

Development and potential application of a new set of Atlantic bluefin tuna EST-SSRs in survival success during the farming cycle

Ivana RADONIĆ¹, Željka TRUMBIĆ², Tanja ŠEGVIĆ- BUBIĆ¹, Leon GRUBIŠIĆ¹ and Ivona MLADINEO¹

¹ Institute of Oceanography and Fisheries, Šetalište I. Meštrovića 63, 21000 Split, Croatia

² University of Split, University Department of Marine Studies, Ruđera Boškovića 37, 21000 Split, Croatia

Corresponding author: radonic@izor.hr

Handling Editor: Panagiotis KASAPIDIS

Received: 30 November 2018; Accepted: 10 March 2020; Published on line: 13 May 2020

Abstract

The capture-based aquaculture of Atlantic bluefin tuna (*Thunnus thynnus*) starts with the catch of wild individuals, slow-operating transport to the rearing facilities, introduction to rearing cages, and adaptation to the captive environment. All these steps bear significant risks for the onset of stressful conditions that can consequently result in unexpected mortalities. In order to explore whether survival success throughout the farming cycle might be monitored at the genetic level and linked to certain immunity and stress response-associated genes, we developed a new set of 13 EST-SSRs for *T. thynnus* and subsequently analysed 334 samples of juvenile wild-caught tuna and captive-reared adults during two consecutive farming cycles in the Adriatic Sea. The results showed a low F_{ST} value (0.005) with similar allele frequencies and no major allele loss between the investigated groups. The two tested approaches for the identification of loci under selection did not indicate a departure from neutrality for any of the 13 EST-SSRs, suggesting that the latter could not be considered adaptive in the studied context. These results are in agreement with other studies aiming to detect adaptive signals in *T. thynnus*, stressing the problems associated with sampling design for a species with complex migratory behaviour and reproduction, and the specific zootechnical practices in place at farms. Nonetheless, the characteristics of the 13 new polymorphic loci reported here contribute to a broadening of the existing EST-SSRs resource, as a useful asset for future genetic studies of *T. thynnus*.

Keywords: Atlantic bluefin tuna; microsatellites; EST-SSRs; aquaculture; Adriatic Sea.

Abbreviations ATT: Adriatic *Thunnus thynnus*; A_r : Allelic Richness; DAPC: Discriminant Analyses of Principal Components; EST: Expressed Sequence Tag; F_{IS} : Wright's Inbreeding Coefficient; F_{ST} : Fixation Index; H_o : Observed Heterozygosity; H_e : Expected Heterozygosity; HWE: Hardy-Weinberg Equilibrium; ICCAT: International Commission for the Conservation of Atlantic Tunas; LD: Linkage Disequilibrium; MISA: MicroSatellite Identification Perl tool script; N_{pr} : Number of Private Alleles; PCR: Polymerase Chain Reaction; SNP: Single Nucleotide Polymorphism; SSR: Simple Sequence Repeats.

Introduction

Atlantic bluefin tuna (*Thunnus thynnus*), one of the largest pelagic fish of the family Scombridae, is an important species for commercial and recreational fisheries. In the past, its numbers have suffered a decline due to overfishing (Taylor *et al.*, 2011), therefore the species has consequently come under the protective management of the International Commission for the Conservation of Atlantic Tunas (ICCAT). The Commission divided the population into the Eastern and Western stocks separated by the 45° meridian (Carlsson *et al.*, 2004) with concomitant division in the Gulf of Mexico and the Mediterranean stocks, based on two principal zones of spawning and nurseries (Rooker *et al.*, 2007; Riccioni *et*

al., 2010a). The distinct genetic differentiation between the two stocks (Vinas *et al.*, 2011) has been recognized through many tagging and genetic studies (Block *et al.*, 2005; Rooker *et al.*, 2007, 2014; Boustany *et al.*, 2008; Albaina *et al.*, 2013).

