ICES Journal of Marine Science, 2023, **80**, 941–952
DOI: 10.1093/icesjms/fsad012
Advance access publication date: 25 February 2023
Original Article

Weak range-wide population structure in the blackfin tuna (*Thunnus atlanticus*) revealed by analysis of genome-wide SNPs

Pavel Dimens[1], Lionel Reynal[2], Cedric Pau[2], Freddy Arocha[3], Fabio Hazin[4], Pollyana Roque[4], Nancie J. Cummings[5], James S. Franks[6], Kenneth L. Jones[7], and Eric A. Saillant[1,*]

[1]School of Ocean Science and Engineering, The University of Southern Mississippi, Ocean Springs, MS 39564, USA
[2]IFREMER Délégation de Martinique, 97231 Le Robert, La Martinique, France
[3]Instituto Oceanográfico de Venezuela, Universidad de Oriente, Cumana 6101, Venezuela
[4] Fisheries & Aquaculture Department, Universidade Federal Rural de Pernambuco, Dom Manoel de Medeiros, s/n. Recife 52171-900, Brazil
[5]Southeast Fisheries Center, 75 Virginia Beach Dr., Key Biscayne, FL 33149, USA
[6]Center for Fisheries Research and Development, The University of Southern Mississippi, Ocean Springs, MS 39564, USA
[7]Department of Cell Biology, University of Oklahoma Health Sciences Center, 975 NE 10th St., BRC 458, Oklahoma City, OK 73104, USA
* *Corresponding author:tel: (1) 228-818-8007;* e-mail: eric.saillant@usm.edu.

Blackfin tuna (*Thunnus atlanticus*) is a small tuna distributed in the western Atlantic Ocean where it is exploited by growing recreational and commercial regional fisheries. In this work, genome-wide genetic variation was analysed to investigate the occurrence of stock subdivision. A *de novo* assembly of the blackfin tuna genome was generated using Illumina paired-end sequencing data and applied as a reference for population genomic analysis of specimens from nine localities (average sample size per locality $n = 72$) spanning most of the blackfin tuna distribution range. A total of 2139 single-nucleotide polymorphisms were discovered and genotyped using the double-digest restriction associated DNA sequencing. Pairwise exact homogeneity tests were significant in 24 out of 36 population pairs and significant spatial autocorrelation of genotypes was observed for specimens collected within 2250 km of each other. However, divergence among locality samples was very low (pairwise F_{ST} range 0.0002–0.0025) and significant temporal variations were detected in localities sampled multiple times. Approaches to detect cryptic groups *de novo* were unsuccessful. Additional sampling is warranted to determine if multiple stocks need to be defined for management and assess temporal and spatial patterns of gene flow connecting them.

Keywords: blackfin tuna, double-digest RAD sequencing, population genetics, single nucleotide polymorphism, stock structure, *Thunnus atlanticus*.

Introduction

Tunas (family Scombridae) are highly specialized fast-swimming pelagic predators known to migrate large distances annually (Mariani *et al*., 2016; Pecoraro *et al*., 2016; Reglero *et al*., 2017). Their adaptations to fast swimming and long distance travel (Graham and Dickson 2004) suggest populations could be connected over broad distances, possibly at the scale of entire oceanic basins. While some smaller bodied tunas such as the skipjack tunas have a circumglobal distribution (Ely *et al*., 2005) and could display long distance connectivity as shown in large tunas, others such as the blackfin tuna and the longtail tuna have more restricted ranges (West Atlantic and Indian-West Pacific Ocean, respectively; Collette and Nauen, 1983), suggesting that the spatial scope of connectivity is more limited due to behavioural or physiological capacity.

The blackfin tuna (*Thunnus atlanticus*) is a small tuna growing to approximately 100 cm and weighing up to 21 kg (Collette *et al*., 2022), making it the smallest of the *Thunnus* genus. The species occupies the narrowest geographic range of all Atlantic true tuna species. It is restricted to the western Atlantic basin where it has been reported from Massachusetts to as far south as Brazil, although it is mostly found in tropical and sub-tropical waters where the temperature is likely to exceed 20°C. In the United States, blackfin tunas are abundant throughout the Gulf of Mexico and South Atlantic Bight regions (Collette and Nauen, 1983). They can be found at depths between 20 and 700 m but are most common between 40 and 50 m (Maghan and Rivas, 1971). They are known to form large schools with skipjack tuna and their distribution has been linked to several factors such as water clarity, steepness of the continental shelf, and plankton concentrations correlated with terrestrial runoff and upwelling zones (De Sylva *et al*., 1987). Their diet consists of surface and deep-sea fish, squid, and arthropods including amphipods, shrimps, and crabs (Frimodt and Dore, 1995; Collette *et al*., 2022). Spawning occurs from late spring to early fall when water temperatures are at or above 27°C, with a peak of activity in the early summer months (Idyll and De Sylva, 1963; Juárez, 1978; Richardson *et al*., 2010; Bezerra *et al*., 2013).

Blackfin tunas are harvested by commercial and recreational fisheries across their range (Mathieu *et al*., 2013). Historically, they were not popular for recreational fishing in the United States, but they have been increasingly targeted in recent years by recreational fishers (Saillant *et al*., 2022) with most landings occurring along the US east coast, off the Florida Keys and around Puerto Rico (Personal Communication of the National Marine Fisheries Service, Fisheries Statis-

Received: 1 June 2022; **Revised:** 2 December 2022; **Accepted:** 12 December 2022

tics Division). Commercial landings in the United States are negligible (Saillant *et al.*, 2022), although the species may be captured as bycatch of other tuna fisheries. It is harvested commercially using longlines and purse seines in the Caribbean and South America with highest landings recorded in Cuba, the Dominican Republic, the Lesser Antilles, Venezuela, and Brazil (Mathieu *et al.*, 2013). Blackfin tunas are managed at the basin level under the international jurisdiction of the International Commission for the Conservation of Atlantic Tunas (ICCAT, https://www.iccat.int/en/#) for international waters and is of high interest to the Western Central Atlantic Fisheries Commission because of importance to artisanal/small-scale fisheries by members relayed by domestic management entities such as the highly migratory species division of the National Oceanic and Atmospheric Association (NOAA) in the United States for captures within the Exclusive Economic Zone. Considering the rising popularity of blackfin tunas in the United States and other countries exploiting them in western Atlantic waters, stock structure needs to be documented to design appropriate units for management. However, there is currently no clear information on population subdivision in this species.

Based on available records of sexually mature individuals, eggs, or larvae, Mathieu *et al.* (2013) suggest that blackfin tunas reproduce over most of their distribution range, thus possibly forming a metapopulation composed of many demes. Mark—recapture studies by Luckhurst *et al.* (2001) in Bermuda and Singh–Renton and Renton (2007) in St. Vincent and the Grenadines revealed some instances of sedentary behaviours where some blackfin tunas were recaptured in the tagging area, sometimes after very long periods (4 years). However, long-distance movement was also suggested by Luckhurst *et al.* (2001) for individuals tagged in the Bermuda Islands where recaptures only occurred during the summer months while blackfin tunas were absent during cold months and hypothesized to move south during those periods. These results suggest that gene flow across geographic populations of blackfin tuna may be partially restricted by some degree of sedentary behaviour.

Information on genetic stock structure is limited to a study by Saxton (2009), comparing the Gulf of Mexico and the US east coast using six microsatellites and sequence of the control region of mitochondrial DNA and a more comprehensive study by Saillant *et al.* (2022) using 13 microsatellite markers surveyed in nine geographic population from Brazil to North Carolina. Saxton (2009) reported significant divergence between the US east coast and the Gulf of Mexico. Saillant *et al.* (2022) reported very weak divergence across the sampling surface with a possible isolation of the Brazilian population from the rest of the range and a weak pattern of isolation by distance. Both studies were limited by the small numbers of genetic loci used, which prevented assessing potential adaptive patterns, and by the lack of or incomplete temporal replication of sampling. The advent of next-generation sequencing and the development of genotyping by sequencing methods have enabled cost-effective generation of high-density genome scans including thousands of genetic loci (Peterson *et al.*, 2012). The Restriction site-associated DNA (RAD) sequencing methods have become the most popular genotyping option in molecular ecology studies due to their immediate applicability to non-model species (O'Leary *et al.*, 2018). The reliability of genotyping and the number of polymorphic loci that can be recovered are dependent on rigorous data filtering and are improved when a reference genome is available and used to map RAD sequencing reads (Shafer *et al.*, 2016), an approach implemented in studies of other tuna species (Laconcha *et al.*, 2015; Pecoraro *et al.*, 2018; Vaux *et al.*, 2021).

This study addresses the limitations of previous population genetic studies of blackfin tuna by employing high-density genome scans to describe the genetic variation in geographic populations across the species' distribution range and multiple sampling years. A draft reference genome was developed and used to map RAD sequences obtained from population samples and both neutral and non-neutral patterns of structure were investigated to assess comprehensively genetic structure accounting for local adaptation of populations.

