et al (2022). Once all the ages were calculated for 2023 individuals, a subsample of 5-6 individuals per birth year, i.e. 15-18 per sampled year, were randomly selected for extraction and subsequent sequencing (Table 1).

$$t = \frac{1}{-k^{(i)}} \ln\left(1 - \frac{L_i(t) - \varepsilon_i}{L_\infty^{(i)}}\right) + t_0$$

Equation 1: Rearranged von Bertalanffy growth model to estimate age of an individual. K represents the Brody growth coefficient, L∞ is the asymptotic length and *t*0 is the theoretical age at length zero (von Bertalanffy, 1938). Li(*t*) is the straight fork length at age *t* for the *i*th individual and *t* is the calculated age of the individual. εi represents the error associated with the *i*th observation and was not included in age estimations as it is unknown.

Table 1. Atlantic bluefin tuna (*Thunnus thynnus*) subsampled from SABS tuna archive. Birth year was determined by the age of the individual compared to the year the tuna was caught. Ages for individuals in 2013 and 2018 were determined by otoliths; ages for individuals in 2023 were estimated using a von Bertalanffy growth model and straight fork lengths converted from curved fork lengths. Location columns represent how many Atlantic bluefin tuna were sampled from that location; Location 1 includes Port Hood, Murray Harbour and Ballantynes cove; Location 2 includes North Lake and Morell; Location 3 represents Tignish.

Year Caught	Birth Year	Location 1	Location 2	Location 3
2013	2001	6		
	2002	3	3	
	2003	4	1	1
2018	2006	1	5	
	2007	1	2	3
	2008	1	2	3
2023	2011		6	
	2012		6	
	2013		5	

DNA extraction and purification

DNA was extracted and purified from the selected samples (n= 53) using the BIO-TEK Omega E.Z.N.A® Tissue DNA Kit according to the manufacturer's instructions with 25-30mg of muscle tissue (with the exception of three samples that were slightly under 25mg) and 50 µL of elution buffer (25 µL x 2). DNA extracts were analyzed using a NanoDrop spectrophotometer (Thermo Scientific) to obtain nucleotide concentrations. If the concentration was less than 20 (ng/µl), the sample was not used and was randomly replaced by another sample that matched the characteristics of the unused sample (i.e, same birth year and location).

Data analysis

To evaluate DNA degradation over time, the extraction yield (mass of nucleic acids in ng) was determined for each sample by multiplying the volume of the yield (50 µl) by the concentration of nucleic acids (ng/ µl). An ANOVA was used to test for a relationship between extraction yield (ng) and the year individuals were caught (extraction yield ~ year). A Tukey's HSD test was then run to see the breakdown of specifically which years had significantly different extraction yields from one another. Because there was three samples extracted from less than 25 mg of muscle tissue, an ANCOVA was run to assess if the amount of tissue (<25 or between 25-30 mg) influenced the extraction yield adjusting for year caught and the interaction between year caught and muscle tissue mass (extraction yield ~ year * mass of tissue). The interaction term was included due to the possibility that older samples may be more sensitive to muscle tissue mass imputed. Summary statistics (mean, standard deviation, minimum and maximum) were also calculated for extraction yield per year and in total across years. Significance was determined if the P-value was lower than the threshold of $\alpha = 0.05$.

Sequence Bank

DNA sequencing

The extracted DNA (40 ul of the extraction yield) was sent overnight on dry ice to the Génome Québec sequencing centre for low coverage whole genome sequencing (lcWGS) with a target average depth of 4x (Figure 1). All samples were normalized to 100 ng of genomic DNA during shotgun library preparation at the sequencing centre. LcWGS was most suitable for this study as distributing a set quantity of sequencing effort over a larger number of samples, each with lower sequencing depth, enhances the accuracy of population-level inferences (Lou et al., 2021). Sequencing was performed using the Illumina NovaSeq Paired End 150 protocol with standard Illumina dual-index adapters, which produces reads 150 bp in length. It is important to note here that 'low coverage' (in LcWGS) refers to low read depths per sample. This differs from the term 'coverage' or 'reference coverage' used throughout the rest of this paper, representing the proportion of the reference genome covered by the reads produced for an individual (Figure 1).

Processing and alignment

Sequence data were delivered on Nanuq, the server used by Génome Québec. Sequences were then uploaded to the Siku server of the ACENET supercomputer (Digital Research Alliance of Canada), along with the sequencing adapter sequences used and the chromosome-level reference genome for *Thunnus thynnus* (NCBI, 2024a). To align sequences with the reference genome, adapter sequences were first trimmed off the raw reads using Trimmomatic (Bolger et al., 2014) to produce trimmed reads, followed by alignment of the trimmed reads to the indexed reference genome. Reads associated with a mapping quality of less than 20 were filtered out using bowtie2 (Langmead et al., 2012) (Physalia Courses, 2024). PCR duplicates were then identified and removed by trimming overlapping regions of reads using Picard Tools MarkDuplicates (Broad Institute, n.d.) and BamUtil clipOverlap (Jun et al., n.d.,) (Physalia Courses, 2024). Read depth was estimated per individual using the module samtools depth (Li et al., 2009). Summary statistics of read depth per individual were computed in RStudio (Physalia Courses, 2024).

Data analysis

The numbers of raw reads produced were assessed temporally by an ANOVA (number of raw reads ~ year) on RStudio. Summary statistics (mean, standard deviation, minimum and maximum) were produced for number of reads, raw and trimmed, as well as read depth and coverage values (Figure 1). Read depth values that summary statistics were produced for were the mean (including and excluding zero read sites), median and standard deviation of read depth across sites in the genome per individual. Summary statistics were computed per year and in total across all years.

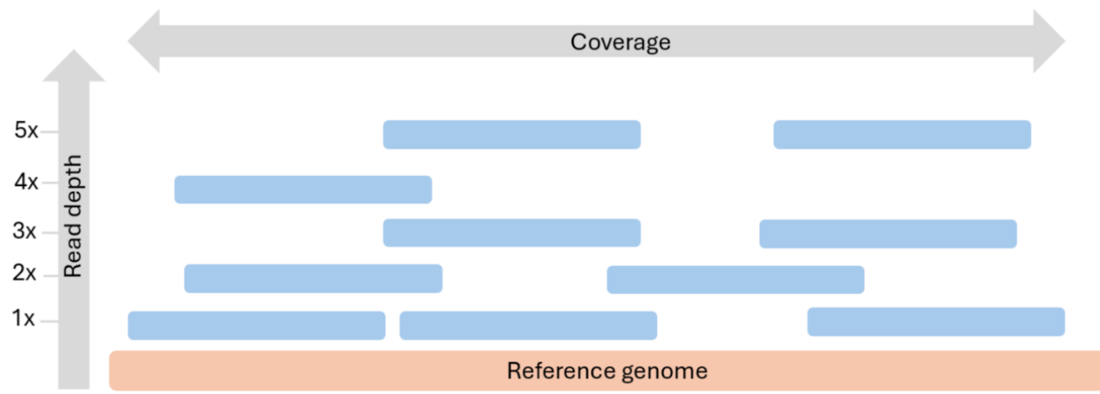

Figure 1. Visual representation of read depth and coverage. The blue boxes represent fragments of one individual's genome (i.e. reads) being aligned against the reference genome (orange).