Croatia was the first country to start tuna capture-based aquaculture in the Mediterranean Sea (Katavic *et al.*, 2003; Miyake *et al.*, 2003) with the objective of increasing the commercial value of fish by increasing fat content in the muscle, as desirable on the sushi and sashimi market (Miyake *et al.*, 2003; Mylonas *et al.*, 2010). The process generally involves the capture of juvenile or adult wild tuna of over 30 kg by purse seines in spring, and their intensive feeding in sea cages for periods ranging between 3 and 24 months to increase tuna total biomass

(Ottolenghi, 2008; Mylonas *et al.*, 2010). Only Croatia is allowed to catch immature tuna (8-20 kg body weight and <130 cm fork length) in the Adriatic Sea that consequently necessitate an extended period of farming (18 to 36 months), in contrast to practices in other Mediterranean countries (Miyake *et al.*, 2003; Ottolenghi, 2008; Mylonas *et al.*, 2010).

The catch of juvenile wild tuna, slow-operating transport to farming facilities, translocation from transport into rearing cages, and subsequent adaptation to cage conditions, combined with other biotic and abiotic factors, can prove stressful for fish and lead to unexpected mortality (Meyer, 1991; Magnadóttir, 2006; Ottolenghi, 2008; Mylonas *et al.*, 2010). Stress from the hauling process, traumatic injuries in juvenile tuna caused by collision with cage walls or net entanglement (Miyashita *et al.*, 2000) may increase susceptibility to infection and exposure to pathogens (Kirchhoff, 2012; Rauta *et al.*, 2012), hampering growth (Mladineo *et al.*, 2006) and ultimately causing large losses during the production cycle (Evans, 2015; Buentello *et al.*, 2016). Mortality during capture, transport and the initial post-capture adaptation period can vary between 2 to 30%, although losses can be greatly decreased with suitable handling (Katavic *et al.*, 2003; Nakada, 2008; Evans, 2015).

As a highly abundant form of repetitive DNA in eukaryotic and prokaryotic genomes, microsatellites or Simple Sequence Repeats (SSR) have become widely used in genetic-based fisheries management and ecology studies, useful in defining stocks, and investigating genetic diversity within and among fish populations, including tuna (Broughton & Gold, 1997; Carlsson *et al.*, 2004, 2007; Riccioni *et al.*, 2010a, 2013, 2017; Antoniou *et al.*, 2017). SSRs are distributed throughout the genome in both coding and non-coding regions (Tautz & Renz, 1984; Tóth *et al.*, 2000; Li *et al.*, 2002; Chistiakov *et al.*, 2006). Although they are generally considered selectively neutral, they may be functionally relevant, particularly those located near or in coding regions. The majority of SSRs used in fish aquaculture studies have been non-coding, in contrast to EST-SSRs developed from Expressed Sequence Tags (EST) of the cDNA libraries associated with expressed genes. Moreover, the polymorphisms of the latter within a coding region may influence the expression of functional genes, and consequently may affect the fitness of individuals associated with the corresponding performance (Wren *et al.*, 2000; Liu & Cordes, 2004; Chistiakov *et al.*, 2006; Gemayel *et al.*, 2010).

Therefore, the scope of this study was to explore whether the survival success of Atlantic bluefin tuna throughout the farming cycle might be monitored at the genetic level and linked to several immunity and stress response-associated genes. We employed the existing molecular Atlantic bluefin tuna resources to develop a set of *T. thynnus*-specific EST-derived microsatellite markers (EST-SSRs), and tested the potential genetic differences between groups of wild-caught juveniles at the beginning of the farming cycle and captive-reared adults at the end of the farming cycle, during two consecutive farming cycles in the Adriatic Sea.

Material and Methods

Fish sampling and DNA isolation

In total, 334 caudal fin clips of Atlantic bluefin tuna *T. thynnus* were collected during June 2009, December 2011, June 2012 and December 2014 from a tuna facility located off the southwest coast of the island Brač in the Adriatic Sea (Fig. S1). Sampling was conducted during two consecutive farming cycles: at the introduction of juvenile wild-caught tuna in cages in spring 2009 and 2012 (Wild-caught2009 and Wild-caught2012) and at the harvest after the farming period of the same groups in winter 2011 and 2014 (Harvest2011 and Harvest2014) (Table 1). Juvenile wild-caught are groups that contain both: i) individuals that will survive up to the harvesting season, and ii) those that do not make it up to the harvesting season due to presumably disadvantageous or non-adaptive alleles. Harvested groups contain only the adult 'survivors' that successfully completed the farming cycle.

