

DR. GREGORY NEILS NEILS PUNCHER (Orcid ID : 0000-0002-4372-8435)

DR. ALESSIA CARIANI (Orcid ID : 0000-0002-5513-0086)

Article type : Resource Article

Title: Spatial dynamics and mixing of bluefin tuna in the Atlantic Ocean and Mediterranean Sea revealed using next generation sequencing.

Authors: GREGORY N. PUNCHER^{1,2,3*}, ALESSIA CARIANI^{1*}, GREGORY E. MAES^{4,5,6}, JEROEN VAN HOUDT⁵, KOEN HERTEN^{5,6}, RITA CANNAS⁷, NAIARA RODRIGUEZ-EZPELETA⁸, AITOR ALBAINA^{9,10}, M. ANDONE ESTONBA⁹, MOLLY LUTCAVAGE¹¹, ALEX HANKE¹², JAY ROOKER^{13,14}, JAMES S. FRANKS¹⁵, JOSEPH M. QUATTRO¹⁶, GUALTIERO BASILONE¹⁷, IGARATZA FRAILE⁸, URTZI LACONCHA^{8,9}, NICOLAS GOÑI⁸, AI KIMOTO¹⁸, A. DAVID MACÍAS¹⁹, FRANCISCO ALEMANY¹⁹, SIMEON DEGUARA²⁰, SALEM W. ZGOZI²¹, FULVIO GARIBALDI²², ISIK K. ORAY²³, F. SAADET KARAKULAK²³, NOUREDDINE ABID²⁴, MIGUEL N. SANTOS²⁵, PIERO ADDIS⁷, HARITZ ARRIZABALAGA⁸, FAUSTO TINTI¹

Author affiliations:

- 1. Dept. of Biological, Geological and Environmental Sciences / Laboratory of Genetics and Genomics of Marine Resources and Environment (GenoDREAM), University of Bologna, Via S. Alberto 163, 48123 Ravenna, Italy*
- 2. Dept. of Biology, Marine Biology Research Group, Ghent University, Krijgslaan 281, Campus Sterre - S8, B-9000, Ghent, Belgium*
- 3. Dept. of Biology, University of New Brunswick, 100 Tucker Park Road, E2L 4L5, Saint John, NB, Canada*
- 4. Centre for Sustainable Tropical Fisheries and Aquaculture, Comparative Genomics Centre, College of Science and Engineering, James Cook University, Townsville, Australia*
- 5. Centre for Human Genetics, Genomics Core, KU Leuven - UZ Leuven, Herestraat 49, 3000 Leuven, Belgium*
- 6. Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven (KU Leuven), B-3000 Leuven, Belgium*

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1755-0998.12764

This article is protected by copyright. All rights reserved.

7. *Department of Life & Environmental Sciences (DISVA), University of Cagliari, via Fiorelli 1, 09126 Cagliari, Italy*
8. *AZTI Tecnalia, Marine Research Division, Herrera Kaia, Portualdea z/g 20110, Pasaia, Gipuzkoa, Spain*
9. *Laboratory of Genetics Faculty of Science & Technology, Dpt. Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Barrio Sarriena s/n, 48940 Leioa, Spain*
10. *Environmental Studies Centre (CEA), Vitoria-Gasteiz City Council, Pintor Teodoro Dublang 25, 1008, Vitoria-Gasteiz, Spain*
11. *School for the Environment and Large Pelagics Research Center, University of Massachusetts, Boston, PO Box 3188, Gloucester, MA 01931*
12. *Department of Fisheries and Oceans, 531 Brandy Cove Road, St. Andrews, E5B 2L9, New Brunswick, Canada*
13. *Department of Marine Biology, Texas A&M University at Galveston, 1001 Texas Clipper RD, Galveston, Texas 77554*
14. *Department of Wildlife and Fisheries Sciences, Texas A&M University, 2258 TAMU, College Station, Texas 77843*
15. *Gulf Coast Research Laboratory, Center for Fisheries Research and Development, University of Southern Mississippi, 703 East Beach Drive, Ocean Springs, MS 39564, USA*
16. *Dept. of Biological Sciences, University of South Carolina, Columbia, South Carolina, USA, 29208*
17. *National Research Council, Institute for Marine and Coastal Environment, Detached Unit of Capo Granitola, Torretta Granitola 91021, Trapani, Italy*
18. *National Research Institute of Far Seas Fisheries, 5-7-1 Orido Shimizu, Shizuoka 424-8633, Japan*
19. *Instituto Español de Oceanografía, Centro Oceanográfico de Baleares, Muelle de Poniente, s/n, Apartado de Correos 291, 07015Palma, Spain*
20. *Federation of Maltese Aquaculture Producers (FMAP) 61, St. Paul Street, VLT 1212 Valletta, Malta*
21. *Marine Biology Research Center, P.O. Box 30830, Tripoli-Tajura, Libya*

22. Dept. of Earth, Environmental and Life Sciences, University of Genoa, DISTAV - C. so Europa, 26 - 16132 Genova, Italy

23. Faculty of Fisheries, Istanbul University, Ordu St., No.200, 34470, Laleli-Istanbul, Turkey

24. National Institute of Fisheries Research, Regional Centre of Tangier, PO 5268, Morocco

25. Instituto Português do Mar e da Atmosfera, Av. 5 de Outubro, 8700-305 Olhão, Portugal

* These authors contributed equally.

Correspondence: Gregory N. Puncher, University of New Brunswick, 100 Tucker Park Drive, Saint John, NB E2L 4L5, Canada, Tel: 506-648-5985, Email: gregory.puncher@unb.ca

Keywords: Single Nucleotide Polymorphisms, *Thunnus thynnus*, population structure, Reduced Representation Sequencing, mixed stock analysis, origin assignment

Running title: Mixed stock analysis of Atlantic bluefin tuna.

Abstract:

The Atlantic bluefin tuna is a highly migratory species emblematic of the challenges associated with shared fisheries management. In an effort to resolve the species' stock dynamics, a genome-wide search for spatially informative single nucleotide polymorphisms (SNPs) was undertaken, by way of sequencing reduced representation libraries. An allele frequency approach to SNP discovery was used, combining the data of 555 larvae and young-of-the-year (LYOY) into pools representing major geographical areas and mapping against a newly assembled genomic reference. From a set of 184,895 candidate loci, 384 were selected for validation using 167 LYOY. A highly discriminatory genotyping panel of 95 SNPs was ultimately developed by selecting loci with the most pronounced differences between western Atlantic and Mediterranean Sea LYOY. The panel was evaluated by genotyping a different set of LYOY (n= 326) and from these 77.8% and 82.1% were correctly assigned to western Atlantic and Mediterranean Sea origins, respectively. The panel revealed temporally persistent differentiation among LYOY from the western Atlantic and Mediterranean Sea ($F_{ST} = 0.008$, $p=0.034$). The composition of six

Accepted Article

mixed feeding aggregations in the Atlantic Ocean and Mediterranean Sea was characterized using genotypes from medium (n = 184) and large (n = 48) adults, applying population assignment and mixture analyses. The results provide evidence of persistent population structuring across broad geographic areas and extensive mixing in the Atlantic Ocean, particularly in the mid-Atlantic Bight and Gulf of St. Lawrence. The genomic reference and genotyping tools presented here constitute novel resources useful for future research and conservation efforts.

1. Introduction

Characterization of wide and small-scale population connectivity patterns of large fish species with extensive migrations remains one of the key challenges preceding the definition of reliable conservation and management units. The International Commission for the Conservation of Atlantic Tunas (ICCAT) manages the Atlantic bluefin tuna (BFT; *Thunnus thynnus*) as two stocks, each composed of a single population with the following spatial dynamics: a western group that spawns in the Gulf of Mexico and forages in the north western Atlantic Ocean and an eastern group that spawns in the Mediterranean Sea and forages in the Mediterranean and north eastern Atlantic Ocean (Fromentin & Powers 2005). Several tagging and otolith studies have demonstrated that these two stocks mix extensively and form large feeding aggregations in the Atlantic Ocean, particularly in the vicinity of the Mid-Atlantic Bight (Block *et al.* 2001, 2005; Wilson *et al.* 2011). This evidence, combined with results from various otolith chemistry and genetic studies, suggests that the population structure of BFT is much more complex than the current two-stock management model (Rooker *et al.* 2008a, b, 2014; Galuardi *et al.* 2010; Aranda *et al.* 2013; Riccioni *et al.* 2010, 2013; Fraile *et al.* 2014; Cermeño *et al.* 2015). The

Accepted Article
resolution of the species' population structure and mixed-stock analysis is a high priority for managers and an overdue improvement for the suboptimal assessment models currently employed (ICCAT 2011). It is estimated that the eastern stock is much more productive (10x) than the western stock and as such, the risk of over-exploitation of western BFT in mixed fisheries is much higher (ICCAT 2015).

Genetic tools used for population structure studies are able to infer the spatial dynamics, migratory movements and kinship of commercially targeted species in order to delineate management units (Ovenden *et al.* 2015). These details are essential for accurate estimations of individual stock contributions to mixed fisheries, which are particularly vulnerable to mismanagement due to a tendency for over-exploitation of less productive stocks (Habicht *et al.* 2010, Taylor *et al.* 2011, Ying *et al.* 2011). In the past 15 years, several molecular techniques have been employed in an effort to gain a better understanding of the population structure and spatial dynamics of BFT (see ICCAT 2013 for a summary).

Sequences from the mitochondrial DNA control region (mtDNA CR) provided the first evidence of differentiation between the western Atlantic and Mediterranean stocks (Alvarado Bremer *et al.* 1999); although, the effectiveness of this marker was called into question by a later study that used samples collected during multiple seasons (Ely *et al.* 2002). Several studies published since then have provided a variety of estimates of differentiation of the two stocks using allozymes (Pujolar *et al.* 2003), mtDNA CR sequences and microsatellites (Carlsson *et al.* 2007; Boustany *et al.* 2008); while other studies have not found any significant differentiation (Alvarado Bremer *et al.* 2005). Compared with the estimated mean and median F_{ST} values of 57 marine fish species (0.062 and 0.020, respectively; Waples 1998), the aforementioned estimates

of differentiation for the eastern and western stocks of BFT have been much lower (fixation indices ranging from 0.005 to 0.012).

Several studies have provided evidence of varying levels of differentiation between groups of BFT within the Mediterranean Sea across a range of geographic scales using mtDNA CR sequences and microsatellites (Carlsson *et al.* 2004; Carlsson *et al.* 2007; Riccioni *et al.* 2010, 2013). Conversely, several other studies have been unable to detect any genetic structure in the Mediterranean Sea using mtDNA CR sequences and microsatellites (Boustany *et al.* 2008; Viñas *et al.* 2011; Vella *et al.* 2016), EST-derived microsatellites (Riccioni *et al.* 2017) and single nucleotide polymorphisms (SNPs; Antoniou *et al.* 2017).