Methods

Development of a draft reference genome assembly

Fin tissue from a single representative individual captured in the north central Gulf of Mexico was used to isolate genomic DNA using the MagBind Blood and Tissue kit (Omega Bio-Tek, catalogue number M6399-01). The sample was sequenced on the Illumina NovaSeq6000 platform to obtain 150 bp paired end reads.

The raw Illumina reads were trimmed using fastp (v0.20.0, Chen *et al.*, 2018) as described in Supplementary Materials File S1. The filtered reads were then used to estimate the size of the blackfin tuna genome using the k-mer frequency counting method (https://bioinformatics.uconn.edu/genome-size-estimation-tutorial/). The k-mer frequency distribution in illumina reads was calculated using the program jellyfish (Marçais and Kingsford, 2011), for k-mer sizes varying from 17 to 25.

Trimmed short reads were assembled using SparseAssembler (Ye *et al.*, 2012), with a k-mer size of 90. DB2OLC (Ye *et al.*, 2016) was then used on the short-read contigs to perform a consensus with a k-mer size of 31. The trimmed raw reads were mapped onto the assembly using BWA-MEM (Li and Durbin, 2009) and then used to polish the consensus with Pilon (Walker *et al.*, 2014). The trimmed short reads were then mapped onto the assembly again and used to identify and remove haplotigs using purge_haplotigs (Roach *et al.*, 2018). Finally, trimmed short reads were mapped to the obtained assembly for a final round of polishing with Pilon. Details of each assembly step are provided in Supplementary File S1. Assemblies were assessed using metrics produced by Quast (Gurevich *et al.*, 2013). Genome completeness was assessed using the *Eukaryota* database of Benchmark Universal Single Copy Orthologs (BUSCO, Simão *et al.*, 2015).

Sampling from blackfin tuna populations

A total of 650 adult blackfin tuna samples from nine geographic localities were analysed during the study (Figure 1, Table 1). Localities surveyed were offshore the US east coast (South Carolina—SCA), the Florida Keys (KEY), the north central Gulf of Mexico in the area of Pensacola (PNS), the Western Gulf of Mexico in the area of Corpus Christi (TX), the US Caribbean (offshore Puerto Rico—PR), the French Antilles (La Martinique—MRT), Venezuela (southern Caribbean—VZ), and Brazil (offshore Baía Formosa, and St. Peter and St. Paul Archipelago—BRZ and BRZ_SP) providing samples from across the species range. Samples were taken post-mortem from fish carcasses through fishery dependent

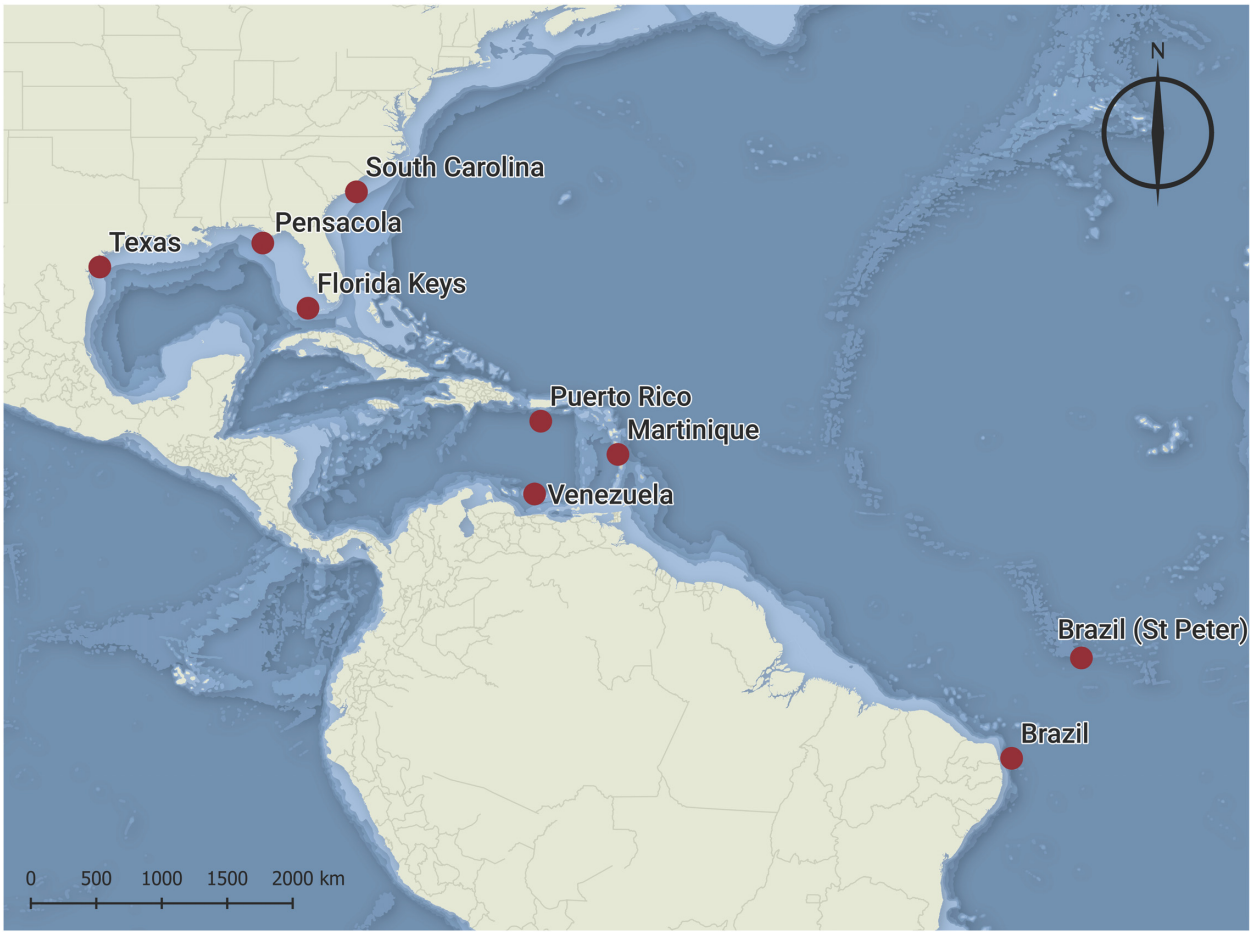

Figure 1. Sampling localities for blackfin tuna.

Table 1. Number of samples obtained for each locality and sampling year. Values in parenthesis reflect the number of samples remaining after sequence quality filtering.

Locality	ID	2015	2016	2017	2018/19
South Carolina	SCA	–	50 (23)	49 (29)	–
Florida Keys	KEY	–	48 (32)	–	41 (24)
Pensacola	PNS	–	46 (30)	–	–
Texas	TX	–	–	51 (28)	–
Puerto Rico	PR	–	44 (23)	34 (16)	–
La Martinique	MRT	–	64 (39)	–	–
Venezuela	VZ	50 (31)	50 (14)	–	–
Brazil Baia Formosa	BRZ	–	46 (6)	–	49 (17)
Brazil St. Peter/Paul	BRZ-SP	–	–	–	28 (14)

sampling. Localities were sampled between 2015 and 2019, aiming to sample each locality in two consecutive years within that range to allow assessing temporal stability of spatial patterns of structure, with a target of 50 specimens per locality per year (thus, 100 samples per locality). The actual sample sizes per locality and sampling year are reported in Table 1. The two northern Gulf of Mexico localities, La Martinique Island, and St. Peter and St. Paul archipelago could only be sampled once (in 2016 for La Martinique and Pensacola, 2017 for Texas, and at the end of 2018 for St. Peter and St. Paul). Available capture coordinates are archived with genotype data in the University of Southern Mississippi public online repository Aquila (doi: 10.18785/rwpg.ds.01). Exact coordinates were not known for samples collected during port sampling from fishermen. Captured location was assumed proximal to the landing port in those cases (within ~100–150 km, the typical range of offshore fishing trips in US waters). Tissue samples, a 1 cm^2 fin clip or 0.5 cm^3 muscle sample, were taken from each fish and stored in either 20% DMSO-EDTA, 95% ethanol, or Sarkosyl urea lysis buffer (8 M urea, 1% Sarkosyl, 20 mM sodium phosphate, and 1 mM EDTA) until DNA isolation. Sampling targeted fish of reproductive size (FL > 50 cm).