Inversion location

The location of the inversion was investigated by creating a similarity matrix (dot plot) on D-GENIES (Cabanettes et al., 2018) comparing two ABFT genomes: the reference, fThuThy2.1 (a Mediterranean individual, target; NCBI, 2024a) and ASM4338166v1 (a Gulf of Mexico individual, query; NCBI, 2024b).

After generating the dot plot, the contigs were sorted to align the query sequences in the same order as the target because the chromosomes were numbered differently between the two genomes. Noise on the dotplot (very small similarity signals that occur frequently but are not contributing to the analysis) was hidden. Each matched chromosome was examined separately in search of the inversion, which is visible in a similarity matrix as a segment that is perpendicular to the similarity signal observed for the collinear parts of the genome. Lengths of inversions were measured in pixels from the dot plot using ImageJ and calibrated to the scale of the matrix axes.

Once an inversion with comparable size to the region of linkage disequilibrium found in Diaz-Arce et al. (2024) was located, an inversion second closest in size to the region of LD was identified to confirm the most accurate match. Lastly, for clarity in visualization, accession numbers for each chromosome were converted to their corresponding chromosome number for each genome using Adobe Illustrator.

Results

Tissue, DNA, and sequence banks

The tissue bank consists of muscle tissue from 53 *Thunnus thynnus* individuals: 18 from 2013, 18 from 2018, and 17 from 2023 (Table 2). The samples are currently stored in 95% ethanol in a standard -20ºC freezer. There are duplicates for 2023 but only single samples for the other two years.

The DNA bank consists of nucleic acids extracted from the same samples. Each sample has a volume of 9 µl (50 ul initial volume minus 40µl sent for sequencing and 1µl used to NanoDrop) and are stored in a standard -20ºC freezer (Table 2). In total, the mean extraction yield (respective to the volume currently in the DNA bank) is 1002.68 ng with a standard deviation of 531.65 ng. The minimum yield is 198.9 ng produced by a 2023 sample and the maximum is 2315.7 ng also produced by a 2023 sample (Table 3). Samples from 2013 have the lowest mean yield, followed by 2018. Samples from 2023 have the highest variation, followed by 2013 (Table 3).

Table 2. Characteristics of samples in the tissue and DNA bank currently stored at -20°C. Concentration, volume, and yield refer only to the DNA bank. Concentration was determined using a NanoDrop spectrophotometer. All birth years for 2023 were estimated using a von Bertalanffy growth model with parameters derived from Stewart et al (2022). Birth years for 2013 and 2018 were determined using otoliths by DFO scientists.

Sample ID	Year caught	Birth year	Month caught	Port landed	Conc (ng/ul)	Volume (ul)	Yield (ng)
3857	2013	2003	10	Port Hood	29.9	9	269.1
3865	2013	2002	9	Port Hood	87.6	9	788.4
3869	2013	2003	10	Morell	91	9	819
3878	2013	2001	10	Port Hood	46.3	9	416.7
3881	2013	2001	10	Port Hood	24.7	9	222.3
3887	2013	2003	10	Port Hood	56.7	9	510.3
3897	2013	2002	10	Port Hood	34.4	9	309.6
3924	2013	2002	10	Port Hood	79.9	9	719.1
3926	2013	2001	10	Port Hood	206.3	9	1856.7
3930	2013	2001	9	Port Hood	72.3	9	650.7
3933	2013	2001	9	Morell	54.3	9	488.7
3956	2013	2003	9	Morell	52.5	9	472.5
3966	2013	2003	9	Morell	49.1	9	441.9
4023	2013	2002	10	Tignish	158.8	9	1429.2
4038	2013	2001	9	Port Hood	44.3	9	398.7
4054	2013	2002	10	Port Hood	52.5	9	472.5
4056	2013	2002	10	Morell	222.5	9	2002.5
4070	2013	2003	10	Port Hood	158.6	9	1427.4
6549	2018	2006	8	North Lake	119.4	9	1074.6
6580	2018	2007	8	Ballantynes Cove	105.9	9	953.1
6588	2018	2007	9	North Lake	146.5	9	1318.5
6608	2018	2008	8	Morell	139.9	9	1259.1
6612	2018	2006	8	North Lake	148.5	9	1336.5
6619	2018	2006	9	North Lake	126	9	1134
6632	2018	2008	9	Tignish	80.2	9	721.8
6635	2018	2006	9	Port Hood	164.5	9	1480.5
6639	2018	2006	9	North Lake	152.5	9	1372.5
6644	2018	2007	9	Tignish	73.8	9	664.2
6655	2018	2008	9	North Lake	37.3	9	335.7
6685	2018	2007	10	Tignish	160.5	9	1444.5

6722	2018	2007	9	Tignish	46.1	9	414.9
6734	2018	2008	9	Tignish	105.3	9	947.7
6925	2018	2006	10	North Lake	139.8	9	1258.2
6936	2018	2008	10	North Lake	105.8	9	952.2
6960	2018	2008	10	Murray Harbour	66.8	9	601.2
6968	2018	2007	10	Tignish	157.3	9	1415.7
11276	2023	2011	8	North Lake	170.4	9	1533.6
11279	2023	2011	8	North Lake	182.8	9	1645.2
11284	2023	2011	10	North Lake	181.3	9	1631.7
11290	2023	2011	8	North Lake	257.3	9	2315.7
11306	2023	2011	9	North Lake	90.9	9	818.1
11314	2023	2011	9	North Lake	90.2	9	811.8
11316	2023	2012	9	North Lake	207.4	9	1866.6
11319	2023	2012	9	North Lake	133.9	9	1205.1
11320	2023	2012	8	North Lake	165	9	1485
11326	2023	2012	8	North Lake	153.4	9	1380.6
11327	2023	2012	9	North Lake	220.5	9	1984.5
11330	2023	2012	9	North Lake	129.4	9	1164.6
11339	2023	2013	9	North Lake	69.2	9	622.8
11353	2023	2013	9	North Lake	122.3	9	1100.7
11363	2023	2013	8	North Lake	22.1	9	198.9
11549	2023	2013	10	North Lake	63	9	567
11341	2023	2013	8	North Lake	47.8	9	430.2

Table 3. Summary statistics of the yield (mass of extracted nucleic acids in ng) of all samples in the DNA bank. The yield was determined by multiplying the current volume of the samples (9ul) by their concentration.

	Yield (ng)			
Year	Mean	SD	Min	Max
2013	760.85	544.5728561	222.3	2002.5
2018	1038.05	361.8590126	335.7	1480.5
2023	1221.3	588.5839161	198.9	2315.7
Grouped	1002.684906	531.6542169	198.9	2315.7

The extraction yield significantly decreased as storage duration increased ($p = 0.0322$, df =2. ANOVA), with the decrease in yield driven by the difference in yield between 2013 and 2023 (p = 0.0258, Tukey's HSD) (Figure 2). The three samples with under 25mg of tissue (sample IDs 3966, 4038, and 6722) did not significantly affect extraction yield when adjusting for year (p = 0.6676, df=1, ANCOVA) or overall (p = 0.1225, df=1, ANCOVA). The year individuals were caught did not significantly affect the number of raw reads produced (p = 0.539, df=2, ANOVA) from the normalized extraction yields (Figure 2).