In addition to the molecular markers selected, several important factors need to be considered when investigating the genetic structure of long-lived migratory fishes with ocean-wide distributions. Comparisons between BFT captured in adult feeding aggregations are likely to generate inaccurate and inconsistent allele frequencies, as they are likely to contain a varying proportion of individuals from the eastern and western stocks. As such, investigations into the genetic population structure of BFT should be limited to larvae and/or small young-of-the-year (YOY) captured in spawning areas and include samples from multiple years in order to avoid false positives for divergence due to sampling bias, temporal variation in the survival of larvae, and sweepstake effects (Larson & Julian 1999; Maes *et al.* 2006).

The goal of this study was to develop a panel of SNPs capable of distinguishing BFT larvae and YOY (LYOY) collected in spawning and nursery areas in the western Atlantic Ocean and Mediterranean Sea using reduced representation sequencing (RRS) of DNA libraries. SNPs are commonly used for population genomic analyses and molecular ecology studies due to their high abundance in any given genome, transferability between labs, high throughput potential and

low genotyping error (Elshire *et al.* 2011; Hauser *et al.* 2011; Hess *et al.* 2011; Seeb *et al.* 2011; Kelley *et al.* 2016). They have been used to investigate marine fish population structure (Albaina *et al.* 2013; Diopere *et al.* 2017; Jackson *et al.* 2014; Milano *et al.* 2013; Wang *et al.* 2016) and seafood traceability (Stokstad 2010; Nielsen *et al.* 2012); providing the scientific premises for the monitoring of migratory stocks and further development of DNA-based forensics in fisheries law enforcement. The developed panel of SNPs was then used to differentiate medium and large size adults of eastern and western origin captured at several feeding aggregations in the Atlantic Ocean and Mediterranean Sea. This study is the most geographically comprehensive genetic analysis of BFT to date, with samples coming from as far away as Libya, Turkey, the Gulf of Mexico and the Gulf of St. Lawrence.

2. Materials and Methods

2.1 SNP discovery and validation

2.1.1 Tissue sampling and DNA extractions

A total of 105 larvae and 450 YOY of BFT were captured at four spawning areas in the Mediterranean Sea (Balearic Sea, Tyrrhenian Sea, Strait of Sicily and Levantine Sea) and western Atlantic Ocean (Gulf of Mexico spawning site and adjacent waters off Cape Hatteras) between 2007 and 2012 during the late spring to early autumn months (Figure 1, Table 1). During 2009 and 2012, samples were provided by multiple suppliers in the Gulf of Mexico (2009), Tyrrhenian Sea (2012), and Levantine Sea (2012). Larvae were collected using plankton nets and stored in 96% ethanol, according to standard larval survey practices. In addition to morphological identification, all larvae were identified to species level using molecular techniques described in Puncher *et al.* (2015). YOY were captured with hand and trolling lines. The size range (straight fork length) of BFT in the YOY samples (Min. = 12 cm, Max. = 55cm,

Avg. = 31.3 ± 6.5 cm; Table S1) corresponds with an estimated age of 2 weeks - 12 months (Avg. = 3 months; Diaz and Turner 2007; Restrepo *et al.* 2010). All research was coordinated through the Atlantic-wide Research Programme for Bluefin Tuna” (GBYP) for which ICCAT issued a recommendation allowing the parties involved in this research to collect and sacrifice young BFT for the purposes of genetic research as well as ship samples from one country to another (Cert. No. ICCAT RMA12-049; ICCAT Circulars 2296-12, 2180-14, 3203-15).

Approximately 20 mg of muscle tissue or fin-clip from each of the YOY and complete or partial (either caudal fin or eyeball) larvae was digested overnight in a proteinase K solution. Genomic DNA was extracted from all tissue and purified using Promega’s Wizard®SV96 Genomic DNA Purification kit and vacuum manifold following the manufacturer’s protocols. DNA was eluted with 60µL of distilled water pre-heated to 60°C.

2.1.2 Restriction site associated DNA sequencing

Extracted DNA was first digested with the *ApeKI* restriction enzyme (RE), which has a 5bp recognition site (5'GCWGC - 3'CGWCG), followed by ligation of standard Illumina sequencing adaptors with RE sequencing overhangs (Elshire *et al.* 2011). Libraries were sequenced at the Genomics Core Facility (KULeuven) using a HiSeq2500 (Illumina) 100 bp paired-end module, multiplexing 188 individuals on each sequencing lane with barcodes developed within the project (Table S2). In order to achieve a sufficient depth of sequencing for reliable genotype calling 316 individuals were re-sequenced (274 of which had fewer than 500K reads). After re-sequencing, 90% of all individuals had >500K reads and the average total read count for all individuals was $1,673,390 \pm 1,407,838$. The average read count of re-sequenced individuals increased from 253,139 to 2,006,368 (Table S2).

2.1.3 SNP discovery

Barcode and low-quality edges were removed from the raw Illumina DNA sequence data and *de novo* SNP calling was initially performed using the Universal Network Enabled Analysis Kit pipeline (UNEAK, Lu et al. 2013). Among the 324,433 loci analyzed, SNPs that were not recognized among more than 70% of all individuals within each sample were removed. Allele frequencies among loci and individuals and genotype coverage were then analyzed using the Adegnet package in R (Jombart & Ahmed 2011) and Genetix (Belkhir et al. 2004) and an excess of heterozygosity was observed for the majority of the remaining loci. In case this phenomenon was a consequence of low individual read coverage, we performed additional allele frequency analysis using individuals with more than 2 million read counts each. The *de novo* SNP discovery pipeline was then abandoned for reasons outlined in the supplementary materials (Text S1).

In response to the aforementioned challenge, a genomic reference was produced by assembling data generated by genome sequencing of a single individual captured offshore from the Balearic Islands (NCBI BioProject PRJNA432036). The genomic library was sequenced in 2 × 75-bp paired-end mode on a HiSeq 2000 (Illumina). A total of 94,000,207 paired-end reads were produced and were quality trimmed to 71bp. ABySS (version 1.3.7; Simpson *et al.* 2009) was used to create a *de novo* assembly with standard parameters and an optimal k-mer size of 31. This resulted in an assembly of 6,309,067 contigs and a genome size of 944 Mb. As quality control for the assembled genomic contigs, the original data was mapped using Bowtie2 (version 2.2.0; Langmead & Salzberg 2012) in the local modus. Summary statistics from the *de novo* assembly of the genomic data are provided in Tables S3 and S4.

Barcodes and adaptors were then trimmed from the raw paired-end sequence data of all 555 individuals (allowing for one sequencing error in the barcodes and enzyme restriction site) and sequences de-multiplexed using GBSX (Herten et al. 2015). Overlapping read pairs of each individual in the dataset were then merged using FLASH (version 1.2.7; Magoč & Salzberg 2011) with default parameters, producing longer sequences (180bp), which resulted in much improved mapping quality and data confidence (Table S5, overall retention of 88.9% of reads). The data for all 555 individuals was then pooled according to the main geographic sampling areas, four from the Mediterranean (Balearic Sea, Tyrrhenian Sea, Strait of Sicily and Levantine Sea) and one from the western Atlantic Ocean (merging Gulf of Mexico and Cape Hatteras; Table 1). The five sequence data pools were mapped against the genomic reference using the end-to-end modus in Bowtie2 (version 2.2.0; Langmead & Salzberg 2012).

SNP calling was performed simultaneously for the five pools in one run using FreeBayes (version 9.9.10; Garrison & Marth 2012), based solely on merged single mapping and similar settings to the human genome project (mapping quality >20 and base quality >15; <http://www.genome.gov/10000923>). Reads that were mapped more than once and/or discordantly to the genomic reference were discarded. Results were analyzed using pairwise comparisons of allele read counts per locus. Several studies have demonstrated that this PoolSeq approach is a reliable method for SNP discovery, which offers faithful estimates of population allele frequencies and low false positive rates (Futschik & Schlötterer 2010; Zhu *et al.* 2012; Ozerov *et al.* 2013a). The genomic coordinates and corresponding coverage statistics for the targeted regions were retrieved using BEDtools (version 2.17.0; Quinlan & Hall 2010). Depth of coverage for each location in the genomic reference was normalized for each sample (all individuals captured at a sampling location during the same survey) using the following formula:

$$\text{Normalized depth location} = \frac{\text{Location depth} \times \text{Highest total read number across all samples}}{\text{Total number of reads within sample}}$$

Locations from the genomic reference with a normalized depth of ≥ 50 reads, among all samples, were selected for SNP identification. By performing SNP calling simultaneously on the mapped data of all samples, SNP information for all candidate positions in all samples were obtained, including positions that were homozygous for certain samples.

2.1.4 SNP selection and validation

A minimum read count of 18 per locus and pooled geographic sample was used as a threshold for confidence in SNP identity. Variants identified *in silico* were first selected for validation based upon minimum coverage and Delta value selection as described in Ozerov *et al.* (2013b; $\Delta = [pA_i - pA_j]$, where pA_i and pA_j are the frequencies of allele A in the i^{th} and j^{th} populations). A panel of 384 loci was then identified based upon compliance with BiomarkTM HD Fluidigm array design criteria (minimum flanking sequence length of 60 bp and $< 65\%$ GC content). The panel was validated using a subset of the 555 individuals initially processed for SNP discovery, selecting 167 BFT representative of the five geographical sampling areas (western Atlantic Ocean, Balearic Sea, Tyrrhenian Sea, Strait of Sicily and Levantine Sea; Table 1). SNP-typing was carried out with the BiomarkTM HD platform (Fluidigm) and run with ROX as the passive reference dye and FAM and HEX SNPtypeTM assays for specific allele calling. Individuals with low genotyping coverage ($< 70\%$ SNPs genotyped) and SNP loci with an excess of missing data ($> 10\%$) and with minor allele frequencies of less than 0.01 among all samples were removed from the dataset. In order to select SNPs optimal for discrimination of eastern and western spawning groups, as well as for within-Mediterranean discrimination, a reduced number of candidate SNPs was then identified with BELS (Bromaghin 2008), which uses a backward

elimination procedure that excludes each locus from the dataset in order of increasing impact on differentiation. As a performance measure “maximize mean mixture-composition estimation accuracy” was selected, population proportions were random, mixture sample size was 200 and each simulation used 250 replications. A final panel of 96 SNPs was selected and compared with sequences contained in the NCBI database using the blastn algorithm in BLAST (Altschul *et al.* 1990).

2.2 Assignment to origin and mixed stock analysis.

2.2.1 Collection, DNA extractions and SNP-typing of additional LYOY and adult samples

The efficacy of the selected 96 SNP panel for stock assignment analysis was then assessed by genotyping a second group of LYOY from the western (n=40) and eastern spawning areas (n=292; Table 1). DNA was extracted from 232 adult BFT from two different size classes (medium-adults: 25-100kg; large-adults: >100kg; Table 1) and 6 geographic locations throughout the species' range (Figure 1) and genotyped using the methods outlined above. Combined with the samples used for SNP validation (n=167), the total number of individuals genotyped using the 96 SNP panel was 731 (Figure 1; Table 1).