Sequencing

DNA was isolated using either the Blood & Tissue DNA HDQ 96 Kit or the EZ-96 tissue kit (Omega Bio-Tek, catalogue number D1196-01). After DNA isolation, each sample was

allocated to one of nine pools of sequencing libraries; each pool received an equal number of samples from each sampling location and year to minimize the impacts of sequencing bias that could occur if samples from individual localities and year were sequenced on separate sequencing runs. Samples were prepared for sequencing using a modified version of the Double Digest Restriction Associated DNA protocol (ddRAD-Seq; Peterson *et al.*, 2012). The modifications to the protocol include the use of EcoRI and MspI restriction endonucleases (New England Biolabs), along with custom adapters fitted with 6 bp unique barcodes allowing multiplexing up to 100 unique individuals in the same sequencing run and an 8 bp Universal Molecular Identifier (UMI, EuroFins) to isolate PCR duplicates in downstream analyses. The barcode was included in both the P1 ("forward") and P2 ("reverse") adapters to ensure proper demultiplexing of reverse sequencing reads and prevent errors due to "barcode hopping" (van der Valk *et al.*, 2020). Samples were pooled and size selected using a 300–500-bp window on a Pippin Prep (Sage Science), and the DNA concentration and fragment size distribution of the pool were assessed on a NanoDrop 2000 and an Agilent 2100 BioAnalyzer DNA chip system, respectively. The obtained libraries were sequenced at the University of Colorado Genomics and Microarray Core facility to generate on average 6 million paired-end reads (150 bp × 2) per individual using the Illumina NovaSeq6000 platform.

Data filtering

The raw sequence data were demultiplexed at the sequencing facility and processed using the dDocent pipeline (Puritz *et al.*, 2014). Briefly, raw sequences were trimmed to remove adapter sequences and filtered to remove low quality bases using a sliding window approach in fastp (Chen *et al.*, 2018). Reads were then mapped on the draft reference genome using BWA (Li and Durbin, 2009) and SNPs were called using FreeBayes (Garrison and Marth, 2012). The settings used for read mapping and SNP discovery in dDocent are available in Supplementary File S1.

The resulting raw SNP dataset were initially filtered using VCFtools (Danecek *et al.*, 2011) to retain loci with less than 50% missing data, a minimum genotype quality value as estimated by FreeBayes of 30, a minimum depth of 10, and a minimum allele frequency of 10^{-6}, and to remove individuals with more than 20% missing data. dDocent_filters (provided by dDocent) was then used to remove sites with extreme allelic balance, improperly paired reads, and SNPs called from overlapping forward and reverse reads. The vcfallelicprimatives transformation from vcflib library (https://github.com/vcflib/vcflib) was used to deconstruct multi-nucleotide polymorphisms into SNPs. Once markers were decomposed into SNPs, indels and multi-allelic SNPs were removed in VCFtools and a maximum missing data tolerance of 20% was applied. Individuals with significantly high or low heterozygosity were identified using χ^2 tests in vcftools and subsequently removed. Loci departing significantly from Hardy–Weinberg equilibrium within localities were identified in vcftools and removed. The Benjamini–Hochberg false discovery rate correction (Benjamini and Hochberg, 1995) was applied to determine significance of within-population exact tests of Hardy–Weinberg equilibrium with an alpha (-h) of 0.0055 to account for the nine tests (nine locality samples) performed simultaneously for each locus. This method was applied as it calculates heterozygosity on a per-locality basis to minimize the influence of the Wahlund effect on allele frequencies if structure is present. The data was then filtered to retain only loci with a minimum minor allele frequency of 0.01. Close kin dyads were identified with the methods presented below and removed from the dataset, followed by a final treatment of thinning SNPs within 10 kb of each other to reduce possible effects of linkage disequilibrium. A combination of the software PGDSpider2 (Lischer and Excoffier, 2012), and Julia package PopGen.jl (Bezanson *et al.*, 2017; Dimens and Selwyn, 2022) were used to convert the datasets between file formats for subsequent analyses.

Population genetic analysis

Relatedness

Studies of other tunas have revealed the presence of close-kins in the same geographic samples (Anderson *et al.*, 2019a). The presence of groups of close kins in regional samples in this study would affect inferences on population structure by confounding the similarity of members of the same sibling with the similarity of specimens from the geographic population they originate from when compared to other regions. To address this potential source of bias and describe the distribution and frequency of close kins in blackfin tuna samples, pairwise relatedness was estimated using PC-Relate (Conomos *et al.*, 2016), which builds on the KING method (Manichaikul *et al.*, 2010) and is robust against the presence of population structure (Conomos *et al.*, 2016). The pairwise relatedness matrix was partitioned into related and unrelated individuals using PC-Air (Conomos *et al.*, 2015) with default parameters and PC-Relate was performed on the unrelated set of individuals. The resulting eigenvalues were projected onto the subset of related individuals to obtain the relatedness coefficients for each sample pair. The results were validated by simulating 1000 of pairs each of full siblings, half siblings, and unrelated individuals using the PopGenSims.jl Julia package (Dimens and Selwyn, 2022) and performing the full analyses on those simulated data.

The mean number of alleles and expected heterozygosity in each locality sample were computed using Arlequin v.3.5 (Excoffier and Lischer, 2010).

Population structure

Loci potentially impacted by natural selection were identified using an outlier analysis implemented in outFLANK (Lotterhos and Whitlock, 2015) using a q threshold of 0.05. We performed a second outlier analysis using Bayescan (Foll and Gaggiotti, 2008, prior odds 100 and 10000 bootstrap iterations) with *p*-values were adjusted for multiple testing using FDR correction and a false discovery rate of 0.05. A third analysis was also performed with Baypass (Gautier, 2015, 10000 burn in iterations, 10000 evaluations, pilot run length 1000; Olazcuaga *et al.*, 2020), comparing empirical results to results using the same parameters on 10000 randomly simulated loci and evaluating significance at a confidence interval of 99%.Pairwise F_{ST} estimates (Hudson *et al.*, 1992; Bhatia *et al.*, 2013) were computed and their significance assessed using 10000 permutations of alleles in Arlequin v.3.5. Hierarchical analyses of molecular variance (Excoffier *et al.*, 1992) were conducted in Arlequin accounting for geographic localities and sampling year within locality. Significance of covari-

ance components was assessed based on 10000 permutations of haplotypes.

Structuring according to an isolation by distance model was first assessed by calculating the least-cost paths between localities using the R package marmap (Pante and Simon-Bouhet, 2013), followed by testing the correlation between genetic and geographic distances using a Mantel test in Genalex 6.5.1 (Smouse *et al.*, 1986; Peakall and Smouse, 2012). Genetic distance was estimated using the multilocus distance of Smouse and Peakall (1999). The logarithm of geographic distance was used in the computations to account for dispersal in a 2-dimensional habitat. Occurrence of spatial structuring was also examined using spatial autocorrelation analysis in Genalex. This analysis allows detecting patterns of spatial structure (through analysis of correlation of genotypes) even if variation does not follow the strict isolation by distance model across the entire distance range sampled as assumed in Mantel tests. The multilocus spatial autocorrelation coefficient (r), was computed based on geographic least-cost distance and the multilocus genetic distance described by Smouse and Peakall (1999). When spatial autocorrelation is occurring, the estimated value of r among proximal samples differs significantly from zero and decreases with increasing geographic distance. Because the estimation of spatial autocorrelation is influenced by the size of the distance class (Peakall *et al.*, 2003), r was computed based on a series of increasing distances between sampling locations. The distance at which r no longer differs significantly from zero provides an approximation of the distance at which genetic divergence (population structure) can be inferred (Peakall *et al.*, 2003). Significance of r was determined via 1000 random permutations of genotypes among distance classes; significance of spatial autocorrelation coefficients was inferred when the observed estimate of r lied beyond the upper 95% limit of the distribution of r values obtained during the 1000 permutations (Peakall and Smouse, 2012).

A Discriminant Analysis of Principal Components (DAPC) was implemented using the R package Adegenet (Jombart, 2008; Jombart *et al.*, 2010; R Core Team, 2013). The optimal number of principal components to retain was determined using the cross-validation method on samples grouped by locality-year with 100 iterations accounting for 1–300 components. DAPC was first conducted grouping samples a priori in regions (northwestern Atlantic (US East coast samples) Gulf of Mexico (Texas, Florida Keys, and Pensacola), Caribbean Sea (Puerto Rico, La Martinique, and Venezuela), and Brazil (St Peter and St Paul, Baía Formosa) to describe differentiation between these groups.

The following analyses were conducted to discover population clusters within the data set *de novo*. A Principal component analysis was performed on the data using the R language package ape (R Core Team, 2013; Paradis and Schliep, 2019). A spatial principal component analysis was performed using the R packages ape and adegenet (Jombart, 2008), where distances were calculated using a minimum spanning tree (type 4) and geographic distance estimated in marmap as described above (spatial autocorrelation analysis). A DAPC was conducted without a priori grouping samples in localities or regions, but instead inferring the optimal number of genetic clusters present in the data *de novo* by applying k-means clustering. The optimal number of principal components inferred from cross-validation was used in k-means clustering and the DAPC was then performed on the samples with groups reclassified by their K-cluster assignments. Cryptic structure within the sampled range was also examined using model-based Bayesian clustering in fastStructure (Raj *et al.*, 2014) accounting for a range of one to nine clusters.