Fin clips were stored in 96% ethanol prior to DNA analysis. Total genomic DNA was extracted following a modified mammalian DNA isolation method (Laird *et al.*, 1991) and stored in TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0). DNA quantity and quality were assessed by UV spectrophotometer (JENWAY, Bibby Scientific, UK) and each sample was diluted to 10 ng/μl in milliQ water and stored at -80°C.

In silico development of *Thunnus thynnus* EST-SSR markers

EST (Expressed Sequence Tag) libraries of different *T. thynnus* tissues, previously obtained by 454 sequencing by Trumbić *et al.* were used to develop EST-SSRs (Trumbić *et al.*, 2015).

In silico mining of EST-SSRs was performed using the MICroSAtellite Identification Perl tool script (MISA) with default parameters, and primer modelling software Primer3 to design PCR primers flanking the microsatellite regions (Untergasser *et al.*, 2012). In total, 25 loci were randomly selected linked to a set of ESTs with immunity, stress or signal transduction putative roles. Their coding regions were characterized as described

Table 1. Number of caudal fin clip samples from Atlantic bluefin tuna *T. thynnus* collected in two consecutive farming cycles at the tuna facility in the Adriatic Sea (43°28'77.04" N - 16°48'28.06" E).

Name of the population	Number of Samples	Farming Cycle
Wild-caught2009	100	1.
Harvest2011	95	1.
Wild-caught2012	55	2.
Harvest2014	84	2.

previously (Trumbić *et al.*, 2015). Briefly, the longest coding regions were determined by the program getorf from the EMBOSS v6.6.0 package (Rice *et al.*, 2000), confirmed by ESTScan v2 (Iseli *et al.*, 1999) and annotated by similarity searches against GenBank reference databases using BLAST (Altschul *et al.*, 1990). Additionally, coding regions were translated and specifically annotated against GenBank Reference Proteins database of *Danio rerio* with an expectation value cutoff of 10^{-4} . Retrieved annotations were used to recover the Gene Ontology terms for molecular function, biological process and cellular component (Ashburner *et al.*, 2000) through org. Dr.eg.db, the zebrafish genome wide annotation package for R (Carlson, 2018).

Optimisation of the multiplex PCR of *Thunnus thynnus* EST-SSR markers

Of the 25 microsatellites initially selected, 15 successfully amplified a PCR product of the expected size, as checked by 2% agarose gel electrophoreses using SybrSafe (Invitrogen) to visualize DNA. The polymorphic potential of loci was checked by 4% agarose gel electrophoresis over 12 DNA samples from different experimental groups, as above. Two loci were removed from further testing as they were monomorphic. The selected primers were optimized to give the product under the same conditions in multiplex PCR reaction (Neff *et al.*, 2000) and labelled with fluorescent dyes (NED, PET, 6FAM and VIC) (Table 2). Multiplex PCR conditions involved an initial denaturation step at 95 °C for 5 min; followed by 28 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 90 s and extension at 72 °C for 30 s; and final extension at 60 °C for 30 min. Reactions were prepared in a final volume of 5 µl, using 2.5 µl Type-it Microsatellite PCR Kit (Qiagen), 0.2 µM of each primer, 10 ng DNA template and milliQ water. Each reaction was twice diluted and loaded in an ABI3730xl DNA automated sequencer (Applied Biosystems) using the Macrogen service (Korea), while peak height values for each microsatellite allele were scored by two different persons using the GeneMapper v.3.2 software (Applied Biosystems).