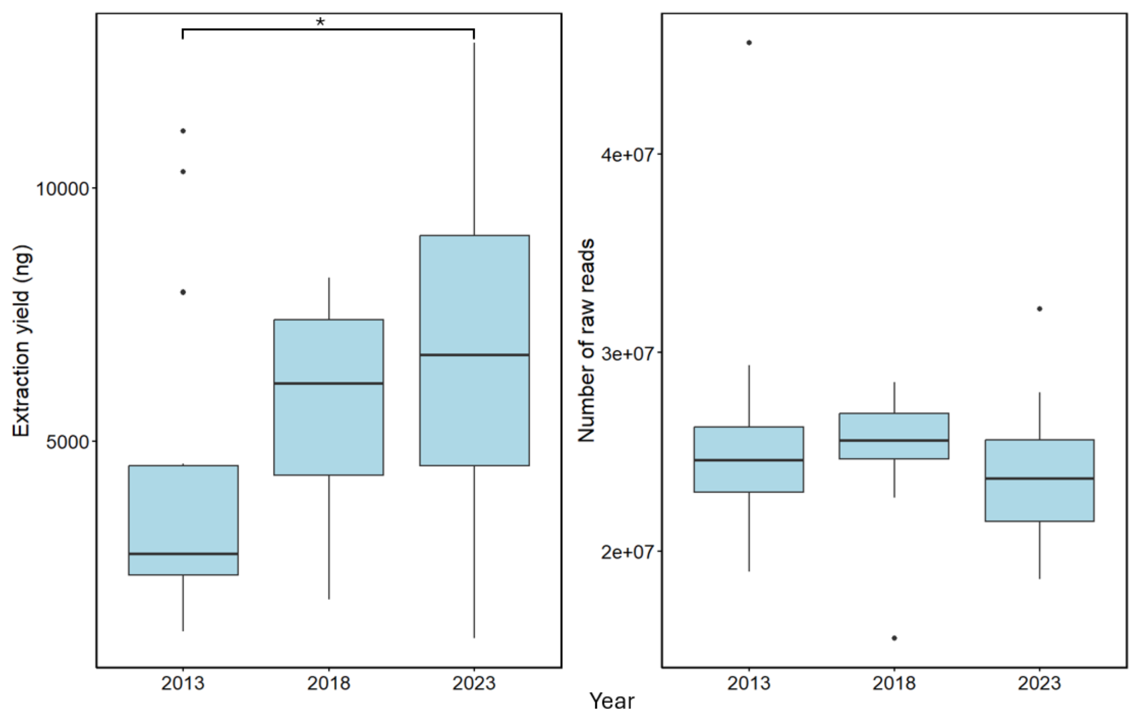

Figure 2. The extraction yield (left) and number of raw reads produced from the normalized extraction yields (right) per the year *Thunnus thynnus* muscle tissue was sampled and subjected to -20ºC cold storage (either in ethanol or frozen as a steak) until extraction in 2024.The asterisk indicates a significant difference (p = 0.0258; α = 0.05).

The sequence bank is an online dataset of all the samples mapped to the reference genome (fThuThy2.1; NCBI, 2024a). All genome sequences were of high quality with quality scores across all bases being between Phred scores of 35 and 40. In total, the mean raw reads (i.e. before adapter clipping) had 18459 more reads than the trimmed read set. The samples were sent for lcWGS with a target average depth of 4x, which is verified with the mean depth across all individuals being 3.94x and the mean excluding zero read sites being 4.35x (Table 4; Table A1). The highest mean across individuals per year of each individual's median and mean (with and without accounting for zero read sites) read depth across sites in the genome, was produced in 2013, followed by 2018. However, the highest variation in read depth was also produced in 2013, followed by 2018 (Table 4).

The highest mean coverage was produced in 2013, followed by 2018. The highest variation in coverage was produced in 2018, followed by 2023 (Table 4). Thus, the increasing trend, as storage duration increases, is seen in mean coverage values as well as mean read depth values.

The average proportion of the reference genome covered each year is over 0.90 with the lowest coverage of 0.849 produced in 2018 (Table 4). All years have a standard deviation of less than 0.02 for proportion of the reference genome covered.

Table 4. Summary statistics across all individuals per year and in total across years ('Grouped') for: (1) number of raw reads, before adapter clipping per individual; (2) number of trimmed reads per individual; (3) mean read depth across sites in the genome per individual; (4) mean read depth excluding zero-read sites per individual; (5) standard deviation of read depth across sites in the genome per individual; (6) Median of read depth across sites in the genome per individual; (7) the proportion of the reference genome covered (coverage).

		Numbers of reads		Read depth values				
Year	Statistic	Raw	Trimmed	Mean	Mean non-zero	SD	Median	Coverage (proportion)
2013	Mean	25350459	25333532	4.09	4.5	2.52	4	0.906
	SD	5687585	5683081	0.894	0.915	0.349	0.907	0.0135
	Min	18948852	18935154	3.03	3.44	2.07	3	0.88
	Max	45580344	45546465	7.14	7.65	3.68	7	0.933
2018	Mean	25257109	25233426	3.89	4.3	2.45	3.78	0.904
	SD	2902025	2898391	0.516	0.509	0.211	0.647	0.0168
	Min	15612039	15601150	2.43	3.86	1.79	2	0.849
	Max	28484287	28447434	4.76	5.17	2.78	5	0.921
2023	Mean	23937731	23923180	3.85	4.25	2.44	3.71	0.903
	SD	3271799	3269519	0.508	0.501	0.199	0.686	0.0139
	Min	18600745	18585356	2.95	3.37	2.05	3	0.875
	Max	32192649	32174536	4.82	5.23	2.85	5	0.922
Grouped	Mean	24865616	24847157	3.94	4.35	2.47	3.83	0.905
	SD	4127762	4123947	0.663	0.669	0.261	0.753	0.0146
	Min	15612039	15601150	2.43	4.86	1.79	2	0.849
	Max	45580344	45546465	7.14	7.65	3.68	7	0.933

Inversion location

The putative location of the albacore-introgressed inversion was found to be on chromosome 24 for both the MED and GOM genomes (accession number NC_089540.1 and CM092385.1 for fThuThy2.1 and ASM4338166v1, respectively) (Figure 3). The size of this inversion is approximately 3.78 Mb long, which is the largest inversion found in the genome and 1.15 Mb longer than the region of linkage disequilibrium identified by Diaz-Arce et al. (2024). The inversion with the second closest size compared to the region of linkage disequilibrium identified by Diaz-Arce et al. (2024) was found on chromosome 15 for both genomes. It spans approximately 0.81 Mb, which is 1.82 Mb smaller than the region of linkage disequilibrium identified by Diaz-Arce et al. (2024). Given that the inversion on chromosome 24 is both the closest in size and the largest detected inversion, it is the most likely candidate for the albacore-introgressed inversion.