Table 2. Characteristics of the *T. atlanticus* de novo genome assembly.

Metric	Value
# contigs (≥0 bp)	203 667
# contigs (≥1 000 bp)	74 166
# contigs (≥5 000 bp)	26 463
# contigs (≥10 000 bp)	13 084
# contigs (≥25 000 bp)	3 117
# contigs (≥50 000 bp)	454
Largest contig	158 390 bp
Total length	514 764 407 bp
Total length (≥1 000 bp)	478 919 186 bp
Total length (≥5 000 bp)	365 013 046 bp
Total length (≥10 000 bp)	270 573 561 bp
Total length (≥25 000 bp)	117 994 980 bp
Total length (≥50 000 bp)	29 616 960 bp
N50	11 597 bp
N75	4 773 bp
L50	10 887
L75	27 556
GC content	39.68%
# N's	47
BUSCO Full (*eukaryota*)	44.55%
BUSCO Partial (*eukaryota*)	18.81%

Results

Genome sequencing and assembly

Illumina sequencing produced 730100108 raw reads for a total of 110245116308 bp. Sequence filtering retained 710740790 reads for a total of 107321859290 bp. The total length of the assembly was 514764407 bp in 203667 contigs with a GC content of 39.71%. A large fraction of the assembly was in small contigs with only 365013046 bp in contigs over 5 kb. The N50 and L50 were 10873 bp and 11820 bp, respectively and the largest contig was 158390 bp (Table 2). The estimate of the size of the Blackfin tuna genome using the k-mer frequency spectrum counting method obtained with varying k-mer sizes were all between 773 and 791 mb (μ = 785121418 bp, σ = 5416115). According to this estimate, the assembly included approximately 65% of the Blackfin tuna genome and sequencing covered the genome at a depth of 139X with filtered reads. BUSCO assessment of the completeness of the assembly indicated that it contained 44.55% and 18.81% full and partial orthologs for the *eukaryota* database, respectively, suggesting an approximate 63% assembly completeness.

Population genomics analysis

The filtering process reduced the data to 2139 biallelic SNPs across 334 samples (Supplementary Table S1). The numbers of individuals retained per geographic population averaged 36 and ranged between 14 and 57.

The KING and PCRelate analyses were performed using four principal components and identified four full-sibling pairs, and one half-sibling pair (Figure 2). Among these observed kin pairs, all full sibling pairs consisted of samples that were in adjacent position in the sample set. These pairs were

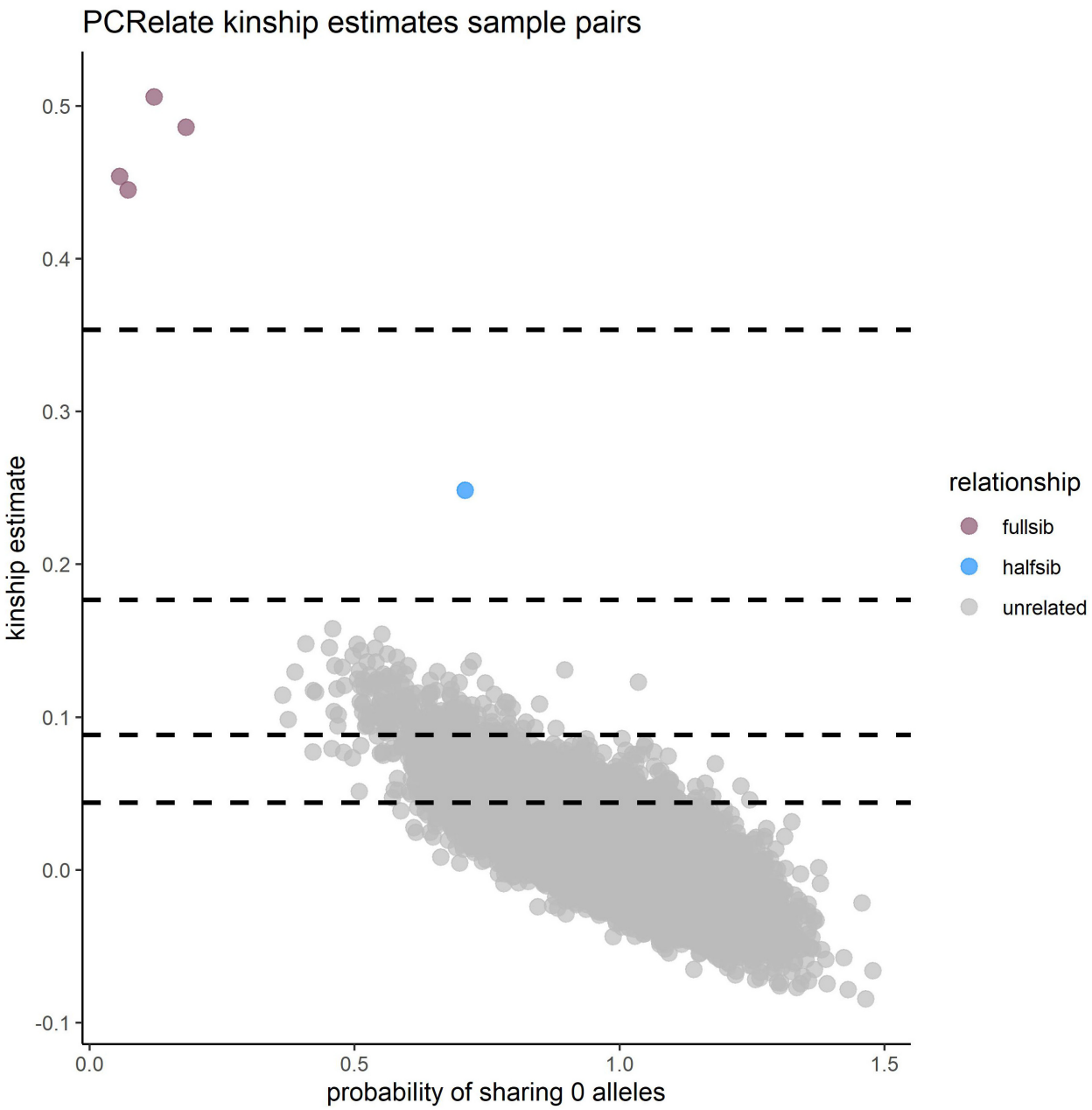

Figure 2. Pairwise relatedness (*r*) estimates between 334 Blackfin tunas collected from nine localities in the western Atlantic Ocean. The distribution of estimates of *r* (*x*-axis) is represented as a function of the estimated probability of sharing no alleles (k0, *x*-axis).

conservatively removed as they possibly reflected contamination during sampling or processing.

OutFLANK detected 55 outlier SNPs (Figure 3) although only eight of these had expected heterozygosity values > 0.1 and were the most robust candidate outliers (Whitlock and Lotterhos, 2015). Bayescan identified five SNP outlier, three of which were also putative outliers identified by outFLANK discussed above. Baypass analysis identified 58 putative SNP outliers, none of whom were corroborated by either Bayescan or outFLANK, and whose heterozygosities were all < 0.1.

Pairwise F_{ST} estimates (Table 3) ranged from 0.0007 (VZ-PNS) to 0.006 (TX-BRZSP). The corresponding exact homogeneity tests were significant ($\alpha = 0.05$) for 24 (of 36) population pairs after FDR correction accounting for a false discovery rate of 5% (Table 3).

Analyses of molecular variance revealed very weak spatial and temporal components of molecular variance (0.04% of molecular variance, $p = 0.253$ for geographic locality, 0.22% of molecular variance, $p = 0.012$ for year of capture nested in locality), which was consistent with the very low F_{ST} estimates reported above.

The Mantel test yielded a non-significant correlation between genetic distance and the logarithm of geographic distance ($r = 0.02$, $p = 0.083$). Spatial autocorrelation analysis runs using variable distance classes in increments of 250 km revealed that the highest correlation of genotypes was observed when samples were aggregated within a 500 km distance and decreased rapidly for distance classes above 750 km (Figure 4). Spatial autocorrelation remained significant for distances up to 2250 km.