Statistical analysis

Scored alleles were checked for genotyping errors associated with microsatellite analysis such as stutter bands, the presence of null alleles and large allele dropout using MICRO-CHECKER 2.2.3 (Van Oosterhout *et al.*, 2004). The presence and frequency of null alleles were additionally examined using FreeNA (Chapuis & Estoup, 2007) following the Expectation Maximization (EM) algorithm. The program applies the ENA (Excluding Null Alleles) correction method to effectively correct for the positive bias induced by the presence of null alleles in the F_{ST} estimation. The bootstrap 95% confidence intervals (CI) for the global F_{ST} values were calculated using 10,000 replicates across the loci.

Genetic diversity indices of the tested groups, as allele frequencies, total number of alleles per locus (N), mean number of alleles per locus, observed heterozygosity (H_o) and expected heterozygosity (H_e) were calculated using ARLEQUIN 3.5 (Excoffier & Lischer, 2010). The allelic richness (A_r), as a standardized measure of the number of alleles per locus independent of the sample size, and F_{IS} , as an estimator of inbreeding, were estimated using FSTAT 2.9.3.2 (Goudet, 2002).

Deviations from the Hardy-Weinberg Equilibrium (HWE) and Linkage Disequilibrium (LD) for all pairs of loci were estimated using ARLEQUIN 3.5. The HWE was tested with 1,000,000 steps in the Markov chain while LD was tested with 10,000 permutations. Significance levels were adjusted using sequential Bonferroni correction (Rice, 1989).

ARLEQUIN 3.5 calculated global and pairwise F_{ST} values. In order to test the partitioning of genetic variation within and among groups, a hierarchical AMOVA was conducted using 2,000 permutations.

The statistical power of sampling design for exact tests given different levels of F_{ST} was estimated using the POWSIM software ver. 4.1 (Ryman & Palm, 2006). Simulations were performed for a scenario involving four tuna groups with F_{ST} ranging from 0.001 to 0.01, using different combinations of effective population size (N_e) and numbers of generations of drift (t) per F_{ST} .

Discriminant Analyses of Principal Components (DAPC), a multivariate non-model based method, was used to infer and visualize genetic structuring between two consecutive farming cycles (Jombart, 2008). The optimal number of principal components best describing most sources of variation in the data with lowest error rate, was determined by a cross-validation procedure using stratified random sampling with 1,000 iterations.

A model-based Bayesian clustering method was used to infer the genetic structure and to identify clusters (K) of genetically similar individuals in the data set by STRUCTURE 2.3.4 (Pritchard *et al.*, 2000; Falush *et al.*, 2003), using the admixture ancestry model, correlated allele frequencies, a burn-in period of 10,000 iterations followed by 100,000 MCMC steps and K values from 1 to 5 with 20 replicates each. To assess the most likely number of clusters, $\ln(P)$ and ΔK values were estimated in Structure Harvester 0.6 (Earl & von Holdt, 2012).

To test whether the loci deviated from neutrality, outlier analyses were conducted with LOSITAN (Antao *et al.*, 2008) and BayeScan 2.1 (Foll, 2012). LOSITAN was run on a stepwise mutation model with the following settings: 50,000 simulations, 95% confidence interval, forced mean F_{ST} and with a 0.05 false discovery rate. Default parameter setting was used for the BayeScan run (prior odds 10, samples size 5,000, thinning interval 10,000, pilot runs 20, pilot run length 5,000 and additional burn-in 50,000), while the decision whether the locus was under selection was based on the magnitude of Bayes Factor ($\log_{10}(BF) > 0.5$: “substantial” evidence for selection; 1.5–2.0: “very strong”; > 2.0 : “decisive”) as suggested by Jeffreys (1998).

Table 2. Characteristics of 13 EST-SSR primers derived from the EST (Expressed Sequences Tags) libraries of Atlantic bluefin tuna *T. thynnus* sampled during two consecutive farming cycles at the tuna facility in the Adriatic Sea.