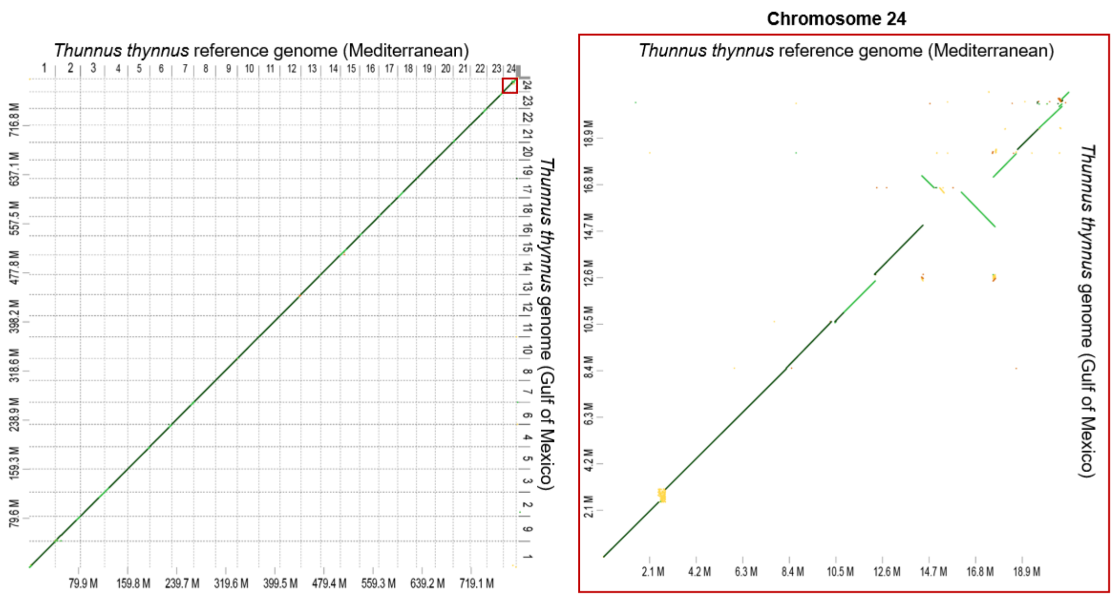

Figure 3. Similarity matrix of *Thunnus thynnus* reference genome (fThuThy2.1, target; NCBI, 2024a) a Mediterranean individual, and another publicly available genome (ASM4338166v1, query; NCBI 2024b), a Gulf of Mexico individual. All accession numbers were changed to their associated chromosome number using Adobe Illustrator.

Discussion

The purpose of this study is to develop genomic resources towards the target application of investigating if a correlation exists between the recently discovered albacore-introgressed inversion and ABFT's species recovery. Tissue, DNA, and high-quality whole genome sequence banks were created, accompanied by the identification of the most likely genomic location of the inversion.

Tissue and DNA bank

Tissue and DNA banks of ABFT individuals spanning a decade were successfully created for the target application. The extraction yield, which was used as a proxy for DNA content, significantly decreased as storage duration increased which is consistent with the findings of Soniat et al (2021). The decreasing trend was driven by the difference of extraction yield between 2013 and 2023. The difference in mean extraction yield between 2013 to 2018 and 2018 to 2023 are within ± 100 ng of each other, whereas the difference between 2013 and 2023 is the sum of those two differences and therefore much larger. Therefore, it is possible that subjecting muscle tissue, either in ethanol or as a steak, to cold storage (-20ºC) for 5 years does not have a significant effect, but for 10 years it does.

One potential limitation of this study is that most of the 2013 and 2018 samples were preserved in ethanol whereas most of the 2023 samples were kept as head steaks, although both were subjected to cold storage. This raises the question of whether the differences in preservation method across years is responsible for the observed differences in extraction yields, rather than storage duration. However, If the storage method was responsible for the trend observed, the opposite trend, DNA content increasing as storage duration increases, would be expected. This is because ethanol is a better preservation technique than typical freezing as it kills microorganisms that decompose tissues, slow down enzymatic processes by removing water, and denatures DNA-degrading enzymes (Dahn et al., 2022; Marquina et al., 2021).

Another potential limitation to this tissue bank is that it is relatively small, with only 53 individuals. However, the precise selection of individuals resulted in an experimental design that is sufficient for the target application. Further, individuals were caught at different ports within the GSL. Although relatively close spatially, it is possible that there are ecological distinctions, (e.g., temperature) among ports. Because ports are not equally represented across years, this could skew results. For example, if the inversion is associated with thermal tolerance, some individuals may be able to reach the more northern ports over other individuals. This could hinder the ability to detect the true temporal trend for the target application. Therefore, it is recommended that potential port bias is taken into account when making inferences using these data.

Sequence bank

The raw reads did not have any significant temporal trend. This is because the DNA input was successfully normalized to 100ng during library preparation prior to sequencing (Randrian, Z, Génome Québec, personal communication, 2025). Normalization is the process where each individual sample is diluted to the same concentration. Thus, the amount of DNA that is sequenced per sample is equal. Therefore, the decreasing trend of DNA yield as storage duration increased did not affect the sequencing results. This finding demonstrates that the samples dating back to 2013 are suitable for genomic studies, as are likely the approximately 14 thousand individuals archived from 2013 to the present in the DFO tissue archive. This opens opportunities for many population genomic studies using the DFO archive.

Next generation sequencing technologies are effective methods for gaining insights to the evolutionary processes acting on genomes. However, the way these technologies are applied influences the genetic inferences that can be made. Here, it was decided to optimize sample size while lowering read depth (4X) rather than maximizing read depth at the expense of a lower sample size.

The whole genomes produced for all individuals are of high quality, with a high level of confidence in accuracy. For all individuals, the probability of a base being called incorrectly ranges between 0.01% - 0.0316%, based on Phred scores (Liao et al., 2017). Furthermore, 84.9% to 93.3% of the reference genome is covered across all individuals (mean = 90.5%). This is very high reference genome coverage relative to other studies, e.g., at 4x read depth 68.52% of the reference genome is covered in *Solanum melongena* (Baraja-Fonseca et al., 2024) at 10x read depth, an average of 74.1% of the reference genome in *Sistrurus catenatus* is covered (Mathur et al., 2023).

The genomes sequenced in this study are of high quality and confidence, produced by spreading an equal amount of low depth sequencing effort across samples and genomes. The sequencing strategy used herein is therefore expected to produce more reliable estimates of the true population(s). These findings suggest that this sequencing strategy could also be used to sequence a much larger sample size from the DFO tissue bank for future population genomic inferences. Variability in read depth and coverage differed slightly across years. It is recommended to consider using larger sample sizes in highly variable years in order to detect true differences between groups due to this enhanced biological variation.

A lower read depth is associated with higher genotype uncertainties; however, a larger sample size is more representative of the genetic variation within the population (Alex Buerkle et al., 2014; Fumagalli et al., 2013; Lou et al., 2021; Sims et al., 2014). Since the target application of this study is to create inferences about the genetic variation within the population, prioritizing sample size improves accuracy and ensures a more representative dataset, without completely compromising sequencing accuracy (Alex Buerkle et al., 2014; Fumagalli et al., 2013; Lou et al., 2021; Sims et al., 2014). Furthermore, a read depth of 4X is sufficient to obtain precise estimates of LD from the dataset (Lou et al., 2021). Therefore, lowering the read depth to prioritize sample size in this study is not expected to compromise the ability to identify the chromosomal inversion. This sequencing strategy is ideal for analyses of allele differentiation and frequency, population structure, such as principal component and admixture analyses, and relative estimation of LD, such as characterization of LD blocks (Lou et al., 2021).

In addition to the limitations described for the tissue bank, the sequenced samples face the limitation of lower base call accuracy than they could have if read depth was prioritized over sample size. Although the probability of a base being called incorrectly is low already, it tends to decrease with a higher read depth (Lou et al., 2021). Therefore, this database and sequencing technique would not be the most appropriate to use for studies that need absolute estimations of rare allele dependent metrics and LD, such as LD decay rates (Lou et al., 2021).