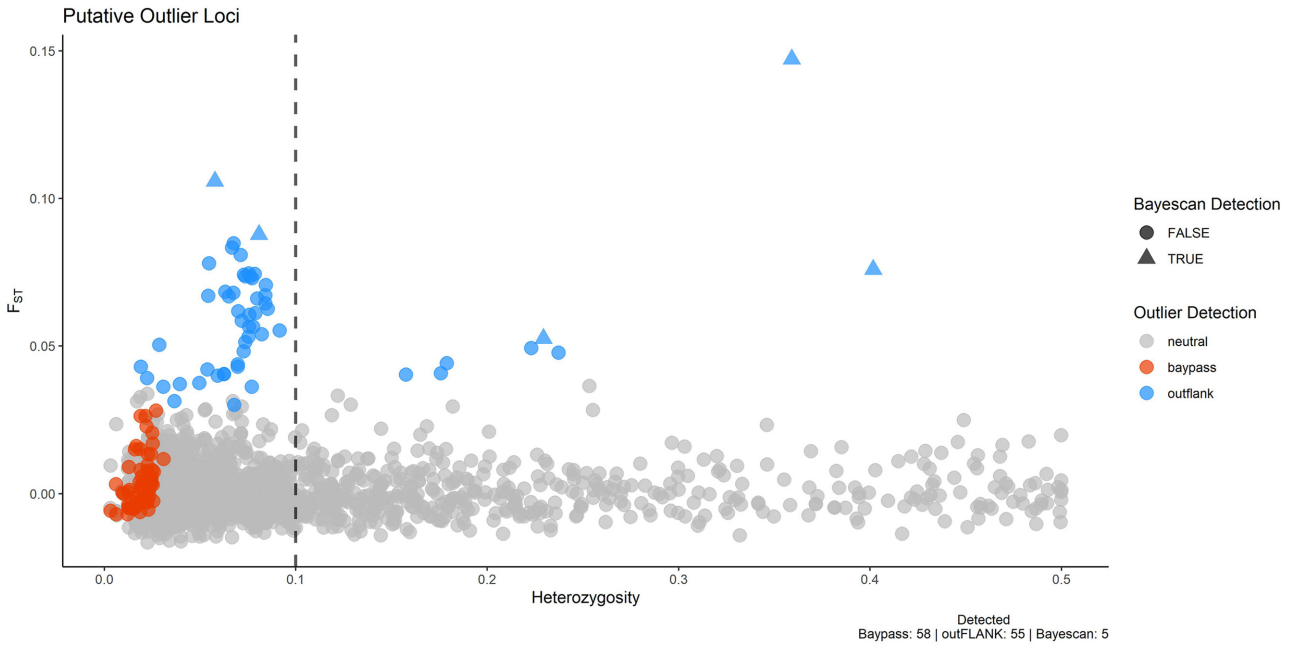

Figure 3. Outlier loci scans using three methods: outFLANK, Bayescan, and Baypass. Points denote loci, with heterozygostiy (*x*-axis) plotted against F_{ST} (*y*-axis). The vertical line represents $H_e = 0.1$.

Table 3. Pairwise F_{ST} estimates of SNP data (below diagonal) and associated *p*-values (above diagonal) comparing samples of Blackfin tuna geographic populations. F_{ST} values are rounded to four decimal places.

	BRZ	BRZSP	KEY	MRT	PNS	PR	SCA	TX	VZ
BRZ		∗0.021	∗0.002	∗0.001	∗0.018	0.2543	0.1471	∗0.001	0.0791
BRZSP	0.0032		∗0.002	∗0.013	0.4655	∗0.033	0.035	∗0.002	0.1461
KEY	0.0028	0.0049		∗0.014	∗0.024	∗0.006	∗0.008	∗0.001	∗0.013
MRT	0.0033	0.0042	0.0014		0.1792	∗0.001	∗0.003	∗0.013	∗0.025
PNS	0.0021	0.0017	0.0016	0.0011		0.038	0.2883	∗0.001	0.5445
PR	0.0016	0.0038	0.0017	0.0024	0.0016		∗0.007	∗0.001	0.1481
SCA	0.0018	0.004	0.0013	0.0018	0.0011	0.0017		∗0.001	0.2012
TX	0.0051	0.0062	0.0045	0.0022	0.0042	0.005	0.0037		∗0.001
VZ	0.0019	0.0031	0.0013	0.0014	0.0007	0.0011	0.0008	0.0039	

An asterisk (∗) denotes a *p*-value significant at $\alpha = 0.05$ after FDR correction.

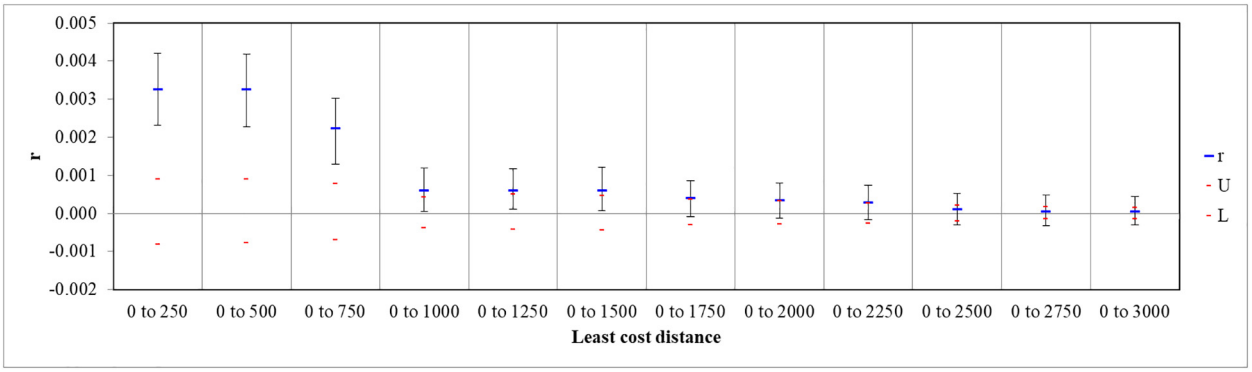

Figure 4. Correlograms illustrating the influence of geographic distance on spatial autocorrelation. Correlation (*r*) of genotypes sampled in proximal locations estimated when the first distance class increases 100 km increments (*x*-axis) visualized against spatial autocorrelation (*r*, *y*-axis). 95% bootstrapped confidence error bars for *r* (black). Red dash symbols represent upper and lower bounds of a 95% *CI* for *r* generated under the null hypothesis of a random geographic distribution of samples.

DAPC accounting for geographic regions taken as *a priori* groups visualized a weak degree of differentiation of the South Carolina samples and the Brazilian samples (southwestern Atlantic) that both diverged from a central group that included the Gulf of Mexico and Caribbean sample groups that appeared to show a greater degree of overlap (Supplemental File S2).

Approaches to detect structure de novo (principal component analysis, spatial principal component analysis, k-means clustering followed by DAPC, Bayesian clustering in FASTStructure) did not reveal interpretable cryptic units within the dataset. The results of these approaches are illustrated in Supplemental File S2.

Discussion

In this work, samples from localities spanning most of the Blackfin tuna's distribution range were characterized using 2139 SNP loci, providing substantially improved inference power compared to previous studies of genetic variation in this species. The dataset also provided a first assessment of loci putatively under divergent selection.

All pairwise F_{ST} values were very low (<0.007) indicating divergence among geographic populations was very weak across the sampled range, a finding consistent with past surveys of Blackfin tuna populations using microsatellite markers alone (Saillant *et al.*, 2022) or in combination with mtDNA sequence variation (Saxton, 2009). Weak divergence across large geographic areas is common in tunas (Barth *et al.*, 2017; Pecoraro *et al.*, 2018; Anderson *et al.*, 2019b) and other large pelagics and likely reflects high gene flow facilitated by their ability to travel long distances, combined with reduced effects of genetic drift. Blackfin tuna display high levels of genetic diversity (Antoni *et al.*, 2014), which suggest they harbor large population sizes and their differentiation under genetic drift is, therefore, expected to be slow. Blackfin tunas, as other species in the region, are presumed to have expanded their range following the last glacial maxima ~20 kya (Ely *et al.*, 2005; Pruett *et al.*, 2005). Accordingly, some geographic populations may be currently isolated, but not have accumulated enough genetic difference to be detectable with present methods, especially if populations experience periodic residual gene flow (Pruett *et al.*, 2005). This scenario is plausible for Blackfin tuna due to the species' high mobility during early life stages (passive dispersal) and adult life stages (active migration).

Clustering methods to detect population structuring de novo did not reveal occurrence of major groups. Spatial structuring was detected during spatial autocorrelation analyses suggesting structuring was related in part to geographic distance. In this study, the slope of the isolation by distance model was not significantly different from zero, but significant spatial autocorrelation of samples collected within 2250 km was observed and the correlation was highest among samples collected within 500–750 km of one another. Isolation by distance was also inferred from the study of variation at microsatellites in a previous study (Saillant *et al.*, 2022) and is consistent with the site fidelity of adults observed in tagging studies and/or limited dispersal at the larval stage. Tagging studies to date were conducted in Bermuda (Luckhurst *et al.*, 2001) and the Southern Caribbean (Singh–Renton and Renton, 2007) and indeed indicated site fidelity of tagged fish (Luckhurst, 2014), tentatively suggesting that site fidelity of adults could contribute to the isolation of geographic stocks and isolation by distance.