2.2.2 Genetic diversity and outlier detection

Kinship analysis was conducted for each temporal and spatial sample of LYOY. Relatedness was evaluated using ML-RELATE (Kalinowski *et al.* 2006), which assigns each pair of individuals to a range of relationships (parent-offspring “PO”, full-sibling “FS”, half-sibling “HS”, and unrelated “U”), the most likely of which was determined by comparing a putative relationship (the highest likelihood value) with an alternative relationship (the second highest likelihood value) using 1,000 simulations. Similar sibship inference was performed with COLONY v. 2.0.5.8 (Jones and Wang 2010) using the full likelihood method for a medium

length run with medium likelihood precision. Genotyping error rates were set at 0.01 per SNP site with no allelic dropout and polygamy among non-inbreeding males and females was assumed. To be classified as belonging to one of the three relatedness categories (unrelated, half-sibs, and full-sibs), dyads had to have $\geq 50\%$ probability of belonging to a given category.

After relatedness assessment, BFT captured in the Gulf of Mexico between 2007 and 2009 were pooled for all calculations, due to the limited number of larvae in each sample. Values of expected and observed heterozygosity and the inbreeding coefficient (F_{IS} ; Weir & Cockerham 1984) were calculated for each sample using Arlequin v3.5.2 (Excoffier & Lischer 2010), allowing for 10% missing data. Departure from Hardy–Weinberg equilibrium (HWE) was assessed for each locus in each group as well as across the entire dataset using the exact test implemented in Genepop v.4.2 (Raymond & Rousset 1995; Rousset 2008) using 1000K Markov chain steps and 100K dememorization steps. Linkage disequilibrium (LD) tests were conducted with Genepop using default values (100 batches of 1,000 iterations each). The false discovery rate (B-H) approach (Benjamini & Hochberg 1995) was used to analyze the resulting LD p-values for false positives using the SGoF+ program (Carvajal-Rodriguez & de Uña-Alvarez 2011).

Outlier loci were identified using the Beaumont and Nichols (1996) FDIST approach, implemented in Lositan (Antao *et al.* 2008) with a 0.995 confidence interval, 50K simulations and a false discovery rate of 0.1. The hierarchical Bayesian method described in Beaumont and Balding (2004), implemented in Bayescan software (Foll & Gaggiotti 2008) was also used to detect loci with strong support for selection ($-\log_{10}PO > 1$) using default settings (5K iterations, 50K burn-in length, 20 pilot runs of 5000).

2.2.3 Stock discrimination

Pairwise comparisons of F_{ST} values between all samples were executed using FSTAT v.2.9.4 (Goudet 1995) and between western and eastern pooled LYOY groups with Arlequin v.3.5.2, testing for significance with an adjusted nominal level of $\alpha=0.05$ and 10K permutations. The performance of the 96 SNP panel, to differentiate western and eastern LYOY pooled groups, was once again analyzed with BELS, using the same settings described above. Genetic heterogeneity and differentiation of LYOY was estimated in Arlequin, using several analyses of molecular variance (AMOVA) in order to examine multiple levels of variance in the dataset. The AMOVAs were performed with 10K permutations and a significance level of 0.05. The following three groupings were explored by assigning samples to various *a priori* groups: 1) western Atlantic Ocean (Gulf of Mexico and Cape Hatteras) vs. Mediterranean Sea, 2) temporal replicates pooled by geographic region (see Table 1), and 3) western Mediterranean Sea (Balearic Sea) vs. central western Mediterranean (Tyrrhenian Sea and Ligurian Sea) vs. central Mediterranean (Strait of Sicily and Malta) vs. eastern Mediterranean Sea (Levantine Sea). Additional analyses were performed using genotypes from two subsets of SNPs selected for trans-Atlantic and intra-Mediterranean differentiation, respectively.

Genetic admixture and stock discrimination was further investigated using STRUCTURE v2.3.4 (Pritchard *et al.* 2000), which employs a Bayesian model-based individual clustering method. STRUCTURE provides an option to use geographic sampling locations as prior information to assist with the clustering of individuals in datasets that have genuine population structure, but the signal is too weak for standard structure models to detect (Hubisz *et al.* 2009). The LOCPRIOR model is suitable when one is assuming that the sampling locations may be informative about ancestry, as is the case with LYOY captured close to their origin.

STRUCTURE analysis was conducted with the following settings: for each simulation of K (1-10), five independent replicates were run, with a 100K burn-in period and 500K Markov Chain Monte Carlo reps after burnin, Admixture Model, LOCPRIOR using the geographical regions defined in Table 1 and lambda inferred for each run. A plot of the mean likelihood L(K) and variance per K value from each of the STRUCTURE analyses was obtained using Structure Harvester (Earl & von Holdt 2012) and used to infer the optimal K value or number of ancestral clusters.

2.2.5 Stock assignment and mixed-stock analysis

A hold-out group (*sensu* Anderson 2010) of 332 LYOY not used for SNP selection or validation was genotyped (section 2.2.1) and assigned to eastern or western origin using both a leave-one-out self-assignment approach as well as using a reference dataset (training set *sensu* Anderson 2010) composed of genotypes from the 167 LYOY used for SNP selection and validation (Table 1). The GeneClass2 software (Piry *et al.* 2004) was employed using the Rannala and Mountain (1997) criterion (0.05 threshold). Individual results are reported as assignment scores, which are the likelihood of belonging to x population divided by the sum of likelihoods of belonging to each of the stocks considered. Individuals with an assignment score <70% were considered “poorly assigned” and are included in the results to demonstrate the proportional contribution of individuals falling within this criteria from each stock. A quality index was also calculated, reflecting the mean value of the scores of each individual in the population it belongs to. The probability of assignment to spawning area of origin was calculated with the Monte Carlo re-sampling probability computation using the Paetkau *et al.* (2004) simulation algorithm and 10,000 simulated individuals in GeneClass2, using the same query and LYOY reference datasets.

Medium and large BFT adults from six feeding aggregations in the Atlantic Ocean and Mediterranean Sea (N = 230; 2 individuals were removed due to low genotyping success) were then assigned to western Atlantic and Mediterranean Sea spawning areas with GeneClass2, using the complete LYOY dataset (N=493) and the Rannala and Mountain (1997) criterion (0.05 threshold). Overall probability of assignment to either spawning area was calculated using the same settings described above and the complete LYOY dataset.

The 100% simulations analysis tool in ONCOR was used to examine the potential accuracy of the mixture analysis using the sample sizes of the LYOY dataset, 100 simulations and 200 fish sampled from a virtual fishery. Mixture analysis was then performed on each adult feeding aggregation using the conditional maximum likelihood method (Millar 1987) used by ONCOR (Kalinowski *et al.* 2007) to estimate mixture proportions. Both the LYOY and adult datasets were re-sampled with 20,000 bootstraps in order to calculate confidence interval values.

2.3 Estimates of effective population size

The effective population size (N_e) of the eastern and western pooled groups was calculated using NeEstimator v. 2.0 (Do *et al.* 2014). A series of five N_e calculations was conducted using equal sample sizes by randomly selecting 80 individuals from the eastern pooled group using the *sample* function in R. All tests were run using the linkage disequilibrium method (Waples & Do 2008), random mating model and without any allele frequency restrictions.

3. Results

3.1 SNP discovery and validation

Statistics related to the assembly of the genomic contigs used as a reference for mapping and variant calling are featured in Tables S3 and S4. After implementation of the mapping and

normalized depth filtering criteria, SNP calling resulted in 2,467,103 SNP positions, of which 236,332 intersected with the >50X coverage mapped locations. After filtering for only bi-allelic variants (among the five geographical pools), a total of 184,895 SNPs was found to be shared among all five geographic sampling areas (minimum 50x coverage across all pools with fragments present with at least 10x coverage in each pool). When selecting variants for SNP assays, preference was given to loci that had Delta values between $1.0 < 0.7$ (high; $n = 64$), $0.7 < 0.6$ (mid; $n = 99$), and $0.6 < 0.5$ (low; $n = 221$) between the five geographical pools.

Technical validation of the 384 variants identified *in silico* (SNP flanking sequences available at <https://doi.org/10.5061/dryad.2f7v9>) had a 75% conversion rate to functional Fluidigm genotyping assays. Further assessment of the remaining 288 loci using genotype data from 167 individuals revealed 77 candidate SNPs with low polymorphism ($MAF < 0.01$) across the entire dataset or a high frequency of missing data ($> 10\%$). Although these values were likely to change with an increase in sample size, these 77 loci were removed in favour of SNPs that are consistently informative. This left a set of 211 SNPs for which the average percentage of missing data per SNP locus was 2.57% (genotype matrix available at <https://doi.org/10.5061/dryad.2f7v9>). According to BELS results, a total of 59 SNPs was selected for discrimination of LYOY from the western Atlantic Ocean and the Mediterranean Sea and another 48 SNPs for discrimination of sampling locations within the Mediterranean (11 of which had already been selected for trans-Atlantic discrimination). In total, 27 loci out of the selected 96 were found to share $> 80\%$ BLAST similarity scores with previously published gene sequences. Annotation details for the 96 SNPs selected for the final genotyping panel are provided in Table S6.

3.2 Assignment to origin and mixed stock analysis.

3.2.1 Genetic diversity and outlier detection

The selected 96 SNP panel provided high genotyping success rate among most of the 731 individuals (99%) genotyped. Only eight individuals were removed from the dataset due to low genotyping coverage (Table 1; genotype matrix available at <https://doi.org/10.5061/dryad.2f7v9>). A single locus was removed from analysis, due to low polymorphism throughout the entire dataset, resulting in a final set of 95 loci (58 candidate markers for trans-Atlantic discrimination and 48 for intra-Mediterranean differentiation).

Kinship analysis of LYOY suggested a very low proportion of related pairs (above the half-sibling level) among samples (Table S7). In fact, the majority of dyads were identified as UNRELATED (range 91.4%-100%), with only a few dyads classified as full and or half siblings.

Observed and expected heterozygosity among LYOY ranged from 0.291 to 0.343 and from 0.316 to 0.346 per sample, respectively (Table S8). Observed F_{IS} values were low and not significant, agreeing with results from the aforementioned kinship analysis. None of the 95 SNPs significantly deviated from the HWE. After applying the false discovery rate to all 3,160 tests there was no evidence of linkage disequilibrium among loci. Moreover, each of the selected SNP loci was contained in separate contigs and as such, physical linkage can also be excluded. When comparing eastern and western LYOY pooled groups, outlier analysis (using Lositan) revealed that three loci (SNP_161, SNP_158, SNP_010) are candidates for positive selection ($F_{ST} = 0.122, 0.082$ and 0.076 , respectively). A BLAST of the sequences flanking each of these SNPs provided a single match (SNP_158) with anaphase promoting complex subunit 2 sequences of several other teleosts (90% nucleotides matched in a ~145 bp sequence; Table S6).