Putative inversion location

The most likely location of the albacore-introgressed inversion was identified to be on chromosome 24, as it was the inversion detected that was closest in length to, yet slightly larger than, the region of LD identified by Diaz-Arce et al (2024). The region of LD, which was detected using RAD-seq, is likely going to be smaller than the inversion itself because LD is determined by the statistical association of a subset of single nucleotide polymorphisms (SNPs) and SNPs are much more sparse in a RAD-seq data set compared to a lcWGS data set. However, how the region of LD fits into the genome is unknown. There are factors other than inversion length that might contribute to a smaller region of LD, such as gene conversion, which is when a region within a sequence is replaced by a copy from a homologous or paralogous sequence (Chen et al., 2007; Fuller et al., 2020). However, there is currently not enough known about this inversion to make any inferences about such mechanistic processes. The precise genomic location of the inversion as well as the loci included and their associations would improve inferences about these processes.

This was a preliminary investigation of the location of the albacore-introgressed inversion identified in previous studies and the findings are circumstantial. Without mapping the previous RAD-seq data to the new reference genome(s), it is not possible to confirm that the inversion detected in this study is in the same location. It is recommended that the inversion is mapped to the reference genome in future studies, including using existing publicly available data from previous studies, to confirm its location across all studies. Identifying the location of the inversion in the genome can provide valuable insights towards its function and evolutionary significance in the recovery of ABFT, as it would allow for elucidation of the associated genes and their evolutionary history. Further, knowing the precise location allows for targeted sequencing of the inversion in future studies, which enables studying the inversion at scale without the high cost of whole genome sequencing strategies (e.g., Hoff et al., 2019).

Future research directions

From the aligned genomes, the next steps in order to investigate the correlation of the inversion with ABFT's recovery are to determine the inversion haplotype frequency at each time point (2013, 2018, 2023). This can be achieved through: (1) callings SNPs, (2) calculating fixation index values for each SNP site, (3) performing a Manhattan plot analysis to locate the outlier SNPs that are characteristic of an inversion, (4) performing a PCA with the genotypes associated with the extracted outlier SNPs, and (5) performing a continuous factorial ANOVA to test for an association between inversion haplotype frequency and time point.

Furthermore, the inverted haplotype frequency can also be modelled as a function of year and stock to investigate if there is a certain stock driving the potential temporal trend. Stock assignment is needed for this, which can be determined with high accuracy by stable isotopes in the otolith (Rooker et al., 2014). All ABFT samples within the archive at DFO have an otolith associated with them, therefore this data could be obtained. Alternatively, stock assignment is also theoretically possible for all three spawning sites using genomic data like that produced in this study (Lou et al. 2021), which is another potential application of the sequence database. This can help reveal the ecological significance of the inversion, as the spawning sites differ in ecological factors (e.g., temperature). Mapping of the inversion and identifying the associated genes would greatly expand this potential site-specific application as it would allow more concrete associations between inversion prevalence and other factors, such as local adaptations and migratory behaviours.

Conclusion

The target application of investigating the correlation between the albacore-introgressed inversion and ABFT's species recovery will inform effective conservation strategies by improving our understanding of evolutionary trajectories in this multi-million-dollar fishery (DFO, 2022). It will also increase the understanding of the influence of introgression in population recovery in general. Considering that this introgression event occurred with an inversion, it provides a unique opportunity to understand how a 'package' of potentially advantageous alleles that are originally selected for the niche of one species can function in the recovery of a different species.

Introgression can aid in recovery by releasing populations from their adaptive limits and/or by decreasing genetic load (Brauer et al., 2023; Hamilton et al., 2016; Hedrick et al, 2010, 2013; Johnson et al., 2010). Inversions can additionally facilitate recovery by suppressing recombination, enabling advantageous alleles to be inherited together rather than accumulating them over multiple generations (Barth et al., 2017; Kirkpatrick, 2010; Oomen et al., 2020). Together, these two factors can serve as powerful evolutionary drivers for the recovery of depleted populations, highlighting the importance of further research using the genomic resources developed in the present study.

Contributions to Thesis

Rebekah A. Oomen and Abigail K. Scher conceptualized the project. Dheeraj S. Busawon collected samples that are in the DFO archive. Abigail K. Scher subsampled muscle tissue using code provided by Nathan D. Stewart. Alex Hanke provided funding from the months of May to August. Nathalie LeBlanc processed and aligned the sequences to the reference genome. Rebekah A. Oomen provided mentorship throughout.

Bibliography

Aitken, S. N., Whitlock, M. C. (2013). Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annual Review of Ecology, Evolution, and Systematics, 44, 367–388.

Albaina, A., Iriondo, M., Velado, I., Laconcha, U., Zarraonaindia, I., Arrizabalaga, H., Pardo, M. A., Lutcavage, M., Grant, W. S., Estonba, A. (2013). Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure. Animal Genetics, 44(6), 678–692.

Alemany, F., Quintanilla, L., Velez-Belchí, P., García, A., Cortés, D., Rodríguez, J. M., Fernández de Puelles, M. L., González-Pola, C., López-Jurado, J. L., CLIOTOP International Symposium La Paz Mexico 20071203-20071207. (2010). Characterization of the spawning habitat of Atlantic bluefin tuna and related species in the Balearic Sea (western Mediterranean). Progress in Oceanography, 86(1-2), 21–38.

Alex Buerkle, C., Gompert, Z. (2013). Population genomics based on low coverage sequencing: how low should we go? Molecular Ecology, 22(11), 3028–3035.

Arrizabalaga, H., Dufour, F., Kell, L., Merino, G., Ibaibarriaga, L., Chust, G., Irigoien, X., Santiago, J., Murua, H., Fraile, I., Chifflet, M., Goikoetxea, N., Sagarminaga, Y., Aumont, O., Bopp, L., Herrera, M., Marc Fromentin, J., Bonhomeau, S. (2015). Global habitat preferences of commercially valuable tuna. Deep-Sea Research Part II, 113, 102–112.

Ailloud, L.E., Lauretta, M.V., Hanke, A.R., Golet, W.J., Allman, R.J., Siskey, et al. (2017). Improving growth estimates for Western Atlantic bluefin tuna using an integrated modeling approach. Fish. Res. 191, 17–24.

Baraja-Fonseca, V., Arrones, A., Vilanova, S., Plazas, M., Prohens, J., Bombarely, A., Gramazio, P. (2024). Benchmarking of low coverage sequencing workflows for precision genotyping in eggplant. bioRxiv.

Barth, J. M. I., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., Jakobsen, K. S., Johannesson, K., Jorde, P. E., Knutsen, H., Moksnes, P.-O., Star, B., Stenseth, N. C., Svedäng, H., Jentoft, S., André, C. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. Molecular Ecology, 26(17), 4452–4466.

Ben, J. N., Devaughn, F., Thomas, H. M. (2020). Transforming ocean conservation: applying the genetic rescue toolkit, 11(2), 209–209. https://doi.org/10.3390/genes11020209

Berdan, E. L., Barton, N. H., Butlin, R., Charlesworth, B., Faria, R., Fragata, I., Gilbert, K. J., Jay, P., Kapun, M., Lotterhos, K. E., Mérot, C., Durmaz Mitchell, E., Pascual, M., Peichel, C. L., Rafajlović, M., Westram, A. M., Schaeffer, S. W., Johannesson, K., Flatt, T. (2023). How chromosomal inversions reorient the evolutionary process. Journal of Evolutionary Biology, 36(12), 1761–1782.

Block, B. A., Teo, S. L. H., Walli, A., Boustany, A., Stokesbury, M. J. W., Farwell, C. J., Weng, K. C., Dewar, H., Williams, T. D. (2005). Electronic tagging and population structure of Atlantic bluefin tuna. Nature, 434(7037), 1121–1127.