The 750 km distance at which spatial structure was highest is close to the maximum distance separating the main geographic areas (US east coast, Gulf of Mexico, Caribbean Sea, and Brazil). Disentangling the role of isolation by distance and that of possible discontinuities within the range is difficult with this dataset because of the distribution of the sampled localities. Characterizing additional localities within the range to increase sampling density would be helpful to formally determine whether discontinuities occur between the four groups or if the genetic structure is truly primarily explained by an isolation by distance model. For example, divergence of the Brazilian group from the northern hemisphere populations was suggested by previous microsatellite data (Saillant *et al.*, 2022) and was suggested by the pairwise F_{ST} involving the two Brazilian localities and the DAPC using *a priori* groups. Sampling blackfin tuna populations between north Brazil and the Southern Caribbean Sea would be useful to assess whether a discontinuity is occurring and further assess patterns and rates of gene flow between these two regions.

Information on the geographic location of capture was limited for some of the localities where samples were obtained from fishing boats at landing who did not communicate the exact coordinates of captures (captures were assumed to have occurred within 150 km of the landing port in those cases). Therefore, comparisons of genotypes collected at small distances were lacking from the dataset and may have prevented detection of isolation by distance in Mantel tests. In non-equilibrium situations, isolation by distance establishes first at short distance scales (Robledo–Arnuncio and Rousset, 2010) and reaches a plateau when geographic distance between samples exceeds $0.56\sigma/\sqrt{2\mu}$, where σ is the standard deviation of parental position relative to offspring position and μ is the mutation rate (Rousset, 2008). Future studies incorporating a larger number of proximal localities with accurate capture coordinates would be valuable to refine the isolation by distance model and estimate dispersal distance parameters.

Divergence between the US East coast and the Gulf of Mexico was suggested by an earlier study using mitochondrial DNA and six heterologous microsatellites (Saxton, 2009) but not confirmed in the study of Saillant *et al.* (2022) with 13 homologous microsatellites, who only reported a weak isolation by distance pattern and no subdivision within Blackfin tunas sampled north of Brazil. This study showed marginal divergence between Gulf of Mexico and US east coast during pairwise homogeneity tests. These inconsistencies likely reflect, in part, that previous datasets had insufficient power to detect the very fine divergence between the two groups. Connectivity between the Gulf of Mexico, Caribbean and East US coast could occur at the larval stage through passive transport; *Thunnus* larvae tend to be widely distributed in the continental shelf and the continental slope in the north central Gulf of Mexico (Cornic *et al.*, 2017). The loop current which becomes the Florida current and then the Gulf Stream (http://oceancurrents.rsmas.miami.edu/atlantic/atlantic.html) was discussed to promote favourable conditions for *Thunnus* larvae when it extends farthest north (Cornic *et al.*, 2017). Larvae distribution overlapped with the current itself and some larvae (e.g. located East of the Mississippi river) could, therefore, be transported to the Florida Keys or the east coast of the United States. Survival of *Thunnus* larvae is hypothesized to be pro-

moted in frontal zones (Lang *et al.*, 1994) where food is available, and if spawning occurs near these mesoscale structures, proximity to eddies promotes continued passive larval dispersal along with suitable foraging habitat for early life stages (Bakun, 2006). Eddies spinning off the loop current promote opportunities for movement as they propagate, typically westward (i.e. from central Gulf to the western Gulf (Damien *et al.*, 2021), but larvae caught within the main Loop Current and transported over long distances towards the US east coast would be expected to be outside of the favourable conditions promoted by eddies and may have low survival. Accordingly, recruitment would be promoted in the northern Gulf (yet with mixing within the Gulf), contributing to isolate this group from the US east coast.

The low F_{ST} between localities was a major challenge in this study and likely contributed to the lack of significance in most of the spatial analyses. Low F_{ST} has been shown to create clustering inaccuracies (Miller *et al.*, 2020) and F_{ST} values in the range of those obtained here are incompatible with detection of subdivision in structure (Chen *et al.*, 2007). This issue can be overcome in future studies by increasing the sampling density as discussed above as the power to detect isolation by distance patterns is improved when samples separated by short distances are included during estimation (Leblois *et al.*, 2003). Increasing the density of the genome scan with methods such as low coverage whole genome sequencing (Clucas *et al.*, 2019) would also improve the likelihood of detecting structure related to local adaptation when it occurs. Population structure of other tunas was indeed revealed by markers under selection, even when groups were homogeneous at neutral markers (e.g. Pecoraro *et al.*, 2018).

In this study, only eight outlier loci identified in outFLANK had heterozygosity > 0.1 and can be considered robust candidate loci experiencing selection (Whitlock and Lotterhos, 2015). The small number of candidate outlier loci found in this work suggests that, if they exist, the genomic regions affected by divergent selection and local adaptation may be very limited. However, considering the number of loci surveyed in this genome scan (2139) and estimates of the size of blackfin tuna genome (774 Mb), the average interval between markers was expected to be 362 kb such that a selected locus would be expected to be within 181 kb of one of the markers surveyed in this study. Genomic regions affected by selection may have remained undetected considering the average size of linkage blocks in studies of other fish is only a few kilobases (Lowry *et al.*, 2017). We note that three pairs of the 55 candidate outliers identified by outFLANK were SNPs defined on the same genomic contigs, which strengthens the inference of selection at these loci. Increasing the density of the genome scan is warranted to capture a greater fraction of adaptive variation in the species. The estimates of pairwise F_{ST} identified in this work are on average 27.5 times higher in the outlier dataset than the neutral dataset, and sampling a greater proportion of the genome not only will provide more information on local adaptation but will also improve the power to detect population subdivision. Greater genomic sampling can be achieved with a more complete genome assembly and a larger set of loci across the entire genome such as those derived from low coverage whole-genome sequencing (Therkildsen and Palumbi, 2017; Clucas *et al.*, 2019). Information on the genomic proximity of genetic loci would also allow performing a sliding window analysis where F_{ST} is assessed in groups of markers located in the same genomic regions. This approach is expected to reduce the occurrence of false positive outliers by observing the lack of signal in neighboring loci (Hohenlohe *et al.*, 2010; Bourret *et al.*, 2013). The draft reference assembly generated in this study was incomplete and highly fragmented due to the type of sequencing data used to generate the assembly and information on the genomic proximity of candidate outlier loci was limited to only those sharing the same contig. A more complete genome assembly would also facilitate performing sliding window analyses to identify putative genomic regions of selection, if they exist.

The marginal evidence for divergent selection and local adaptation may also be related to the high levels of gene flow in blackfin tuna. High gene flow is expected to counterbalance the differentiation caused by divergent selection and local adaptation, effectively preventing local adaptation from occurring, or limiting it to loci affected by strong selective pressures (Lenormand, 2002; Conover *et al.*, 2005; Cheviron and Brumfield, 2009). Genomic studies of other marine species revealed the occurrence of outliers in metapopulations that were also exhibiting structure at neutral markers (Nielsen *et al.*, 2009; Bradbury *et al.*, 2010; Limborg *et al.*, 2012; Laconcha *et al.*, 2015). However, outlier loci were also discovered in metapopulations where no significant spatial structure was observed at neutral loci (Lamichhaney *et al.*, 2012; Pujolar *et al.*, 2014; Grewe *et al.*, 2015). It is possible that blackfin tunas utilize their high capacity for movement to select habitats with favourable characteristics across their range leading to little or no local selection, although the wide range utilized by the species suggests that regional populations would differ by some environmental characteristics such as the differences in reproductive season timing discussed above for the South American group.

Finally, while a few close-kin pairs were identified (four full siblings, one half sibling), the full sibling pairs identified were all adjacent samples in our sample set. Each of the four sibling pairs co-occurred in the same locality, which could indicate true co-location of siblings bit could also reflect sample contamination. The remaining half sibling pair (SCA—VZ) may be a true close-kin pair, which would suggest dispersal between the US east coast and northern South America.

A mechanisms potentially contributing to the incidence of co-located kins include some behavioural cohesion ("close kin co-dispersal"), where larvae spawned by the same parents remain together through early life stages and in some cases may stay together through sexual maturity (Anderson *et al.*, 2019b). Another potential explanation is a sweepstake recruitment where cohorts of a regional population include a disproportionate contribution of a few siblings (Hedgecock and Pudovkin, 2011). Further sampling to investigate patterns of relatedness across various life stages is warranted to better evaluate the occurrence of co-located close-kins in blackfin tuna.

In conclusion, this study indicates very weak divergence among geographic stocks and no clear cryptic subdivision in major groups within the sampled range. Isolation by distance, although weak, seems to be a structuring factor as spatial autocorrelation was found in this and a previous studies of blackfin tuna but needs to be formally evaluated with a more even sampling. Increased sample sizes in regional populations and higher marker density would be valuable to evaluate if discrete units occur within the range that would warrant separate management and describe their temporal and geographic distribution.