Locus	Primer sequence (5'- 3') with fluorescently tagged forward primer for use in multiplex PCR		Repeat motif	Allele size range (bp)	Position*	Gene symbol and annotation
ATT1	PET -	(F)GTCATCATCTTCGAGAGCGTC (R)GTGTCGGCCATCTTGTTGTAG	(TCCG) ₇	98-154	Coding	slc6a8; solute carrier family 6 (neurotransmitter transporter, creatine), member 8
ATT2	PET -	(F)AGAATGTGTTTGGCTGCTGATA (R)TTCGTCTCTTTCTTTCACGACA	(AAAC) ₅	148-188	Coding	nampta; nicotinamide phosphoribosyltransferase-like
ATT3	NED -	(F)GGGACACAGATTGATAAAGACAC (R)CCACTGACGCCATGACAC	(CAGA) ₆	81-129	5'	scn4ba; sodium channel, voltage-gated, type IV, beta a
ATT5	VIC -	(F)ACGACATCCAGACCAAGAGAAG (R)AGGAAACAAACCACCTGAGTGA	(ACTGA) ₆	192-252	Coding	polr2d; polymerase (RNA) II (DNA directed) polypeptide D
ATT11	VIC -	(F)ATACAGGTTGGTCTGCCGTC (R)AGTGTCCGTCATCACACAGC	(CA) ₁₁	154-199	3'	cfI; complement factor I
ATT12	NED -	(F)GATTGGTGTCTACTGGTGTGC (R)CAAGGAGACAAAGACCCGAG	(GT) ₈	127-159	5'	mapk6; mitogen-activated protein kinase 6
ATT13	6FAM	(F)AATGCATGTGGGATCAACAA (R)TCTGGCTGAATGGGAAAGTC	(AC) ₁₁	185-229	3'	ctsf; cathepsin F
ATT15	VIC -	(F)TCCAAGAAGAAGAAGGCGAA (R)TATGCAGGTTTTCTCCACCC	(GAG) ₅	114-144	Coding	nkaf; NFKB activating protein-like
ATT18	NED -	(F)AGGAAGGAGAGCTGGGTTTC (R)GGGGCTGATGAGATCAAGAA	(AAC) ₁₁	211-277	3'	grb2b; growth factor receptor-bound protein 2b
ATT19	NED -	(F)ACGGCACATCTCACAAATGA (R)TGAGTAACGATGGCTTCTGC	(AAC) ₅	181-211	3'	tnfr1; TNF receptor 1 and NCK interacting kinase a
ATT22	6FAM	(F)CAAACAGAAAAAGAGCCGTC (R)GCACAGACAGCTGCATTGAG	(ATC) ₈	236-284	Coding	ppara; peroxisome proliferator-activated receptor delta a
ATT23	VIC -	(F)TCTGCAACAGCCAGTAGAGC (R)GCAACAGCAGCAAGTGAGTG	(AC) ₁₀	256-296	3'	sch1; SHC-transforming protein 1-like
ATT25	6FAM	(F)ATTACAGTGTGCGGAGCATC (R)TGTCAGTGAATGTGACTGCTG	(CAG) ₈	134-182	Coding	ldlra; low density lipoprotein receptor a

*Putative position in respect to inferred EST coding region using ESTScan.

Results

In silico mining of the EST-SSRs in Thunnus thynnus

Microsatellite repeat motifs, excluding mononucleotides, were identified in 9.2% of 10,093 annotated EST sequences available from a previous study (Trumbić *et al.*, 2015). Di- and tri-nucleotide perfect repeats were the most numerous, comprising 71% of SSRs (Table 3). The most common motif among di-nucleotide repeats was AC (Fig. S2), as generally found in vertebrates (Tóth *et al.*, 2000), including sea bass *Dicentrarchus labrax* and sea bream *Sparus aurata* (Louro *et al.*, 2010), common carp *Cyprinus carpio* (Ji *et al.*, 2012), Atlantic cod *Gadus morhua* (Carlsson *et al.*, 2013) and other fish species (Ju *et al.*, 2005). The CG repeat was not detected, similarly to the situation in sea bass and sea bream (Louro *et al.*, 2010), while CAG was the most prevalent among tri-nucleotide SSRs. After inspection of repeats found in