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Brauer, C. J., Sandoval-Castillo, J., Gates, K., Hammer, M. P., Unmack, P. J., Bernatchez, L., Beheregaray, L. B. (2023). Natural hybridization reduces vulnerability to climate change. Nature Climate Change.

Bremer,A. J. R., Viñas, J., Mejuto, J., Ely, B., Pla, C. (2005). Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. Molecular Phylogenetics and Evolution, 36(1), 169–187.

Broad Institute. (n.d.). Picard Toolkit: MarkDuplicates. Broad Institute. https://broadinstitute.github.io/picard/

Cabanettes F, Klopp C. (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6:e4958

Chen, I.-C., Lee, P.-F., Tzeng, W.-N. (2005). Distribution of albacore (Thunnus alalunga) in the Indian Ocean and its relation to environmental factors. Fisheries Oceanography, 14(1).

Chen, J.-M., Cooper, D. N., Chuzhanova, N., Férec, C., Patrinos, G. P. (2007). Gene conversion: mechanisms, evolution and human disease. Nature Reviews. Genetics, 8(10), 762–775.

Chow, S., Kishino, H. (1995). Phylogenetic relationships between tuna species of the genus Thunnus (Scombridae: Teleostei): Inconsistent implications from morphology, nuclear and mitochondrial genomes. Journal of Molecular Evolution, 41(6), 741–748.

Dahn, H. A., Mountcastle, J., Balacco, J., Winkler, S., Bista, I., Schmitt, A. D., Pettersson, O. V., Formenti, G., Oliver, K., Smith, M., Tan, W., Kraus, A., Mac, S., Komoroske, L. M., Lama, T., Crawford, A. J., Murphy, R. W., Brown, S., Scott, A. F., et al. (2022). Benchmarking ultra-high molecular weight DNA preservation methods for long-read and long-range sequencing. GigaScience.

Davies, C. A., Gosling, E. M., Was, A., Brophy, D., Tysklind, N. (2011). Microsatellite analysis of albacore tuna (Thunnus alalunga): population genetic structure in the North-East Atlantic Ocean and Mediterranean Sea. Marine Biology : International Journal on Life in Oceans and Coastal Waters, 158(12), 2727–2740.

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.

Díaz-Arce, N., Gagnaire, P.-A., Richardson, D. E., Walter, J. F. I., Arnaud-Haond, S., Fromentin, J.-M., Brophy, D., Lutcavage, M., Addis, P., Alemany, F., Allman, R., Deguara, S., Fraile, I., Goñi, N., Hanke, A. R., Karakulak, F. S., Pacicco, A., Quattro, J. M., Rooker, J. R., Arrizabalaga, H. (2024). Unidirectional trans-Atlantic gene flow and a mixed spawning area shape the genetic connectivity of Atlantic bluefin tuna. Molecular Ecology.

DFO. (2022). Government of Canada celebrates successful outcomes at the International Commision for the Conservation of Atlantic Tunas meeting.

Edelman, N. B., Mallet, J. (2021). Prevalence and Adaptive Impact of Introgression. Annual Review of Genetics, 55, 265–283.

Fitzpatrick, S. W., Bradburd, G. S., Kremer, C. T., Salerno, P. E., Angeloni, L. M., Funk, W. C. (2020). Genomic and Fitness Consequences of Genetic Rescue in Wild Populations. Current Biology, 30(3), 517–522.

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.

Fromentin, J.-M., Powers, J. E. (2005a). Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. Fish and Fisheries, 6(4), 281–306.

Fromentin, J.-M., Ravier, C. (2005b). The East Atlantic and Mediterranean bluefin tuna stock: looking for sustainability in a context of large uncertainties and strong political pressures. Bulletin of Marine Science -Miami-, 76(2), 353–362.

Fuller, Z. L., Koury, S. A., Leonard, C. J., Young, R. E., Ikegami, K., Westlake, J., Richards, S., Schaeffer, S. W., Phadnis, N. (2020). Extensive Recombination Suppression and Epistatic Selection Causes Chromosome-Wide Differentiation of a Selfish Sex Chromosome in Drosophila pseudoobscura. Genetics, 216(1), 205–226.

Fumagalli, M., Orlando, L. (2013). Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. PLoS ONE, 8(11).

Galuardi, B., Royer, F., Golet, W., Logan, J., Neilson, J., Lutcavage, M. (n.d.). Complex migration routes of Atlantic bluefin tuna (Thunnus thynnus) question current population structure paradigm. Canadian Journal of Fisheries and Aquatic Sciences, 67(6), 966.

Hedrick, P. W., Fredrickson, R. (2010). Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. Conservation Genetics, 11(2), 615–626.

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology, 22(18), 4606–4618.

Hoff, S. N. K., Baalsrud, H. T., Tooming-Klunderud, A., Skage, M., Richmond, T., Obernosterer, G., Shirzadi, R., Tørresen, O. K., Jakobsen, K. S., Jentoft, S. (2019). Long-read sequence capture of the haemoglobin gene clusters across codfish species. Molecular Ecology Resources, 19(1), 245–259.

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. Molecular Ecology, 29(14), 2535–2549.

ICCAT. (2024). Report for the biennial period, 2022-23 part II (2023). SCRS. International Commission for the Conservation of Atlantic Tunas. 2, 75-85.

ICCAT. (2021). Report for the biennial period, 2020-21 part II (2021). SCRS. International Commission for the Conservation of Atlantic Tunas. 2.

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature genetics*, *53*(3), 288-293.

Johnson, W. E., Onorato, D. P., Roelke, M. E., Land, E. D., Cunningham, M., Belden, R. C., McBride, R., Jansen, D., Lotz, M., Shindle, D., Howard, J., Wildt, D. E., Penfold, L. M., Hostetler, J. A., Oli, M. K., O'Brien, S. J. (2010). Genetic restoration of the Florida panther. Science (New York, N.Y.), 329(5999), 1641–1645.

Jun, G., Wing, M. K., Abecasis, G. R., Kang, H. M. (n.d.). BamUtil: clipOverlap. University of Michigan.

Kerr, L. A., Cadrin, S. X., Secor, D. H., Taylor, N. G. (2016). Modeling the implications of stock mixing and life history uncertainty of Atlantic bluefin tuna. Canadian Journal of Fisheries and Aquatic Sciences, 74(11), 1990–2004.

Kirkpatrick, M., Barton, N. (2006). Chromosome inversions, local adaptation and speciation. Genetics, 173(1), 419–434.

Kirkpatrick, M. (2010). How and Why Chromosome Inversions Evolve. PLoS Biology, 8(9).

Kulathinal, R. J., Stevison, L. S., Noor, M. A. F., Nachman, M. W. (2009). The Genomics of Speciation in Drosophila: Diversity, Divergence, and Introgression Estimated Using Low-Coverage Genome Sequencing. PLoS Genetics, 5(7).

Langmead, B., Salzburg, S. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods. 9: 367-359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079.

Liao, P., Satten, G. A., Hu, Y.-J. (2017). PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. Genetic Epidemiology, 41(5), 375–387.

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.

Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., Sedlazeck, F. J. (2019). Structural variant calling: the long and the short of it. Genome biology, 20, 1-14.