3.3 Genetic differentiation among LYOY samples

Based on pairwise comparisons using the 95 SNP panel, there was no differentiation among samples collected from different years at the same sampling location (Table S9). For this reason, multi-year geographic samples were combined. BFT caught during different sampling seasons in Cape Hatteras (YOY) and the Gulf of Mexico (larvae), including larvae collected in the Gulf of Mexico in 2010 (WAGM10V), which were not used for SNP panel selection, showed no genetic differentiation, providing further support to pool temporal and geographical samples in this region. Within the Mediterranean Sea, the only pairwise comparison that resulted in significant differentiation was between Malta 2013 and the Levantine Sea 2011 ($F_{ST} = 0.010$; Table S9). After pooling temporal replicates by geographic area, western Atlantic Ocean samples significantly differed from all Mediterranean Sea samples ($F_{ST} = 0.007-0.012$, $p \leq 0.002$; Table 2). BELS analysis revealed that only 24 loci were required for optimal differentiation of these two groups of samples (Table S6). Analyses of pairwise comparisons using 58 and 24 SNP panels showed consistent differentiation between western and eastern BFT ($F_{ST} = 0.012-0.024$, $p < 0.002$; $F_{ST} = 0.032-0.040$, $p < 0.002$, respectively). Pairwise comparisons using the 95 and 48 SNP panels detected no temporally stable significant differences between Mediterranean Sea samples (data not shown).

Among the different pooled groups analysed using AMOVAs, the only scenario that provided significant statistical support was that of separate Mediterranean Sea and Gulf of Mexico populations (Tables S10 and S11). Pooled groups of LYOY from these two locations were significantly different using panels of 95, 58 and 24 loci ($F_{ST} = 0.008$, $p = 0.034$; $F_{ST} = 0.014$, $p = 0.035$ and $F_{ST} = 0.034$, $p < 0.001$, respectively). Differentiation of eastern and western pooled groups remained significant when assessing only LYOY that were not used for the SNP

panel development and validation (N=326, after removal of 6 individuals with low genotyping coverage), using these same panels ($F_{ST} = 0.004$, $p = 0.009$; $F_{ST} = 0.008$, $p < 0.001$, $F_{ST} = 0.024$, $p < 0.001$, respectively).

Bayesian cluster analysis suggested that modern BFT stocks are descendants of two distinct ancestral populations ($K=2$), as currently represented by the samples collected in the western Atlantic Ocean/Gulf of Mexico and Mediterranean Sea (Figure 2).

3.4 Population assignment and admixture analysis

3.4.1 Larvae and young-of-the-year

Among the LYOH used for SNP selection, 75.0% from the western Atlantic ($n=33/44$) and 77.2% from the Mediterranean ($n=95/123$) were correctly assigned to origin (Table S12). A total of 11.3% and 14.6% of individuals captured in the western Atlantic and Mediterranean, respectively, were considered poorly assigned (assignment scores $< 70\%$; Table S12).

Assignment of LYOH not included in the SNP discovery, validation and selection processes (hold-out group *sensu* Anderson 2010; $N=326$) demonstrated consistent efficacy of the panel. However, this hold-out dataset did yield a lower quality index of 70.9%, with 74.8% of individuals correctly assigned to origin. Probability computation with Monte-Carlo re-sampling assigned 83.4% of individuals from this hold-out group to correct origin with an average probability of assignment of 82.9%. Overall 66.7% and 71.7% of individuals captured in western and eastern spawning areas were correctly assigned to origin with average assignment scores of 89.6% and 87.1%, respectively (Table S12). A total of 19.4% and 20.3% of individuals assigned to the western and eastern stocks were considered poorly assigned (Table S12). Similar results were obtained when assigning individuals in the hold-out group to their origin using the reference dataset ($n=167$; training set *sensu* Anderson 2010; Table S13). A total of 72.7% of

individuals in the hold-out group was assigned to origin, while 14.4% was considered poorly assigned.

An analysis of the SNP panel's ability to assign LY0Y to origin using all LY0Y individuals (N=493) was also conducted; however, in an effort to avoid high grading bias, summary statistics reported in Table 3 are limited to individuals not used for SNP selection (N=326; refer to Table S14 for assignment scores of all LY0Y samples). Among these individuals, 66.7% of western and 72.4% eastern LY0Y were correctly assigned to origin with average assignment scores of 86.4% and 89.0%, respectively (Table 3). The proportion of poorly assigned individuals collected from the western Atlantic Ocean (22.2%) was higher than those collected in the Mediterranean Sea (17.3%; Figure 3; Table 3).

3.4.2 Adults

The average probability of assignment of all adult individuals (N = 230) to spawning areas of origin was estimated at 86.8%. Average assignment scores for adults to the western Atlantic Ocean and Mediterranean Sea were 75.2% and 86.6%, respectively, with an overall average of 84.5% (Table 4; Figure S3). The proportion of individuals assigned to either western or eastern spawning areas was calculated and plotted on a map illustrating geographic distribution of mixing rates (Figure 4; Table 4).

An evaluation of the power of assignment revealed that the actual sample sizes from the western Atlantic Ocean (n = 80) and Mediterranean Sea (n = 413) included in the LY0Y dataset and used for genetic stock identification of adults were likely to provide $87.9 \pm 6.4\%$ and $98.7 \pm 1.1\%$ accuracy (Table S15). Mixture analysis of feeding aggregations provided similar results (Table S16) as the individual assignment of adults; however, the 95% confidence intervals provided by the ONCOR analysis are extremely wide and should be interpreted with caution, as

these measures are sensitive to unbalanced sample sizes. For example, the proportion of individuals assigned to western origin is much reduced, in comparison to other methods, at locations where the western fraction is low. This is likely due to the lower number of individuals from the western Atlantic in the LYOY dataset.

Both methods used to analyse the composition of adult feeding aggregations revealed that the central Atlantic Ocean, Gulf of St. Lawrence and coastal Morocco are areas hosting the highest amount of mixing between stocks (Table 4). Conversely, the samples from Portugal are entirely of Mediterranean Sea origin. The contribution of adult BFT from each stock to the feeding aggregation in Morocco appears to vary slightly from one year to the next.

3.5 Estimates of effective population size

The effective population size estimated for the pooled western Atlantic group (3437.2, confidence interval = 469.6 - ∞) was lower than the N_e estimated for the pooled eastern Atlantic group (4682.2, confidence interval = 1840.7 - ∞). Results from tests using equal sample sizes ($n=80$) were so variable (Table S17) that no conclusions can be drawn from their consideration.

4. Discussion

4.1 De novo SNP discovery

The difficulties that were encountered during our initial SNP discovery efforts using a de novo approach were not unique. Several studies have shown that *ApeKI* library preparations, coupled with Illumina sequencing, results in a low, biased and error prone coverage of genome sequences (De Donato et al. 2013; Lu et al. 2013). The restriction site targeted by *ApeKI* (GC[A/T]GC) is relatively short and contains a degenerate base, allowing frequent fragmentation of the genome. Short fragments can easily be incorrectly aligned, resulting in paralogs that appear as heterozygous loci. Moreover, this can result in a high number of missing data in post-

sequencing datasets, irrespective of the quality of template DNA (De Donato et al. 2013). Using these methods researchers can expect to have ~30% missing data for any given individual in a dataset, even with high depth coverage (Beissenger et al. 2013). For this reason, many recent GBS studies have used alternative restriction enzymes with longer base pair cut sites (*Pst*I, *Eco*T22I) for studies of wild non-model species (Alcaide et al. 2014; De Kort et al. 2014; Li et al. 2014). A more in depth discussion about challenges associated with the *de novo* SNP discovery methods that were used is featured in the supplementary materials (Text S1).

4.2 SNP genotyping

Since we chose to pool all sequence data according to geographical origin of individuals, the focus on allele frequencies was shifted to overall read counts. This is not a new approach and has been advised for tasks requiring reduced representation library construction and less ambitious sequencing efforts (Gautier *et al.* 2013). Pooling of genotype data from individuals sequenced using NGS platforms can provide accurate estimates of population allele frequencies with similar accuracy to individual-based analyses; however, it also reduces capacity for detection of rare alleles (Cutler & Jensen 2010).

4.3 Population assignment and mixture analysis

Kinship analysis indicated that the relations among LY0Y collected from each location were sufficiently distant so as not to cause concern for duplication of genotypes within the dataset. The sampling techniques used (sampling on multiple dates at different locations) avoided the collection of patches of closely related individuals which can confuse population structure patterns and lead to under-estimation of genetic diversity.

Juvenile bluefin tuna of greater than 12 months age (>53cm) have been observed to cross the Atlantic Ocean (Cort & Liorzou 1995; Rooker *et al.* 2008b; Fraile *et al.* 2014). In this study,

we have used younger YOY as genetic references to avoid the inclusion of migrants in our spawning location reference dataset (Table S1). For example, the age of the YOY captured in Cape Hatteras was approximately 2-5 months (Min. = 12cm; Max. = 37cm; Avg. = 29.6cm; Diaz & Turner 2007). It is unlikely that these young fish have crossed the Atlantic Ocean, since large fast-swimming adult BFT require 3-4 months to complete the journey (Mather 1980; Block *et al.* 2001; Wilson *et al.* 2011).

The range of F_{ST} values that we have calculated (0.009 – 0.034) based on the separation of the eastern and western LYOY samples are similar to those calculated by a previous study that used mtDNA CR sequences ($\Phi_{ST} = 0.013$) and microsatellites ($F_{ST} = 0.006$; Carlsson *et al.* 2007). However, it must be noted that the SNPs used to calculate F_{ST} values were selected based on their capacity to differentiate the two stocks. The first effort to characterize BFT populations in the western and eastern Atlantic Ocean using SNPs produced the most significant evidence for differentiation of populations to date (Albaina *et al.* 2013). In that study, 17 SNPs revealed significant differentiation of YOY captured in the western Atlantic Ocean and western Mediterranean Sea ($F_{ST} = 0.116 \pm 0.078$, $p < 0.01$) as well as juveniles from the Bay of Biscay ($F_{ST} = 0.120 \pm 0.091$, $p < 0.01$). These F_{ST} estimates are an order of magnitude higher than previous estimates using other molecular approaches. Due to the limited number of western Atlantic individuals used in their study (N=15), the results should be verified with an increased sample size. Moreover, at least 80 SNPs are recommended for the detection of low levels of differentiation among highly mobile species (Hauser *et al.* 2011; Ryman *et al.* 2006).

Variability of F_{ST} values among locations sampled over multiple years, throughout the entire dataset, is indicative of temporal genetic variability among cohorts. This suggests a reconsideration of the conclusions reached by previous studies that have used single-season

sample collections and supports the argument that multi-year sampling is a necessity for studies investigating the population structure of BFT as well as other migratory marine fishes.

Considering the highly migratory nature of juvenile and adult BFT, the weak signal of differentiation (low yet statistically significant F_{ST} values) between the two stocks is not surprising. Only a handful of individuals (<1%) migrating between spawning areas each generation can eliminate most genetic evidence of stock structure (Waples 1998). The F_{ST} values generated here are relatively low when compared to other taxa; however, these values are consistent with what is to be expected from a highly migratory marine species with a larval phase and recently generated populations (Ward *et al.* 1994; Alvarado Bremer *et al.* 2005).

The broad distribution of individual assignment scores that we calculated can be expected from highly variable candidate source populations (see Tables S10 and S11 for within population variation; Ogden and Linacre 2015). In order to improve the assignment of juvenile and adult BFT to Gulf of Mexico and Mediterranean Sea spawning areas, additional LYOY are required from the Gulf of Mexico in order to account for this variation.