Acknowledgements

This paper is dedicated to the memory of Professor Fabio Hazin who passed away from Covid complications in 2021. Dr Hazin was a pioneering and outstanding scientist who actively participated in several international negotiations dedicated to the sustainable management of fisheries resources, for the conservation of tunas and tuna-like species in the Atlantic Ocean and its adjacent seas. His great knowledge, pragmatism, and diplomacy will be sorely missed. He left an important legacy to friends, colleagues, students, and the international community.

The authors would like to thank Capt. M. Hanke (Light Tackle Fishing Paradise, Fajardo Puerto Rico), Brett Falterman, and LDWF personnel for help with samples from Puerto Rico and the north central Gulf of Mexico, Christopher Manley and Capt. Mike Gutierrez for help with sampling in the western Gulf of Mexico, Captain Myles Colley for assistance with collection in the north central Gulf of Mexico, Amy Dukes (SCDNR) for help with collection in South Carolina, and Kelly Sullivan and the personnel at FRWI for help with collection in the Florida Keys. The authors would also like to thank Brooke McPeak and the other members of the USM Aquaculture and Fisheries Genomics Laboratory for help in the laboratory.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

The following Supplementary material is available at ICESJMS online. Detailed supplementary protocols for processing of sequencing data featuring links to custom scripts used for this study are provided in Supplemental Material S1. Supplemental File S2 includes supplementary tables and figures.

Conflict of interest

The authors have no conflict of interest to declare.

Author contributions

E.S. and PD: conceptualization, data analysis, visualization, data curation, and writing—original draft, PD: data acquisition, ES: project administration and funding acquisition, PD and KLJ: software, LR, CP, FA, FH, PR, JSF, and NJC: resources, all authors: writing—review and editing, KLJ: methodology.

Data availability

The raw illumina sequencing reads obtained during genome sequencing are deposited in the NCBI Sequence Read Archive under the BioProject #PRJNA504598. The geographic collection location information for each sample and the SNP data file used for analyses in a vcf file format are made available in the public repository aquila (doi: 10.18785/rwpg.ds.01)

References

Anderson, G., Lal, M., Hampton, J., Smith, N., and Rico, C. 2019a. Close kin proximity in yellowfin tuna (*Thunnus albacares*) as a driver of population genetic structure in the tropical western and Central Pacific Ocean. Frontiers in Marine Science, 6: 341.

Anderson, G., Hampton, J., Smith, N., and Rico, C. 2019b. Indications of strong adaptive population genetic structure in albacore tuna (*Thunnus alalunga*) in the southwest and central Pacific Ocean. Ecology and Evolution, 9: 10354–10364.

Antoni, L., Luque, P. L., Naghshpour, K., Reynal, L., and Saillant, E. A. 2014. Development and characterization of microsatellite markers for blackfin tuna (*Thunnus atlanticus*) with the use of Illumina paired-end sequencing. First Break, 112: 322–325.

Bakun, A. 2006. Fronts and eddies as key structures in the habitat of marine fish larvae: opportunity, adaptive response and competitive advantage. Scientia Marina, 70: 105–122.

Barth, J. M. I., Damerau, M., Matschiner, M., Jentoft, S., and Hanel, R. 2017. Genomic differentiation and demographic histories of Atlantic and Indo-Pacific yellowfin tuna (*Thunnus albacares*) populations. Genome Biology and Evolution, 9: 1084–1098,

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57: 289–300.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. 2017. Julia: a fresh approach to numerical computing. SIAM Review, 59: 65–98.

Bezerra, N. P. A., Fernandes, C. A. F., Albuquerque, F. V., Pedrosa, V., Hazin, F., and Travassos, P. 2013. Reproduction of blackfin tuna *Thunnus atlanticus* (Perciformes: Scombridae) in Saint Peter and Saint Paul Archipelago, Equatorial Atlantic, Brazil. Revista de Biologia Tropical, 61: 1327–1339.

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

Bourret, V., Kent, M. P., Primmer, C. R., Vasemägi, A., Karlsson, S., Hindar, K., McGinnity, P. *et al.* 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Molecular Ecology, 22: 532–551.

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V. R. *et al.* 2010. Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. Proceedings of the Royal Society B Biological Sciences, 277: 3725–3734.

Chen, C., Durand, E., Forbes, F., and François, O. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology Notes, 7: 747–756.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34: i884–i890.

Cheviron, Z. A., and Brumfield, R. T. 2009. Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. Evolution Internation Journal of Organic Evolution, 63: 1593–1605.

Clucas, G.V., Lou, R.N., Therkildsen, N.O., and Kovach, A.I. 2019. Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. Evolutionary Applications, 12: 1971–1987.

Collette, B. B., and Diemens, P. 2022. *Thunnus atlanticus*. The IUCN Red List of Threatened Species 2022 e.T155276A46931209 https://dx.doi.org/10.2305/IUCN.UK.2022-1.RLTS.T155276A46931209.en

Collette, B. B., and Nauen, C. E. 1983. FAO Species Catalogue. Vol. 2. Scombrids of the World. An Annotated and Illustrated Catalogue of Tunas, Mackerels, Bonitos and Related Species Known to Date. 2. FAO Fisheries Synopsis number 125. Food and Agricultural Organization, Rome. 137 pp.

Conomos, M. P., Miller, M. B., and Thornton, T. A. 2015. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genetic Epidemiology, 39: 276–293.

Conomos, M. P., Reiner, A. P., Weir, B. S., and Thornton, T. A. 2016. Model-free estimation of recent genetic relatedness. American Journal of Human Genetics, 98: 127–148.

Conover, D. O., Arnott, S. A., Walsh, M. R., and Munch, S. B. 2005. Darwinian fishery science: lessons from the Atlantic silverside (*Menidia menidia*). Canadian Journal of Fisheries and Aquatic Sciences, 62: 730–737.

Cornic, M., Smith, B. L., Kitchens, L. L., Alvarado Bremer, J. R., and Rooker, J. R. 2017. Abundance and habitat associations of tuna larvae in the surface water of the Gulf of Mexico. Hydrobiologia, 806: 1–18.

Damien, P., Sheinbaum, J., Pasqueron de Fommervault, O., Jouanno, J., Linacre, L., and Duteil, O. 2021. Do Loop Current eddies stimulate productivity in the Gulf of Mexico? Biogeosciences, 18: 4281–4303.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E. *et al.* 2011. The variant call format and vcftools. Bioinformatics, 27: 2156–2158.

De Sylva, D. P., Rathjen, W. F., and Higman, J. B. 1987. Fisheries development for underutilized Atlantic tunas: blackfin and little tunny. NOAA Technical Memorandum NMFS-SEFC-191. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Washington, DC.

Dimens, P. V., and Selwyn, J. 2022. BioJulia/PopGen.Jl: V0.8.0 (v0.8.0). Zenodo.https://doi.org/10.5281/zenodo.6450254

Ely, B., Viñas, J., Alvarado Bremer, J. R., Black, D., Lucas, L., Covello, K., Labrie, A. V. *et al.* 2005. Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (*Thunnus albacares*) and the skipjack tuna (*Katsuwonus pelamis*). BMC Evolutionary Biology, 5: 19.

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

Excoffier, L., Smouse, P. E., and Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, 131: 479–491.

Foll, M., and Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics, 180: 977–993.

Frimodt, C., and Dore, I. 1995. Multilingual Illustrated Guide to the World's Commercial Coldwater Fish ("Fishing News" Books). Wiley-Blackwell, Oxford.

Garrison, E., and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.https://doi.org/10.48550/arXiv.1207.3907

Gautier, M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics, 201: 1555–1579.

Graham, J. B., and Dickson, K. A. 2004. Tuna comparative physiology. Journal of experimental biology, 207(23): 4015–4024.

Grewe, P. M., Feutry, P., Hill, P. L., Gunasekera, R. M., Schaefer, K. M., Itano, D. G., Fuller, D. W. *et al.* 2015. Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. Scientific Reports, 5: 16916.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics, 29: 1072–1075.

Hedgecock, D., and Pudovkin, A. I. 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. Bulletin of Marine Science, 87: 971–1002.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics, 6: e1000862.

Hudson, R. R., Slatkin, M., and Maddison, W. P. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics, 132: 583–589.

Idyll, C. P., and De Sylva, D. P. 1963. Synopsis of Biological Data on the Blackfin Tuna *Thunnus atlanticus* (Lesson) 1830 (Western Atlantic). Food and Agricultural Organization of the United Nations, Rome.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24: 1403–1405.

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

Juárez, M. 1978. Distribución de las larvas de la familia Scombridae en aguas adyacentes a las Bahamas. Revista Cubana de Investigaciones Pesqueras, 3: 69–77.