Mather, F.J, Mason Jr J.M., Jones, A. (1995). Historical document: life history and fisheries of Atlantic bluefin tuna. Miami, FL: National Oceanographic and Atmospheric Administration. Technical Memorandum NMFS-SEFSC-370

Marquina, D., Mateusz, B., Ronquist, F., Łukasik, P. (2021). The effect of ethanol concentration on the morphological and molecular preservation of insects for biodiversity studies. PeerJ, 9, e10799.

Mathur, S., Mason, A. J., Bradburd, G. S., Gibbs, H. L. (2023). Functional genomic diversity is correlated with neutral genomic diversity in populations of an endangered rattlesnake. Proceedings of the National Academy of Sciences, *120*(43).

Mérot, C., Oomen, R. A., Tigano, A., Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. Trends in Ecology & Evolution, 35(7), 561–572.

Muhling, B. A., Lee, S.-K., Lamkin, J. T., Liu, Y. (2011). Predicting the effects of climate change on bluefin tuna (*Thunnus thynnus*) spawning habitat in the Gulf of Mexico. ICES Journal of Marine Science, 68(6), 1051–1062.

National Center for Biotechnology Information (NCBI). (2024a). Genome assembly fThuThy2.1. National Library of Medicine (US), National Center for Biotechnology Information.

National Center for Biotechnology Information (NCBI). (2024b). Genome assembly ASM4338166v1. National Library of Medicine (US), National Center for Biotechnology Information.

Oomen, R. A., Kuparinen, A., Hutchings, J. A., Waples, R. (2020). Consequences of Single-Locus and Tightly Linked Genomic Architectures for Evolutionary Responses to Environmental Change. Journal of Heredity, 111(4), 319–332.

Physalia Courses. (2024). Population genomic inference from low-coverage whole-genome sequencing data. https://github.com/nt246/physalia-lcwgs

Porch, C. E., Bonhommeau, S., Diaz, G. A., Arrizabalaga, H., Gary, M. (2019). The journey from overfishing to sustainability for Atlantic bluefin tuna, *Thunnus thynnus*. The future of bluefin tunas: Ecology, fisheries management, and conservation, 3-44.

Pujolar, J. M., Roldán, M. I., Pla, C. (2003). Genetic analysis of tuna populations, *Thunnus thynnus thynnus* and *T. alalunga*. Marine Biology: International Journal on Life in Oceans and Coastal Waters, 143(3), 613–621.

Reglero, P., Santos, M., Balbín, R., Laíz-Carrión, R., Alvarez-Berastegui, D., Ciannelli, L., Jiménez, E., Alemany, F. (2017). Environmental and biological characteristics of Atlantic bluefin tuna and albacore spawning habitats based on their egg distributions. Deep-Sea Research Part II, 140, 105–116.

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 of the United States of America, 113(12), 3299–3304.

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., et al. (2019). Determining natal origin for improved management of Atlantic bluefin tuna. Frontiers in Ecology and the Environment, 17(8), 439–444.

Rodriguez-Marin E, Ortiz M, Ortiz de Urbina J.M., Quelle P, Walter J, Abid N, et al. (2015) Atlantic bluefin tuna (Thunnus thynnus) Biometrics and Condition. PLoS ONE 10(10): e0141478.doi:10.1371/journal.pone.0141478

Rooker, J. R., Arrizabalaga, H., Fraile, I., Secor, D. H., Dettman, D. L., Abid, N., Addis, P., Deguara, S., Karakulak, F. S., Kimoto, A., Sakai, O., Macías, D., Santos, M. N. (2014). Crossing the line: migratory and homing behaviors of Atlantic bluefin tuna. Marine Ecology Progress Series, 504, 265–276.

Rooker, J., Bremer, J. R. A., Block, B., Dewar, H., de Metrio, G., Corriero, A., Kraus, R., Prince, E., Rodríguez-Marín, E., Secor, D. (2007). Life History and Stock Structure of Atlantic Bluefin Tuna (*Thunnus thynnus*). Reviews in Fisheries Science, 15(4), 265–310.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nature Reviews. Genetics, 15(2).

Soniat, T. J., Sihaloho, H. F., Stevens, R. D., Little, T. D., Phillips, C. D., Bradley, R. D. (2021). Temporal-dependent effects of DNA degradation on frozen tissues archived at -80°C. Journal of Mammalogy, 102(2), 375–383.

Stewart, N. D., Busawon, D. S., Rodriguez-Marin, E., Siskey, M., Hanke, A. R. (2022). Applying mixed-effects growth models to back-calculated size-at-age data for Atlantic bluefin tuna (*Thunnus thynnus*). Fisheries Research, 250.

Suarez-Gonzalez, A., Lexer, C., Cronk, Q. C. B. (2018). Adaptive introgression: a plant perspective. Biology Letters, 14(3).

Twyford, A. D., Ennos, R. A. (2012). Next-generation hybridization and introgression. Heredity, 108(3), 179–189.

Valencia-Montoya, W. A., Elfekih, S., North, H. L., Meier, J. I., Warren, I. A., Tay, W. T., Gordon, K. H. J., Specht, A., Paula-Moraes, S. V., Rane, R., Walsh, T. K., Jiggins, C. D., Singh, N. (2020). Adaptive Introgression across Semipermeable Species Boundaries between Local Helicoverpa zea and Invasive *Helicoverpa armigera* Moths. Molecular Biology and Evolution, 37(9), 2568–2583.

Viñas, J., Tudela, S., Kayser, M. (2009). A Validated Methodology for Genetic Identification of Tuna Species (Genus Thunnus). PLoS ONE, 4(10).

von Bertalanffy, L. (1938). A quantitative theory of organic growth (Inquiries on growth laws. II). Human Biology, 10(2), 181–213.

Worm, B., Tittensor, D. P. (2011). Range contraction in large pelagic predators. Proceedings of the National Academy of Sciences of the United States of America, 108(29), 11942–11947.

Wellenreuther, M., Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. Trends in Ecology & Evolution, 33(6), 427–440.

Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., Tallmon, D. A. (2015). Genetic rescue to the rescue. Trends in Ecology & Evolution, 30(1), 42–49. https://doi.org/10.1016/j.tree.2014.10.009