4.4 Assignment power and confidence

The extent to which individual LYOY were successfully assigned to origin (79.3% for all LYOY, 77.3% for the hold-out group, considering all assignment scores) is similar to the results of studies using various properties of otoliths for assignment. Rooker *et al.* (2003) were able to assign YOY to eastern and western origins with an 85% success rate, using trace elements (Li, Mg, Ca, Mn, Sr and Ba) in otoliths. Using stable isotopes ($\delta^{13}C$ and $\delta^{18}O$), Schloesser *et al.* (2010) and Rooker *et al.* (2014) were able to assign YOY from eastern and western nurseries to their origin with overall success rates of 84% and 83%. With the same markers, Hanke *et al.* (2016) also encountered an overall assignment success rate of 83% when assigning adult BFTs

caught in the Northwest Atlantic Ocean from 1975-1977 and 2011-2013. Similarly, Brophy *et al.* (2015) were able to distinguish between eastern and western BFT using otolith shape, once again with an 83% success rate. The results of these five studies combined suggest the existence of an unknown source of error that accounts for a 16-20% misclassification rate.

While genetic and otolith analyses provide similar assignment success rates, it is important to note that only genetic analysis is effective for the identification of butchered market tuna with heads removed. Moreover, the collection of tissues suitable for genetic analysis (scales, fin clips, spines, muscle) is much easier, non-invasive and doesn't require the killing of fish. Genetic assignment results are informative at an evolutionary scale while other methods (eg. morphology) provide information that is mainly informative at an ecological time scale and on each individual's pre/post reproductive movements. Multi data source information should be ultimately combined or integrated rather than compared.

In comparison to population assignment analysis, results from the admixture analysis appear to underestimate the proportion of individuals from western spawning/nursery areas. Given the much higher productivity of the eastern stock and the tendency of maximum likelihood estimations to eclipse estimates of less productive stocks by more dominant ones (Millar 1990; Fraile *et al.* 2014), we discuss the results from both population assignment and admixture analyses.

4.5 Evidence of mixing

Both tagging and otolith chemistry indicate that stocks mix extensively in the Atlantic Ocean, particularly around the mid-Atlantic Bight, the central north Atlantic and off the coasts of Morocco and western Europe (Block *et al.* 2001, 2005; Rooker *et al.* 2008b, 2014; Galuardi *et al.* 2010; Wilson *et al.* 2011; Fraile *et al.* 2014). The extent of these oceanic movements appears

to be dependent upon BFT origin and ontogeny (Mather 1980; Block *et al.* 2005). Although adults are capable of crossing the Atlantic in 3-4 months (Block *et al.* 2001; Wilson *et al.* 2011, 2015), significant numbers of juvenile BFT (>3 up to 44kg) also make the trip (Mather 1980; Rooker *et al.* 2008b, 2014; Graves *et al.* 2015). According to various tagging studies, BFT from the east are more likely to cross the Atlantic than their counterparts in the west (Block *et al.* 2005; Taylor *et al.* 2011). Our results are corroborated by previously published synopses stating that stock mixing is highest in the western Atlantic Ocean, with the eastern stock subsidizing the western stock by 10-50% (Taylor *et al.* 2011).

Evidence of BFT reproduction taking place in a third, previously undocumented spawning area in the Slope Sea (western Atlantic Ocean) has been provided by multiple sources (Mather *et al.* 1974; Baglin 1976; Mather *et al.* 1995; Goldstein *et al.*, 2007; Heinisch *et al.*, 2014); however, the extent to which this occurs remains unknown. Recently, Richardson *et al.* (2016) provided strong evidence of this third spawning area by collecting larvae in the western Slope Sea. Combined with the presence of reproductively active medium-sized BFT (120-220cm) in the same area, this finding supports a scenario of size-structured spawning migrations in the western Atlantic spawning groups, as suggested by life history modelling (Chapman *et al.* 2011). In such a scenario, medium-sized BFT may feed and spawn in the Slope Sea, while large adults from the eastern and western stocks have a wider spatial distribution that includes feeding areas in the North Atlantic Ocean and spawning areas in the Mediterranean Sea and Gulf of Mexico, respectively. Moreover, BFT captured in the Mediterranean Sea and identified as western in origin may come from the Slope Sea and not just from the Gulf of Mexico. This might account for the relatively weak levels of differentiation between the currently recognized western and eastern stocks. These revelations may provide an explanation for the aforementioned 16-20%

of individuals that cannot be assigned to either spawning group using otolith characteristics, or the techniques described herein. Moreover, lower assignment scores in some individuals may be an indication that gene flow is occurring in the Slope Sea.

4.5.1 North western Atlantic Ocean

The size fraction of YOY BFT collected from Cape Hatteras (fork length = 29.6 ± 5.7 cm, total weight = 0.5 ± 0.2 kg) was likely too small to contain any of the juvenile trans-Atlantic migrants documented by various studies (Mather 1980; Rooker *et al.* 2008a,b). Based on our results, the adult fishery in the Gulf of St. Lawrence appears to contain a significant fraction of eastern migrants (>50%). A recent study, using otolith microchemistry assignment techniques, uncovered similarly high numbers of eastern migrants in the Gulf of St. Lawrence (26%), trans-Atlantic Bight (35%) and off the coast of Newfoundland (41%) and Nova Scotia (22%; Hanke *et al.* 2016 and Table S18). Although proportions may vary with time and sampling sites, estimates of the annual contribution of eastern BFT to the catch show a large increase in each of these regions in 2014 (Table S18). Rooker *et al.* (2014) has also estimated the proportion of eastern migrants in the Gulf of St. Lawrence mixed stock to be as high as $44.0 \pm 16.8\%$. Conversely, an earlier analysis of archived BFT otoliths caught in the same area suggested that eastern migrants were all but absent from the Gulf of St. Lawrence during 1975-1984 and 2006-2007 (Schloesser *et al.* 2010).

Catches of BFT in the Gulf of St. Lawrence indicate a steady increase in abundance since 2010 (Vanderlaan *et al.* 2014) with a growing proportion of the catch composed of smaller fish. A tendency for fish size to be inversely related to the probability of eastern origin has been demonstrated (Hanke *et al.* 2016). This putative growing proportion of eastern migrants in the

Gulf of St. Lawrence could be an indication of a poleward shift (MacKenzie *et al.* 2014; Dell *et al.* 2015; Robinson *et al.* 2015) of eastern migrants in the western Atlantic Ocean.

4.5.2 Central North Atlantic

Genotyping of samples collected in the central North Atlantic Ocean revealed that the majority of large (83.0% overall; 59.0% when individuals with assignment scores of <70% are excluded) and medium-sized (88.0% overall; 67.0% when individuals with assignment scores of <70% are excluded) adults were of eastern origin. Using fish caught in the same area and season, Brophy *et al.* (2015) estimated that the mixed fishery contained approximately $94 \pm 7\%$ BFT of eastern origin. An analysis of otoliths from 202 adults captured in the same area in 2010-2011 concluded that eastern migrants constitute between $79.5 \pm 6.5\%$ (based on maximum likelihood estimations) and $90.7 \pm 5.3\%$ (2011 alone) of the mixed stock in that region (Rooker *et al.* 2014).

4.5.3 North eastern Atlantic and Mediterranean Sea

Similar to some otolith-based studies (Rooker *et al.* 2008b, 2014) we have also identified a small number of western-type individuals among samples collected in the Mediterranean Sea. These individuals could be incorrectly assigned, due to the low number of individuals contained in the Western Atlantic LYOY reference pool (N=80; Tables 3 and S13-15). Alternatively, they could be an indication of previously undocumented mixing within spawning areas, as suggested by the two western-type adults caught in the Gulf of Sirte (Libya). The highest number of western migrants among the adult individuals collected in the eastern Atlantic was captured off the coast of the Canary Islands (21%) and Morocco (16-32%). Previously, it has been estimated that $6.1 \pm 4.7\%$ to $22.6 \pm 7.0\%$ of BFT captured off the coast of Morocco are of western origin (Rooker *et al.* 2014; Brophy *et al.* 2015).

4.6 Inter-annual variation

The results from the population assignment and mixture analysis of adult samples suggest that the contribution of each stock to feeding aggregations varies over time and across age classes. A previous genetic study determined that the mixed fishery in the North Atlantic Ocean, south of Iceland, contains BFT of significant genetic and morphological heterogeneity (Carlsson *et al.* 2006). Distinct groups of BFT appear to enter North Atlantic waters at different times throughout the fishing season. Our results may provide an origin for these different cohorts of tuna. Rooker *et al.* (2014) observed considerable inter-annual variability in the proportion of western BFT in the central North Atlantic Ocean, ranging from $36.1 \pm 9.6\%$ in 2010 to $9.3 \pm 5.3\%$ in 2011. Inability to take these spatio-temporal patterns of stock variation into consideration for policy design and management imperils the welfare of the species as well as the sustainable exploitation of both stocks (Taylor *et al.* 2011). Other fisheries that target Baltic Sea cod (Hüssy *et al.* 2013; Eero *et al.* 2014) and various salmon stocks on the west coast of North America (Hess *et al.* 2014; Larson *et al.* 2014) already consider this information when assigning temporal and regional quotas. This would require that the routine monitoring of BFT fisheries by way of catch estimates and morphological measurements should also include aspects of genetic and otolith analyses.

4.7 Effective population size

Our estimates of N_e for the eastern pooled group (4682, CI = 1841 - ∞) was slightly higher than other previously published short term N_e estimates for the Adriatic Sea and Tyrrhenian Sea which were calculated with pseudo-likelihood (752, 95% CI = 429–1,853; 326, 95% CI = 193–695, respectively) and likelihood-based (682, 95% CI: 363–1575; 405, 95% CI = 221–974, respectively) methods using the mutation rate of eight microsatellites (Riccioni *et al.*

2010). Our estimates are similar to values calculated for another closely related tuna species occupying the same ecological niche in the North Atlantic Ocean, the albacore tuna, which were generated using a panel of 75 SNPs (5466 - 23330; Laconcha et al. 2015).

Caution should be taken when interpreting our estimations of N_e , however, given the apparent admixture of the two groups and the assumption of closed populations when estimating N_e using the linkage disequilibrium method (Waples & England 2011). Moreover, the inconsistency of the N_e estimates and large confidence intervals generated by subsets of 80 individuals from the eastern pooled group suggests that a higher sample size is required to generate accurate estimates.

5. Conclusion

Developing genomic resources for Atlantic Bluefin tuna is of crucial importance for the continued modernization of techniques used for management of the species as well as other tuna species that are equally threatened by over-exploitation. The genomic reference presented here is of use for the mapping and positioning of current and future genomic markers. Annotation and comparison of our reference against other tuna or teleost genomes will allow better functional characterization of tuna genes linked to conservation and relevant life-history traits. Genome-wide SNP data, and especially outlier loci, have broad applications in future studies, allowing for local analyses and the validation of population structure at micro-scale levels. Moreover, the development of a database containing SNP allele frequencies will allow for additional membership analyses of LY0Y captured at spawning locations and adults caught in mixed feeding aggregations. Finally, the various resources that we have developed can be used for seafood traceability and fisheries forensics applications (Bernatchez *et al.* 2017).