Table 3. Counts of microsatellite repeats (di-, tri-, tetra-, penta-, hexa- and compound motives) found in the Atlantic bluefin tuna *T. thynnus* EST sequences (targets)

SSR type	Count	Frequency	Targets
Di-	364	0.377	347
Tri-	332	0.344	319
Tetra-	24	0.025	24
Penta-	4	0.004	4
Hexa-	4	0.004	3
Compound	237	0.246	232
Total	965	1	929

compound formation, the distribution of microsatellite motif counts remained similar, though with increased additional abundance of the GTAC tetranucleotide (Fig. S2). Most SSRs contain approximately 10 or fewer repeats, with few exceeding 20. In the latter case, the most expanded was the GAC motif with 50 repeats found in compound formation (data not shown).

Genetic diversity

A total of 10% of randomly selected tuna samples were genotyped twice in separate PCR reactions, and the genotypes compared to estimate error in the present study. The genotyping error rate by direct count was 0.2%. There was 0.92% missing data in the whole dataset. The potential presence of null alleles was detected by MICROCHECKER at loci ATT13 and ATT15. However, they were retained in further analysis since the F_{ST} values gave similar results with or without the use of ENA correction: $F_{ST} = 0.005$ with 95% CI 0.0021–0.0079 without ENA and 0.0027–0.0087 with ENA, and estimated frequencies of null alleles were < 0.08 . Three loci (ATT3, ATT12 and ATT13) showed significant deviation from Hardy-Weinberg equilibrium (HW) in more than one population, with strong tendencies toward heterozygote excess. No LD was observed among the genotypes of 13 loci, suggesting that the analysis could be performed assuming statistical independence of loci.

Population descriptive statistics per locus, including results for the HWE test, are presented in Table S1. Twelve loci were polymorphic with a mean expected heterozygosity > 0.49 in all four sample groups, and the locus ATT5 was monomorphic in only one group (Wild-caught2012). The observed (H_o) and expected (H_e) heterozygosities varied from 0.49 to 0.60 and 0.49 to 0.52, respectively, and did not differ significantly between the Wild-caught and Harvest groups ($p > 0.05$). A slightly higher mean number of alleles per locus (N) was observed in tuna at the end of both farming cycles; i.e. Harvest2011 vs. Wild-caught2009 and Harvest2014 vs. Wild-caught2012. Allelic richness A_r varied between 2 and 16.52 globally, and was not significantly different between the Wild-caught and Harvest groups ($p > 0.05$). The highest mean value of A_r was detected in the Harvest2011 group (5.67 ± 3.68), followed in descending order by the groups Wild-caught2009 (5.41 ± 3.69), Harvest2014 (5.33 ± 2.95) and Wild-caught2012 ($A_r = 4.77 \pm 3.23$).

Tuna from the first farming cycle and tuna from the second farming cycle showed 12 private alleles (N_{pr}) over 85 and 15 N_{pr} over 77 in total, respectively. For both farming cycles, harvested tuna showed a slightly higher mean number of N_{pr} in comparison to wild-caught specimens (Table S1). The mean F_{IS} value across groups and loci was -0.072 , and did not differ significantly from zero ($p > 0.05$). Single-locus F_{IS} ranged from -0.454 to 0.196 in Wild-caught groups, and from -0.480 to 0.526 in Harvest groups, and was significantly lower than zero in some cases, indicating a slight excess of heterozygosity.

Genetic differentiation among populations

Evaluation of statistical power estimated a 89% probability (χ^2 , Fisher's test) of detecting a structure for $F_{ST} = 0.002$, stating that the exact tests used herein may detect a real population structure if the true estimates of F_{ST} were at or above this level.

The global F_{ST} value of 0.005 (95% CI= 0.002–0.008) was low but statistically significant ($p < 0.05$). Pairwise F_{ST} across all samples of tuna ranged from 0.0009 to 0.0095, with an absence of statistical significance between groups from the first farming cycle ($p > 0.01$) vs those from the second farming cycle ($p < 0.01$) (Table S2). In addition, AMOVA showed that only 0.25% of the total variance was explained by differences among groups and 99.45% within groups.