Appendix

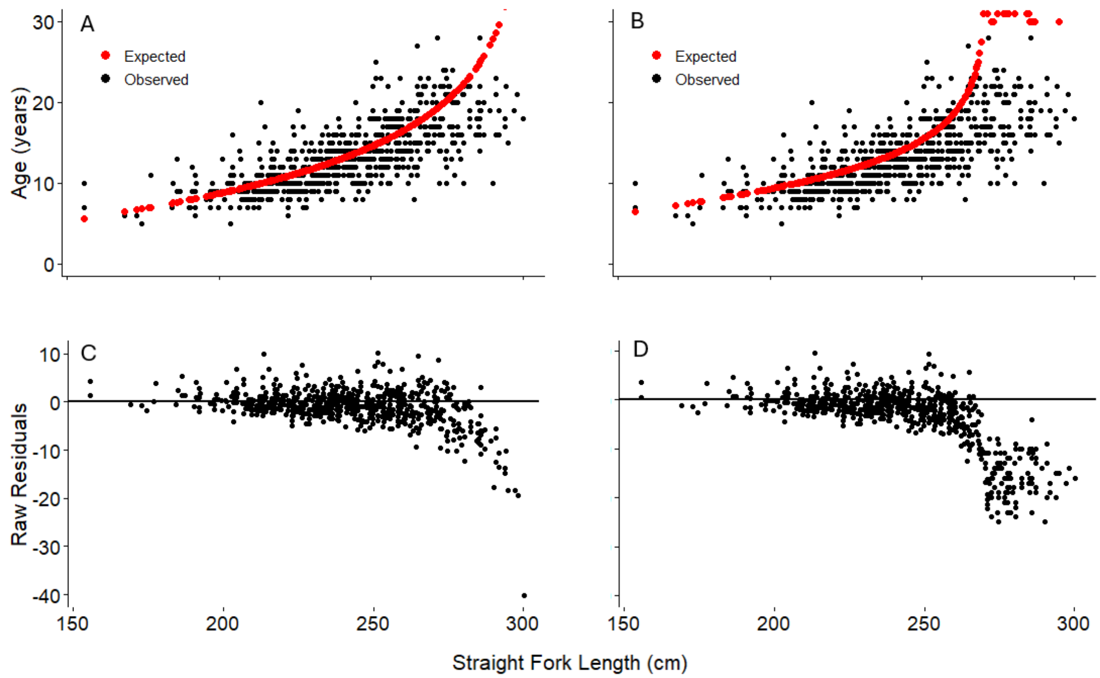

Figure A1: The straight fork length of individuals from the DFO archive caught in 2013 and 2018 and aged using otoliths plotted against their biological age (black) (A, B). The expected values (red line) in (A) represent the age of the individual estimated from a von Bertalanffy growth model using parameters derived by Stewart et al. (2022). The expected values (red line) in (B) represent the age of the individual estimated from a Richards growth curve using parameters derived by Ailloud et al. (2017). (C) displays the raw residuals of the model fit depicted in (A). (D) displays the raw residuals of the model fit depicted in (B).

Table A1: Genomic data of all sampled individuals consisting of (1) number of raw reads, before adapter clipping per individual; (2) number of trimmed reads per individual; (3) mean read depth across sites in the genome per individual; (4) mean read depth excluding zero-read sites per individual; (5) standard deviation of read depth across sites in the genome per individual; (6) Median of read depth across sites in the genome per individual; (7) the proportion of the reference genome covered (coverage).

		Number of Reads		Read depth values				
Year	Sample ID	Raw	Trim	Mean	Mean nonzero	SD	Median	Coverage (proportion)
2013	3857	23964542	23951189	3.928318938	4.32215146	2.4191628	4	0.90888045
2013	3865	26173419	26159746	4.254976408	4.66289352	2.60005033	4	0.91251846
2013	3869	45580344	45546465	7.135852503	7.64926451	3.68017694	7	0.93288087
2013	3878	19468174	19451954	3.147483845	3.55354813	2.10829725	3	0.88572991
2013	3881	26328725	26302807	4.297672181	4.69141007	2.56796895	4	0.91607259
2013	3887	21110243	21094696	3.340489845	3.73941845	2.20408387	3	0.89331801
2013	3897	29345334	29322519	4.946013818	5.34668555	2.80314083	5	0.92506166
2013	3924	26248914	26237237	4.372961307	4.77307348	2.60237852	4	0.91617305
2013	3926	22271331	22257906	3.605507576	4.003104	2.29946351	3	0.90067797
2013	3930	23831634	23818921	3.754307978	4.15261183	2.38515332	4	0.90408353
2013	3933	22825540	22810225	3.684321552	4.08406188	2.34170065	4	0.90212187
2013	3956	18948852	18935154	3.029632209	3.44305504	2.07227647	3	0.87992558
2013	3966	24689254	24674011	3.814731502	4.27710843	2.66905552	4	0.89189497
2013	4023	25124030	25104918	4.179078601	4.58893063	2.58185185	4	0.91068681
2013	4038	27269674	27250901	4.266037236	4.65203868	2.52550223	4	0.91702532
2013	4054	23437477	23423468	3.665546755	4.07769806	2.3682131	4	0.8989255
2013	4056	25226881	25213885	4.190759848	4.5924303	2.56341098	4	0.91253641
2013	4070	24463891	24447567	3.94185317	4.36110619	2.53422178	4	0.90386544
2018	6549	25896426	25882047	4.122652193	4.51278081	2.4986347	4	0.91355029
2018	6580	24637981	24606735	3.868979078	4.25713419	2.39301512	4	0.90882244
2018	6588	23730427	23701993	3.542143994	3.94193626	2.31387888	3	0.89857972
2018	6608	27016260	26981213	4.164116023	4.56994376	2.57451075	4	0.91119634
2018	6612	27291295	27275771	4.377453811	4.77334179	2.59374911	4	0.91706272
2018	6619	27887628	27870474	4.294645681	4.70285718	2.64994026	4	0.91319926
2018	6632	25061168	25039623	3.969721169	4.36729911	2.46065396	4	0.9089648
2018	6635	24905006	24894269	3.959790158	4.35027501	2.44898131	4	0.91023904
2018	6639	24091260	24079436	3.288601301	3.69882504	2.24704895	3	0.8890935
2018	6644	28484287	28430708	4.196524308	4.59065886	2.56221032	4	0.91414423

2018	6655	26668659	26642654	4.161252713	4.56704816	2.61375798	4	0.91114711
2018	6685	26334518	26317667	4.129697281	4.52790761	2.53482963	4	0.91205423
2018	6722	15612039	15601150	2.428148573	2.86000621	1.78553025	2	0.84900115
2018	6734	25984884	25951079	3.821436192	4.20820576	2.38188193	4	0.90809157
2018	6925	25223339	25206327	3.86442895	4.27443401	2.48646024	4	0.90407968
2018	6936	28468529	28447434	4.763595221	5.17438963	2.77748847	5	0.92061007
2018	6960	24636099	24605549	3.824592884	4.22079001	2.39301778	4	0.906132
2018	6968	22698163	22667543	3.279687959	3.7208616	2.33760434	3	0.8814324
2023	11276	22443912	22427534	3.588623584	4.00593963	2.39061668	3	0.89582568
2023	11279	18600745	18585356	2.950497448	3.3714204	2.05090099	3	0.87514967
2023	11284	23722904	23704911	3.863672943	4.25189551	2.38522424	4	0.90869424
2023	11290	22549476	22534913	3.592606024	3.99790069	2.34896561	3	0.89862313
2023	11306	23633053	23619039	3.846618308	4.24594497	2.42215449	4	0.90595105
2023	11314	25898651	25876300	4.191215616	4.59900461	2.58824947	4	0.91133103
2023	11316	26170040	26159002	4.234183103	4.63556191	2.5822872	4	0.91341313
2023	11319	25602390	25589878	4.251897724	4.63672874	2.53314892	4	0.91700377
2023	11320	25405096	25389768	4.104321098	4.50289261	2.52891868	4	0.91148545
2023	11326	20033846	20020811	3.279934747	3.67321167	2.15519367	3	0.89293377
2023	11327	27974253	27948592	4.610766317	5.0205292	2.72619452	5	0.91838253
2023	11330	25559404	25544688	4.160803508	4.55281251	2.50664631	4	0.9138974
2023	11339	21439435	21426938	3.465481321	3.86833478	2.26194831	3	0.89585869
2023	11353	20772333	20761113	3.362403053	3.77980246	2.26118631	3	0.8895711
2023	11363	21503786	21498397	3.23826802	3.69589414	2.34109831	3	0.87617986
2023	11549	23439455	23432285	3.83558739	4.24404484	2.4686991	4	0.90375751
2023	11341	32192649	32174536	4.818352262	5.22542997	2.85076293	5	0.9220968