Acknowledgements

This work was carried out under the provision of the ICCAT Atlantic Wide Research Programme for Bluefin Tuna (GBYP), funded by the European Union, by several ICCAT CPCs, the ICCAT Secretariat and other entities (see: <http://www.iccat.int/GBYP/en/Budget.htm>). The contents of this paper do not necessarily reflect the point of view of ICCAT or of the other funders, which have no responsibility for the contents of this publication, neither do they necessarily reflect the views of the funders and in no way anticipates the Commission's future policy in this area. Additional funds were provided by the MARES Joint Doctorate programme selected under Erasmus Mundus and coordinated by Ghent University (FPA 2011-0016).

The Biogenomics LRD subdivision (<https://bio.kuleuven.be/eeb/lbeg/consulting.html>), Genomics Core Leuven (<https://www.genomicscore.be/>) and Ing. Bart Hellemans (KULeuven, Belgium) provided much appreciated assistance with NGS library preparation, sequencing and advanced bioinformatics. The authors thank the Sequencing and Genotyping unit of SGIker (UPV/EHU) for the technical and human support provided.

The following researchers were essential for the collection of samples: Dheeraj Busawon (Department of Fisheries and Oceans, Canada), Iñigo Onandia (AZTI Tecnalia, Spain), Osamu Sakai (National Research Institute of Far Seas Fisheries, Japan), John Lamkin (National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southeast Fisheries Science Center). The following Gulf Coast Research Laboratory researchers were essential for the collection and identification of larval bluefin tuna from the Gulf of Mexico: Jason Tilley, Bruce Comyns, Eric Hoffmayer, Mae Blake.

Larvae from the Balearic Sea were provided by the ATAME project "Bluefin tuna abundance indices: towards recruitment estimators based on larval ecology" (CTM 2011-29525-

C04-02) and the BLUEFIN project “Modelling bluefin spawning areas variability and population dynamics in the Western Mediterranean” (IEO-SOCIB agreement).

References

- Albaina A, Iriondo M, Velado I, *et al.* (2013) Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure. *Animal Genetics*, 44, 678-692.
- Alcaide M, Scordato ES, Price TD, Irwin DE (2014) Genomic divergence in a ring species complex. *Nature*, 511(7507), 83-85.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman, DJ (1990) "Basic local alignment search tool". *Journal of Molecular Biology*, 215, 403-410.
- Alvarado-Bremer JR, Naseri I, Ely B (1999) A provisional study of northern bluefin tuna populations. *ICCAT Collection of Scientific Papers*, 49,127–129.
- Alvarado Bremer JR, 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, 169-187.
- Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*,10, 701–710
- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, 9(1), 323.
- Antoniou A, Kasapidis P, Kotoulas G, Mylonas CC, Magoulas A (2017) Genetic diversity of Atlantic Bluefin tuna in the Mediterranean Sea: insights from genome-wide SNPs and microsatellites. *Journal of Biological Research-Thessaloniki*, 24(1), 3, DOI 10.1186/s40709-017-0062-2
- Aranda G, Abascal FJ, Varela JL, Medina A (2013) Spawning behaviour and post-spawning migration patterns of Atlantic bluefin tuna (*Thunnus thynnus*) ascertained from satellite archival tags. *PLoS ONE*, 8(10), e76445.
- Baglin RE (1976) A preliminary study of the gonadal development and fecundity of the western Atlantic bluefin tuna. *ICCAT Collection of Scientific Papers*, 5, 279–289.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, 263(1377), 1619-1626.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans, *Molecular ecology*, 13(4), 969-980.
- Beissinger TM, Hirsch CN, Sekhon RS, *et al.* (2013) Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics*, 193, 1073-1081.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 405, logiciel sous Windows TM pour la génétique des populations Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Society Series*, 57, 289-300.

- Bernatchez L, Wellenreuther M, Araneda C, *et al.* (2017) Harnessing the Power of Genomics to Secure the Future of Seafood. *Trends in Ecology & Evolution*, 32(9) 665-680, <https://doi.org/10.1016/j.tree.2017.06.010>.
- Block BA, Dewar H, Blackwell SB (2001) Migratory movements, depth preferences, and thermal biology of Atlantic bluefin tuna. *Science*, 293(5533), 1310-1314.
- Block BA, Teo SL, Walli A (2005) Electronic tagging and population structure of Atlantic bluefin tuna. *Nature*, 434(7037), 1121-1127.
- Boustany AM, Reeb CA, Block BA (2008) Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*). *Marine Biology*, 156, 13-24.
- Bromaghin JF (2008) BELS: backward elimination locus selection for studies of mixture composition or individual assignment. *Molecular Ecology Resources*, 8, 568-571.
- Brophy D, Haynes P, Arrizabalaga H, *et al.* (2015) Otolith shape variation provides a marker of stock origin for north Atlantic bluefin tuna (*Thunnus thynnus*). *Marine and Freshwater Research*, 67(7), 1023-1036, doi: 101071/MF15086.
- Carlsson J, McDowell JR, Diaz-Jaimes P, Carlsson JEL, Boles SB, Gold JR, Graves JE (2004) Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Molecular Ecology*, 13, 3345-3356.
- Carlsson J, McDowell JR, Carlsson JEL, Ólafsdóttir D, Graves JE (2006) Genetic heterogeneity of Atlantic bluefin tuna caught in the eastern North Atlantic Ocean south of Iceland. *ICES Journal of Marine Science*, 63, 1111-1117.
- Carlsson J, McDowell JR, Carlsson JE, Graves JE (2007) Genetic identity of YOY bluefin tuna from the eastern and western Atlantic spawning areas. *Journal of Heredity*, 98, 23-28.
- Carvajal-Rodriguez A, de Uña-Alvarez J (2011) Assessing Significance in High-Throughput Experiments by Sequential Goodness of Fit and q-Value Estimation. *PLoS ONE*, 6(9), e24700.
- Chapman EW, Jørgensen C, Lutcavage ME (2011) Atlantic bluefin tuna (*Thunnus thynnus*): a state-dependent energy allocation model for growth, maturation, and reproductive investment. *Canadian Journal of Fisheries and Aquatic Sciences*, 68(11), 1934-1951.
- Cermeño P, Quílez-Badia G, Ospina-Alvarez A (2015) Electronic tagging of Atlantic bluefin tuna (*Thunnus thynnus*, L) reveals habitat use and behaviors in the Mediterranean Sea. *PLoS ONE*, 10(2), e0116638.
- Cort JL, Liorzou B (1995) Revision del mercado/recaptura de atun rojo (*Thunnus thynnus*) en el Atlántico Este y Mediterráneo. *ICCAT Collection of Scientific Papers*, 44, 293-304.
- Cutler, DJ, Jensen, JD (2010) To pool, or not to pool? *Genetics*, 186(1), 41-43.
- De Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin, IG (2013) Genotyping-by-Sequencing (GBS): A Novel, Efficient and Cost-Effective Genotyping Method for Cattle Using Next-Generation Sequencing. *PLoS ONE*, 8(5), e62137 doi:101371/journal.pone0062137.
- De Kort H, Mergeay J, Vander Mijnsbrugge K, *et al.* (2014) An evaluation of seed zone delineation using phenotypic and population genomic data on black alder *Alnus glutinosa*. *Journal of Applied Ecology*, 51(5), 1218-1227.
- Dell JT, Wilcox C, Matear RJ, Chamberlain MA, Hobday AJ (2015) Potential impacts of climate change on the distribution of longline catches of yellowfin tuna (*Thunnus albacares*) in the Tasman sea. *Deep Sea Research Pt II*, 113, 235-245.

- Diaz GA, Turner SC (2007) Size frequency distribution analysis, age composition, and maturity of western bluefin tuna in the Gulf of Mexico from the US (1981–2005) and Japanese (1975–1981) longline fleets. *ICCAT Collection of Scientific Papers*, 6, 1160-1170.
- Diopere E, Vandamme SG, Hablützel PI, *et al.* (2017) Seascape genetics of a flatfish reveals local selection under high levels of gene flow, *ICES Journal of Marine Science*, <https://doi.org/10.1093/icesjms/fsx160>.
- Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14(1), 209-214.
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359-361.
- Eero M, Hemmer-Hansen J, Hüsey K (2014) Implications of stock recovery for a neighbouring management unit: experience from the Baltic cod. *ICES Journal of Marine Science*, 71(6), 1458-1466.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), e19379.
- Ely B, Stoner DS, Bremer AJ, Dean JM, Addis P, Cau A, Quattro JM (2002) Analyses of nuclear *ldhA* gene and mtDNA control region sequences of Atlantic northern bluefin tuna populations. *Marine Biotechnology*, 4, 583-588.
- Excoffier L, Lischer HEL (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.
- Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977-993.
- Fraile I, Arrizabalaga H, Rooker JR (2014) Origin of Atlantic bluefin tuna (*Thunnus thynnus*) in the Bay of Biscay. *ICES Journal of Marine Science*, 101093/icesjms/fsu156.
- Fromentin JM and Powers JE (2005) Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. *Fish and Fisheries*, 6, 281-306
- Futschik A, Schlötterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, 186, 207–218.
- Galuardi B, Royer F, Golet W, Logan J, Neilson J, Lutcavage M (2010) Complex migration routes of Atlantic bluefin tuna (*Thunnus thynnus*) question current population structure paradigm. *Canadian Journal of Aquatic Sciences*, 67, 966-976.
- Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv*, 12073907.
- Gautier M, Foucaud J, Gharbi K, *et al.* (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology*, 22(14), 3766-3779.
- Goldstein J, Heppell S, Cooper A, Brault S, Lutcavage M (2007) Reproductive status and body condition of Atlantic bluefin tuna in the Gulf of Maine, 2000-2002. *Marine Biology*, 151, 2063–2075.
- Goudet J (1995) FSTAT (Version 12): A computer program to calculate F-statistics. *Journal of Heredity*, 86(6), 485-486.