Laconcha, U., Iriondo, M., Arrizabalaga, H., Manzano, C., Markaide, P., Montes, I., Zarraonaindia, I. *et al.* 2015. New nuclear SNP markers unravel the genetic structure and effective population size of Albacore Tuna (*Thunnus alalunga*). PLoS ONE, 10: e0128247.

Lamichhaney, S., Martinez Barrio, A., Rafati, N., Sundström, G., Rubin, C.-J., Gilbert, E. R., Berglund, J. *et al.* 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *In* Proceedings of the National Academy of Sciences of the United States of America, 109: 19345–19350. PNAS.

Lang, K. L., Grimes, C. B., and Shaw, R. F. 1994. Variations in the age and growth of yellowfin tuna larvae, *Thunnus albacares*, collected about the Mississippi River plume. Environmental biology of fishes, 39: 259–270.

Leblois, R., Estoup, A., and Rousset, F. 2003. Influence of mutational and sampling factors on the estimation of demographic parameters in a "continuous" population under isolation by distance. Molecular Biology and Evolution, 20: 491–502.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends in Ecology and Evolution, 17: 183–189.

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25: 1754–1760.

Limborg, M. T., Helyar, S. J., De Bruyn, M., Taylor, M. I., Nielsen, E. E., Ogden, R., Carvalho, G. R. *et al.* 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). Molecular Ecology, 21: 3686–3703.

Lischer, H. E. L., and Excoffier, L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics, 28: 298–299.

Lotterhos, K. E., and Whitlock, M. C. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology, 24: 1031–1046.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., and Storfer, A. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular Ecology Resources, 17: 142–152.

Luckhurst, B. E. 2014. Elements of the ecology and movement patterns of highly migratory fish species of interest to ICCAT in the Sargasso Sea. Collective Volume of Scientific Papers ICCAT, 70: 2183–2206.

Luckhurst, B. E., Trott, T., and Manuel, S. 2001. Landings, Seasonality, Catch per Unit Effort, and Tag-Recapture Results of Yellowfin Tuna and Blackfin Tuna at Bermuda. pp. 225–234. American Fisheries Society, Bethesda, MD.

Maghan, B. W., and Rivas, L. R. 1971. The blackfin tuna (*Thunnus atlanticus*) as an underutilized fishery resource in the tropical western Atlantic Ocean. FAO Fisheries Report, 71: 163–172.

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W.-M. 2010. Robust relationship inference in genome-wide association studies. Bioinformatics, 26: 2867–2873.

Marçais, G., and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics, 27: 764–770.

Mariani, P., Křivan, V., MacKenzie, B. R., and Mullon, C. 2016. The migration game in habitat network: the case of tuna. Theoretical Ecology, 9: 219–232.

Mathieu, H., Pau, C., and Reynal, L. 2013. Chapter 2.1.10.7 thon a Nageoires Noires. *In* ICCAT Manual. International Commis-

sion for the Conservation of Atlantic Tuna. Updated 2016. [Cited 005/18/2022]. ISBN (Electronic Edition): 978-92-990055-0-7. ICCAT Publications, Madrid.

Miller, J. M., Cullingham, C. I., and Peery, R. M. 2020. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. Heredity, 125: 269–280.

Nielsen, E. E., Hemmer-Hansen, J., Poulsen, N. A., Loeschcke, V., Moen, T., Johansen, T., Mittelholzer, C. *et al.* 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). BMC Evolutionary Biology, 9: 276.

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., and Portnoy, D. S. 2018. These aren't the loci you'e looking for: principles of effective SNP filtering for molecular ecologists. Molecular Ecology, 27: 3193–3206,

Olazcuaga, L., Loiseau, A., Parrinello, H., Paris, M., Fraimout, A., Guedot, C., Diepenbrock, L. M. *et al.* 2020. A whole-genome scan for association with invasion success in the fruit fly Drosophila suzukii using contrasts of allele frequencies corrected for population structure. Molecular Biology and Evolution, 37(8): 2369–2385.

Pante, E., and Simon-Bouhet, B. 2013. marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. PLoS ONE, 8: e73051.

Paradis, E., and Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35: 526–528.

Peakall, R., Ruibal, M., and Lindenmayer, D. B. 2003. Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat *Rattus fuscipes*, Evolution, 57: 1182–1195.

Peakall, R., and Smouse, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics, 28: 2537–2539.

Pecoraro, C., Babbucci, M., Franch, R., Rico, C., Papetti, C., Chassot, E., Bodin, N. *et al.* 2018. The population genomics of yellowfin tuna (*Thunnus albacares*) at global geographic scale challenges current stock delineation. Scientific Reports, 8: 13890.

Pecoraro, C., Zudaire, I., Bodin, N., Murua, H., Taconet, P., Díaz-Jaimes, P., Cariani, A. *et al.* 2016. Putting all the pieces together: integrating current knowledge of the biology, ecology, fisheries status, stock structure and management of yellowfin tuna (*Thunnus albacares*). Reviews in Fish Biology and Fisheries, 27: 1–31.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE, 7: e37135.

Pruett, C. L., Saillant, E., and Gold, J. R. 2005. Historical population demography of red snapper (*Lutjanus campechanus*) from the northern Gulf of Mexico based on analysis of sequences of mitochondrial DNA. Marine Biology, 147: 593–602.

Pujolar, J. M., Jacobsen, M. W., Als, T. D., Frydenberg, J., Munch, K., Jónsson, B., Jian, J. B. *et al.* 2014. Genome-wide single-generation signatures of local selection in the panmictic European eel. Molecular Ecology, 23: 2514–2528.

Puritz, J. B., Hollenbeck, C. M., and Gold, J. R. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ, 2: e431.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. Computer software. R Foundation for Statistical Computing, Vienna.

Raj, A., Stephens, M., and Pritchard, J. K. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics, 197: 573–589.

Reglero, P., Santos, M., Balbín, R., Laíz-Carrión, R., Alvarez-Berastegui, D., Ciannelli, L., Jiménez, E. *et al.* 2017. Environmental and biological characteristics of Atlantic bluefin tuna and albacore spawning habitats based on their egg distributions. Deep Sea Research Part II: Topical Studies in Oceanography, 140: 105–116.

Richardson, D. E., Llopiz, J. K., Guigand, C. M., and Cowen, R. K. 2010. Larval assemblages of large and medium-sized pelagic species in the Straits of Florida. Progress in Oceanography, 86: 8–20.

Roach, M. J., Schmidt, S. A., and Borneman, A. R. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics, 19: 460.

Robledo-Arnuncio, J. J., and Rousset, F. 2010. Isolation by distance in a continuous population under stochastic demographic fluctuations. Journal of Evolutionary Biology, 23: 53–71.

Rousset, F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources, 8: 103–106.

Saillant, E. A., Luque, P. L., Short, E., Antoni, L., Reynal, L., Pau, C., Arocha, F. *et al.* 2022. Population structure of blackfin tuna (*Thunnus atlanticus*) in the western Atlantic Ocean inferred from microsatellite loci. Scientific Reports, 12: 9830.

Saxton, B. L. 2009. Historical Demography and Genetic Population Structure of the Blackfin Tuna (Thunnus atlanticus) from the Northwest Atlantic Ocean and the Gulf of Mexico. Texas A&M University, College Station, TX.

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., and Wolf, J. B. W. 2016. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution, 8: 907–917.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31: 3210–3212.

Singh-Renton, S., and Renton, J. 2007. Cframp's large pelagic fish tagging program. Gulf and Caribbean Research, 19: 99–102.

Smouse, P. E., Long, J. C., and Sokal, R. R. 1986. Multiple regression and correlation extensions of the mantel test of matrix correspondence. Systematic Zoology, 35: 627.

Smouse, P. E., and Peakall, R. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity, 82(Pt 5): 561–573.

Therkildsen, N. O., and Palumbi, S. R. 2017. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Molecular Ecology Resources, 17: 194–208.

van der Valk, T., Vezzi, F., Ormestad, M., Dalen, L., and Guschanski, K. 2020. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. Molecular Ecology Resources; 20(5): 1171–1181.

Vaux, F., Bohn, S., Hyde, J. R., and O'Malley, K. G. 2021. Adaptive markers distinguish North and South Pacific Albacore amid low population differentiation. Evolutionary Applications 14, 1343–1364.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A. *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE, 9: e112963.

Whitlock, M. C., and Lotterhos, K. E. 2015. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F(ST). The American Naturalist, 186: S24–S36.

Ye, C., Hill, C. M., Wu, S., Ruan, J., and Ma, Z. S. 2016. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Scientific Reports, 6: 31900.

Ye, C., Ma, Z. S., Cannon, C. H., Pop, M., and Yu, D. W. 2012. Exploiting sparseness in de novo genome assembly. BMC Bioinformatics, 13: S1.

Handling Editor: Kristi Miller-Saunders