- Graves JE, Wozniak AS, Dickhut RM, Cochran MA, MacDonald EH, Bush E, Arrizabalaga H, Goñi N (2015) Transatlantic movements of juvenile Atlantic bluefin tuna inferred from analyses of organochlorine tracers. *Canadian Journal of Aquatic Sciences*, 72, 625-633.
- Habicht C, Seeb LW, Myers KW, Farley EV, Seeb JE (2010) Summer–fall distribution of stocks of immature sockeye salmon in the Bering Sea as revealed by single-nucleotide polymorphisms. *Transactions of the American Fisheries Society*, 139(4), 1171-1191.
- Hanke A, Busawon D, Rooker JR, Secor DH (2016) Estimates of stock origin for bluefin tuna caught in the western Atlantic fisheries from 1975 to 2013. *ICCAT Collective Volume of Scientific Papers*, 72(6), 1376-1393.
- Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources*, 11(s1), 150-161.
- Heinisch G, Rosenfeld H, Knapp JM, Gordin H, Lutcavage ME (2014) Sexual maturity in western Atlantic bluefin tuna. *Scientific Reports*, 4, 7205.
- Herten K, Hestand MS, Vermeesch JR, Van Houdt JK (2015) GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics*, 16(1), 73.
- Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*, (Suppl 1), 137–149.
- Hess JE, Whiteaker JM, Fryer JK, Narum SR (2014) Monitoring stock-specific abundance, run timing, and straying of Chinook salmon in the Columbia River using genetic stock identification (GSI). *North American Journal of Fisheries Management*, 34(1), 184-201.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322-1332.
- Hüssy K, Bastardie F, Eero M, Hemmer-Hansen J, Mosegaard H, Nielsen JR (2013) Improved management based on stock identification of eastern and western Baltic cod. National Institute of Aquatic Resources, Technical University of Denmark.
- ICCAT (2011) Recommendation by ICCAT concerning the Atlantic-wide research programme for bluefin tuna (GBYP) ICCAT BFT 11-06.
- ICCAT (2013) Report of the 2013 bluefin meeting on biological parameters review (Tenerife, Spain, 07 May – 13 May, 2013) pp 1-75.
- ICCAT (2015) Report for biennial period, 2014-2015 Part I (2014) - Vol 1 Madrid, Spain.
- Jackson AM, Semmens BX, De Mitcheson YS, *et al.* (2014) Population structure and phylogeography in Nassau grouper (*Epinephelus striatus*), a mass-aggregating marine fish. *PLoS ONE*, 9(5), e97508.
- Jombart T, Ahmed I (2011) Adegnet 13-1: new tools for the analysis of genome-wide SNP data *Bioinformatics*, doi: 101093/bioinformatics/btr521.
- Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10, 551–555.
- Kalinowski ST, Wagner AP, Taper ML (2006) ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, 6, 576-579.
- Kalinowski ST, Manlove KR, Tape, ML (2007) ONCOR: a computer program for genetic stock identification Department of Biology, Montana State University 22 pp.

- Kelley JL, Brown AP, Overgaard Therkildsen N, Foote AD (2016) The life aquatic: advances in marine vertebrate genomics. *Nature Reviews Genetics*, 17, 523-534.
- Laconcha U, Iriando M, Arrizabalaga H, *et al.* (2015) New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*). *PLoS One*, 10(6), e0128247.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357-359.
- Larson RJ, Julian RM (1999) Spatial and temporal genetic patchiness in marine populations and their implications for fisheries management. *California Cooperative Oceanic Fisheries Investigations Report*, 94-99.
- Larson WA, Seeb JE, Pascal CE, Templin WD, Seeb LW (2014) Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. *Canadian Journal of Aquatic Sciences*, 71(5), 698-708.
- Li C, Waldbieser G, Bosworth B, Beck BH, Thongda W, Peatman E (2014) SNP discovery in wild and domesticated populations of blue catfish, *Ictalurus furcatus*, using genotyping-by-sequencing and subsequent SNP validation. *Molecular Ecology Resources*, 14(6), 1261-1270.
- Lu F, Lipk, AE, Glaubitz J, *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, 9(1), e1003215 doi:10.1371/journal.pgen.1003215.
- MacKenzie BR, Payne MR, Boje J, Høyer JL, Siegstad H (2014) A cascade of warming impacts brings bluefin tuna to Greenland waters. *Global Change Biology*, 20(8), 2484-2491.
- Maes GE, Pujolar JM, Hellemans B, Volckaert FA (2006) Evidence for isolation by time in the European eel (*Anguilla anguilla* L.). *Molecular ecology*, 15(8), 2095-2107.
- Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957-2963.
- Mather FJ, Mason JM, Jones AC (1974) Distribution, fisheries and life history data relevant to identification of Atlantic bluefin tuna stocks. *ICCAT Collective Volume of Scientific Papers*, 2, 234-258.
- Mather FJ (1980) A preliminary note on migratory tendencies and distributional patterns of Atlantic bluefin tuna based on recently acquired and cumulative tagging results. *ICCAT Collection of Scientific Papers*, 9(2), 478-491.
- Mather FJ, Mason JM, Jones AC (1995) Historical Document: Life History and Fisheries of Atlantic Bluefin Tuna (US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Miami). *NOAA Technical Memorandum NMFS-SEFSC 370*.
- Milano I, Babbucci M, Cariani A, *et al.* (2014), Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular ecology*, 23, 118-135. doi:10.1111/mec.12568
- Millar RB (1987) Maximum likelihood estimation of mixed stock fishery composition. *Canadian Journal of Aquatic Sciences*, 44, 583-590.
- Millar RB (1990) Comparison of methods for estimating mixed stock fishery composition. *Canadian Journal of Aquatic Sciences*, 47(11), 2235-2241.

- Minoche AE, Dohm JC, Himmelbauer H (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*, 12, R112.
- Nielsen EE, Cariani A, Mac Aoidh E, *et al.* (2012) Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications* 3: 851 doi: 101038/ncomms1845.
- Ogden R, Linacre A (2015) Wildlife forensic science: a review of genetic geographic origin assignment. *Forensic Science International: Genetics*, 18, 152-159.
- Ovenden JR, Berry O, Welch DJ, Buckworth RC, Dichmont CM (2015) Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish and Fisheries*, 16(1), 125-159.
- Ozerov M, Vasemägi A, Wennevik V, Niemelä E, Prusov S, Kent M, Vähä JP (2013a) Cost-effective genome-wide estimation of allele frequencies from pooled DNA in Atlantic salmon (*Salmo salar* L). *BMC Genomics*, 14(1), 12.
- Ozerov M, Vasemägi A, Wennevik V, *et al.* (2013b) Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. *PLoS ONE*, 8(12), e82434.
- Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, 13, 55–65.
- Piry, S, Alapetite, A, Cornuet, JM, Paetkau, D, Baudouin, L, Estoup, A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of heredity*, 95(6), 536-539.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.
- Pujolar JM, Roldán MI, Pla C (2003) Genetic analysis of tuna populations, *Thunnus thynnus thynnus* and *T alalunga*. *Marine Biology*, 143, 613-621.
- Puncher GN, Arrizabalaga H, Alemany F, *et al.* (2015) Molecular identification of Atlantic bluefin tuna (*Thunnus thynnus*, Scombridae) larvae and development of a genetic character-based identification key for Mediterranean scombrids. *PLoS ONE*, 10(7), e0130407.
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842.
- Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Science USA* 94, 9197–9201.
- Raymond M, Rousset F (1995) GENEPOP (version 12): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248-249 .
- Restrepo VR, Diaz GA, Walter JF, Neilson JD, Campana SE, Secor D, Wingate RL (2010) Updated estimate of the growth curve of western Atlantic bluefin tuna. *Aquat Living Resources*, 23(04), 335-342.
- Riccioni G, Landi M, Ferrara G, *et al.* (2010) Spatio-temporal population structuring and genetic diversity retention in depleted Atlantic bluefin tuna of the Mediterranean Sea. *Proceedings of the National Academy of Science*, 107(5), 2102-2107.
- Riccioni G, Stagioni M, Landi M, Ferrara G, Barbujani G, Tinti F (2013) Genetic structure of bluefin tuna in the Mediterranean Sea correlates with environmental variables. *PLoS ONE*, 8(11), e80105.

- Riccioni G, Cariani A, Ferrara G, Cannas R, Melis R, Stagioni M, Addis P, Tinti F (2017) Evolutionary constraints limiting the variation of Expressed Sequence Tag-linked microsatellite loci, prevent the detection of local adaptation in Mediterranean Bluefin tuna. *Fisheries Research*, 190, 157-163.
- Richardson DE, Marancik KE, Guyon JR, *et al.* (2016) Discovery of a spawning ground reveals diverse migration strategies in Atlantic bluefin tuna (*Thunnus thynnus*). *Proceedings of the National Academy of Science*, 113(12), 3299-3304.
- Robinson LM, Hobday AJ, Possingham HP, Richardson AJ (2015) Trailing edges projected to move faster than leading edges for large pelagic fish habitats under climate change. *Deep Sea Research Pt II*, 113, 225-234.
- Rooker JR, Secor DH, Zdanowicz VS, De Metrio G, Relini LO (2003) Identification of Atlantic bluefin tuna (*Thunnus thynnus*) stocks from putative nurseries using otolith chemistry. *Fisheries Oceanography*, 12(2), 75-84.
- Rooker JR, Secor DH, De Metrio G, Schloesser R, Block BA, Neilson JD (2008a) Natal homing and connectivity in Atlantic bluefin tuna populations. *Science*, 322, 742-744.
- Rooker JR, Secor DH, De Metrio G, Kaufman AJ, Ríos AB, Ticina V (2008b) Evidence of trans-Atlantic movement and natal homing of bluefin tuna from stable isotopes in otoliths. *Marine Ecology Progress Series*, 368, 231-239.
- Rooker JR, Arrizabalaga H, Fraile I, *et al.* (2014) Crossing the line: migratory and homing behaviors of Atlantic bluefin tuna. *Marine Ecology Progress Series*, 504, 265-276.
- Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, 8, 103-106.
- Ryman N, Palm S, Andre C, *et al.* (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology*, 15, 2031-2045.
- Schloesser RW, Neilson JD, Secor DH, Rooker JR (2010) Natal origin of Atlantic bluefin tuna (*Thunnus thynnus*) from Canadian waters based on otolith $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$. *Canadian Journal of Fisheries and Aquatic Sciences*, 67(3), 563-569.
- Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011) Single nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, 11(s1), 1-8.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Resources*, 19(6), 1117-1123.
- Stokstad E (2010) "To Fight Illegal Fishing, Forensic DNA Gets Local". *Science*, 330(6010), 1468-1469.
- Taylor NG, McAllister MK, Lawson GL, Carruthers T, Block BA (2011) Atlantic bluefin tuna: A novel multistock spatial Model for assessing population biomass. *PLoS ONE*, 6(12), e27693 doi:10.1371/journal.pone.0027693.
- Vella A, Vella N, Karakulak FS, Oray I, Garcia-Tiscar S, Stephanis R. (2016) Population genetics of Atlantic bluefin tuna, *Thunnus thynnus* (Linnaeus, 1758), in the Mediterranean: implications for its conservation management. *Journal of Applied Ichthyology*, 32(3):523-31.
- Vanderlaan AS, Hanke AR, Chassé J, Neilson JD (2014) Environmental influences on Atlantic bluefin tuna (*Thunnus thynnus*) catch per unit effort in the southern Gulf of St Lawrence. *Fisheries Oceanography*, 23(1), 83-100.
- Viñas, J, Gordo, A, Fernández-Cebrián, R, Pla, C, Vahdet, Ü, Araguas, RM (2011) Facts and uncertainties about the genetic population structure of Atlantic bluefin tuna (*Thunnus*

- thynnus*) in the Mediterranean Implications for fishery management. *Reviews in Fish Biology and Fisheries*, 21, 527-541.
- Wang J, Xue DX, Zhang BD, Li YL, Liu BJ, Liu JX (2016) Genome-wide SNP discovery, genotyping and their preliminary applications for population genetic inference in spotted sea bass (*Lateolabrax maculatus*). *PLoS ONE*, 11(6), e0157809.
- Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89, 438-450.
- Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, 3, 244–262.
- Waples RS, England PR (2011) Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics*, 189(2), 633-644.
- Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of fish biology*, 44(2), 213-232.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.
- Wilson SG, Lawson GL, Stokesbury MJW, *et al.* (2011) Movements of Atlantic bluefin tuna from the Gulf of St Lawrence to their spawning grounds. *ICCAT Collection of Scientific Journals*, 66(3), 1247-1256.
- Wilson SG, Jonsen ID, Schallert RJ, Ganong JE, Castleton MR, Spares AD, Boustany AM, Stokesbury MJ, Block BA (2015) Tracking the fidelity of Atlantic bluefin tuna released in Canadian waters to the Gulf of Mexico spawning grounds. *Canadian Journal of Fisheries and Aquatic Sciences*, 72, 1700-1717.
- Ying Y, Chen Y, Lin L, Gao T (2011) Risks of ignoring fish population spatial structure in fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences*, 68(12), 2101–2120 doi:10.1139/F2011-116.
- Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical validation of pooled whole genome population re-sequencing in *Drosophila melanogaster*. *PLoS ONE*, 7(7), e41901-e41901.

Data Accessibility

All sequencing data has been published at the National Center for Biotechnology Information (NCBI) as BioProject PRJNA432036, including the newly assembled genomic reference as well as the raw reads from the reduced representation library sequencing 555 individual Atlantic Bluefin tuna (BioSamples SAMN08408093 - SAMN08408647). SNP flanking regions and alleles, together with individual genotypes are available on the Dryad Digital Repository (DOI: <https://doi.org/10.5061/dryad.2f7v9>).

Author Contributions

Project development and design - Gregory N. Puncher, Alessia Cariani, Gregory E. Maes, Rita Cannas, Aitor Albaina, M. Andone Estonba, Haritz Arrizabalaga, Fausto Tinti.

Manuscript development - Gregory N. Puncher, Alessia Cariani, Gregory E. Maes, Jeroen Van Houdt, Koen Herten, Rita Cannas, Aitor Albaina, M. Andone Estonba, Molly Lutcavage, Alex Hanke, Jay Rooker, Ai Kimoto, Naiara Rodriguez-Ezpeleta, Haritz Arrizabalaga, Fausto Tinti.

Laboratory analysis - Gregory N. Puncher, Alessia Cariani, Gregory E. Maes, Jeroen Van Houdt, Koen Herten.

Data analysis - Gregory N. Puncher, Alessia Cariani, Gregory E. Maes, Jeroen Van Houdt, Koen Herten, Rita Cannas, Naiara Rodriguez-Ezpeleta, Fausto Tinti.

Data management - Gregory N. Puncher, Alessia Cariani, Igaratza Fraile, Urtzi Laconcha, Nicolas Goñi

Sample collection - Aitor Albaina, M. Andone Estonba, Molly Lutcavage, Alex Hanke, Jay Rooker, James S. Franks, Joseph M. Quattro, Gualtiero Basilone, Angela Cuttita, Igaratza Fraile, Urtzi Laconcha, Nicolas Goñi, Ai Kimoto, A. David Macías, Francisco Alemany, Simeon Deguara, Salem W. Zgozi, Fulvio Garibaldi, Isik K. Oray, F. Saadet Karakulak, Nouredine Abid, Miguel N. Santos, Piero Addis, Haritz Arrizabalaga, Fausto Tinti.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/1755-0998.12764](https://doi.org/10.1111/1755-0998.12764)

This article is protected by copyright. All rights reserved.

Table 1. Details of Atlantic bluefin tuna used in this study for: i) *in silico* variant discovery ii) validation of a 384 SNP genotyping panel and then selection of the 96 SNPs panel and, iii) large scale genotyping with a 96 SNP panel. Integers in parentheses in the final column represent individuals removed from analyses due to low genotyping success (<70% SNPs genotyped).

Sampling location	Area code	Age class	Year	Sample code	Sample sizes		
					SNP discovery	SNP validation & selection	96 SNP-typing
Western Atlantic Ocean							
Gulf of Mexico	GOM (WAtl)	Larvae (V)	2007	WAGM07V	11	2	
		Larvae (V)	2008	WAGM08V	14	8	
		Larvae (V)	2009	WAGM09bV	25	10	
		Larvae (V)	2009	WAGM09V	24	8	
		Larvae (V)	2010	WAGM10V			40 (4)
Cape Hatteras	CAP (WAtl)	YOY (0)	2008	WACH080	16	16	
Gulf of St. Lawrence	GSL	Large Adult (L)	2013	WAGSL13L			23 (1)
Central North Atlantic Ocean							
Central North Atlantic	CNAtl	Medium Adult (M)	2012	CACA12M			24
		Large Adult (L)	2012	CACA12L			47 (1)
Eastern Atlantic Ocean							
Canary Islands	CIS	Large Adult (L)	2013	EACI13L			24

Morocco	MRC	Large Adult (L)	2012	EAMO12L			24
		Large Adult (L)	2013	EAMO13L			19
Portugal	PTL	Large Adult (L)	2011	EAPO11L			23
Mediterranean Sea							
Balearic Sea	BAL	YOY (0)	2009	WMBA090	41		
		YOY (0)	2010	WMBA100	45		
		YOY (0)	2011	WMBA110	40	24	
		YOY (0)	2012	WMBA120	40		
		Larvae (V)	2012	WMBA12V			40
		YOY (0)	2013	WMBA130			34
Ligurian Sea	LIG	YOY (0)	2013	WMLI130			33
Tyrrhenian Sea	TYR	YOY (0)	2011	WMTY110	37		
		YOY (0)	2012	WMTY120	40	24	
		YOY (0)	2012	WMTY12b0	40		
		YOY (0)	2013	WMTY130			40
Malta	MLT	YOY (0)	2013	CMMA130			40
Strait of Sicily	STS	Larvae (V)	2008	CMSI08V			25 (2)
		YOY (0)	2011	CMSI110	21	21	
		Larvae (V)	2012	CMSI12V	21		
		YOY (0)	2012	CMSI120	40	27	

		YOY (0)	2013	CMSI130			40
Gulf of Sirte	SRT	Medium Adult (M)	2012	CMSY12M			24
		Large Adult (L)	2012	CMSY12L			24
Levantine Sea	LEV	Larvae (V)	2011	EMLS11V	10		
		YOY (0)	2011	EMLS110	29	27	
		YOY (0)	2012	EMLS120	40		
		YOY (0)	2012	EMLS12b0	21		
		YOY (0)	2013	EMLS130			40
				TOTAL	555	167	564 (8)

Table 2. Pair-wise F_{ST} values among Atlantic bluefin tuna larvae and young-of-the-year samples (N=493) pooled by geographic region using 95 SNPs. F_{ST} values are below and associated p-values above the diagonal. Significant pairwise comparisons are marked with an asterisk. Sample sizes, age class, year of sampling and area code definitions are detailed in Table 1.

	WAtl	BAL	LIG	TYR	MLT	STS	LEV
WAtl		$\leq 0.002^*$	$\leq 0.002^*$	$\leq 0.002^*$	$\leq 0.002^*$	$\leq 0.002^*$	$\leq 0.002^*$
BAL	0.010*		0.155	0.414	0.005	0.136	0.871
LIG	0.008*	0.001		0.512	0.064	0.279	0.433
TYR	0.007*	0.000	0.001		0.207	0.098	0.202
MLT	0.012*	0.004	0.004	0.001		0.024	0.076
STS	0.009*	0.001	0.001	0.001	0.004		0.455
LEV	0.010*	0.000	0.002	0.001	0.002	0.000	

Table 3. Assignment of Atlantic bluefin tuna larvae and young-of-the-year (N=326) not included in the SNP panel selection, captured in spawning and nursery areas in the Western Atlantic (WAtl) and Mediterranean Sea (Med) using genotypes from 95 SNPs. Individuals with an assignment score (A.S.) <70% were considered “poorly assigned” and data are reported separately. All values, aside from sample size (N) are reported as percentages. All calculations were made using GeneClass2. See Table 1 for sampling locations and sample codes.

Sample Location	Sample code	N	WAtl	WAtl	Avg. A.S.	Med	Med	Avg. A.S.
			A.S. >70%	A.S. <70%	to WAtl	A.S. >70%	A.S. <70%	to Med
Western Atlantic	WAGM10V	36	66.7	11.1	86.4	11.1	11.1	72.7
Balearic Sea	WMBA12V	40	7.5	10.0	69.9	72.5	10.0	90.9
Balearic Sea	WMBA130	34	8.8	8.8	68.2	73.5	8.8	92.4
Ligurian Sea	WMLI130	33	12.1	3.0	73.7	69.7	15.2	86.2
Tyrrhenian Sea	WMTY130	40	5.0	12.5	67.0	75.0	7.5	89.4
Malta	CMMA130	40	12.5	2.5	73.6	70.0	15.0	88.2
Strait of Sicily	CMSI08V	23	4.3	13.0	66.2	78.3	4.3	91.6
Strait of Sicily	CMSI130	40	17.5	7.5	76.6	67.5	7.5	88.9
Levantine Sea	EMLS130	40	12.5	5.0	78.3	75.0	7.5	85.4
Total WAtl		36	66.7	11.1	86.4	11.1	11.1	72.7
Total Med		290	10.3	7.6	72.2	72.4	9.7	89.0

Table 4. Assignment of adult Atlantic bluefin tuna (N=230), captured in various feeding aggregations, to natal origin in the Western Atlantic (WAtl) and Mediterranean Sea (Med) using genotypes from 95 SNPs. Individuals with an assignment score (A.S.) <70% were considered “poorly assigned” and data are reported separately. All values, aside from sample size (N) are reported as percentages. All calculations were made using GeneClass2. See Table 1 for sampling locations and sample codes.

Sampling Location	Sample Code	N	WAtl	WAtl	Avg. A.S. to WAtl	Med	Med	Avg. A.S. to Med
			A.S. >70%	A.S. <70%		A.S. >70%	A.S. <70%	
Gulf of St. Lawrence	WAGSL13L	22	36.0	14.0	78.9	41.0	9.0	87.6
Central North Atlantic	CACA12L	46	13.0	4.0	84.4	59.0	24.0	81.0
Central North Atlantic	CACA12M	24	8.0	4.0	70.2	67.0	21.0	86.2
Canary Islands	EACI13L	24	8.0	13.0	71.9	75.0	4.0	90.9
Morocco	EAMO12L	24	8.0	8.0	70.4	79.0	4.0	88.5
Morocco	EAMO13L	19	16.0	16.0	71.4	63.0	5.0	91.1
Portugal	EAPO11L	23	0.0	4.0	51.3	70.0	26.0	84.4
Gulf of Sirte	CMSY12L	24	4.0	0.0	70.0	83.0	13.0	89.0
Gulf of Sirte	CMSY12M	24	4.0	4.0	69.2	79.0	13.0	87.5
Total		230	10.9	7.0	75.2	67.8	14.3	86.6

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1755-0998.12764

This article is protected by copyright. All rights reserved.