Both cluster analyses (STRUCTURE and DAPC) gave concordant results in terms of an absence of population structure within and between the two consecutive farming cycles (Fig. 1 and Fig. S3). Highly admixed membership probabilities of each group within each cluster are presented in Fig. S3.

The two approaches for the identification of loci under selection (LOSITAN and BayeScan) did not indicate any evidence of selection acting on the 13 microsatellites used, suggesting their neutrality in the studied context (Fig. 2). All posterior odds and \log_{10} posterior odd values computed in BayeScan were not significant, as evidenced by q-values that failed to detect any outlier loci (Table S3).

Discussion

Atlantic bluefin tuna (*T. thynnus*) microsatellite, Single Nucleotide Polymorphism (SNP) and mitochondrial DNA-based data have been widely employed to discern genetic differentiation between geographic locations, stock-structure and estimates of effective population size essential for proper conservation and management of the species (Ward, 1995; Clark *et al.*, 2004; Carlsson *et al.*, 2007; Riccioni *et al.*, 2010b; Cuéllar-Pinzón *et al.*, 2016; Vella *et al.*, 2016; Antoniou *et al.*, 2017). Although considered useful for planning sustainable breeding programmes (Chistiakov *et al.*, 2006; Cerdà & Manchado, 2013; Cuéllar-Pinzón *et al.*, 2016), there is still a lack of approaches for employing EST-SSRs related to production (growth rate, stress response and disease resistance) and reproductive traits (sex determination and development rate), which could depict tuna genetic structure and simultaneously characterise genes that control such traits.

A set of 13 EST-SSRs was used to test whether a particular genotype in juvenile wild-caught tuna entering the farming cycle enables better fitness in fish to survive to harvest time. The selected microsatellites were chosen for their proximity to the coding regions of genes associated with different aspects of immune and cellular active-transport system, essential for fish survival (Table S4). Most of the tested EST-SSRs exhibited a relatively low level of genetic diversity, showing 2–18 alleles per

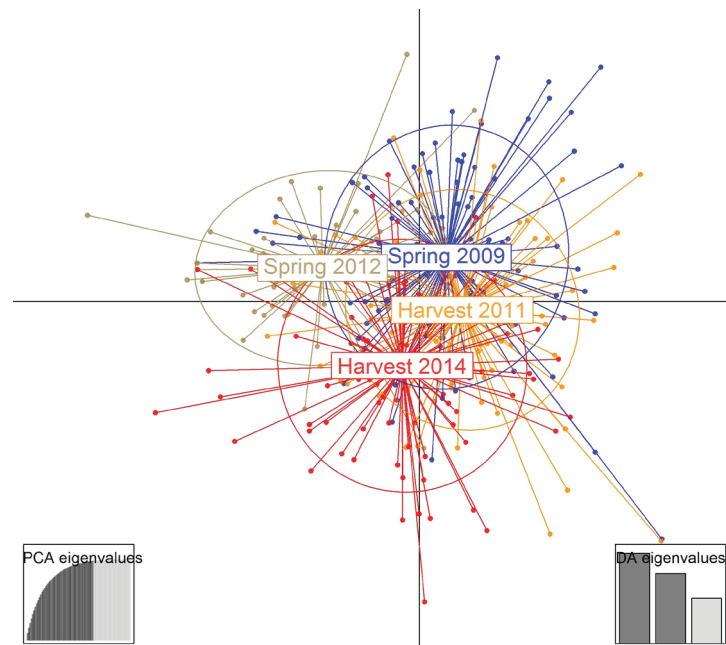


Fig. 1: Scatterplot of the first two principal components of the DAPC analyses of population structure of the Atlantic bluefin tuna *T. thynnus*, sampled during two consecutive farming cycles at the tuna facility in the Adriatic Sea, inferred by 13 EST-SSR loci.

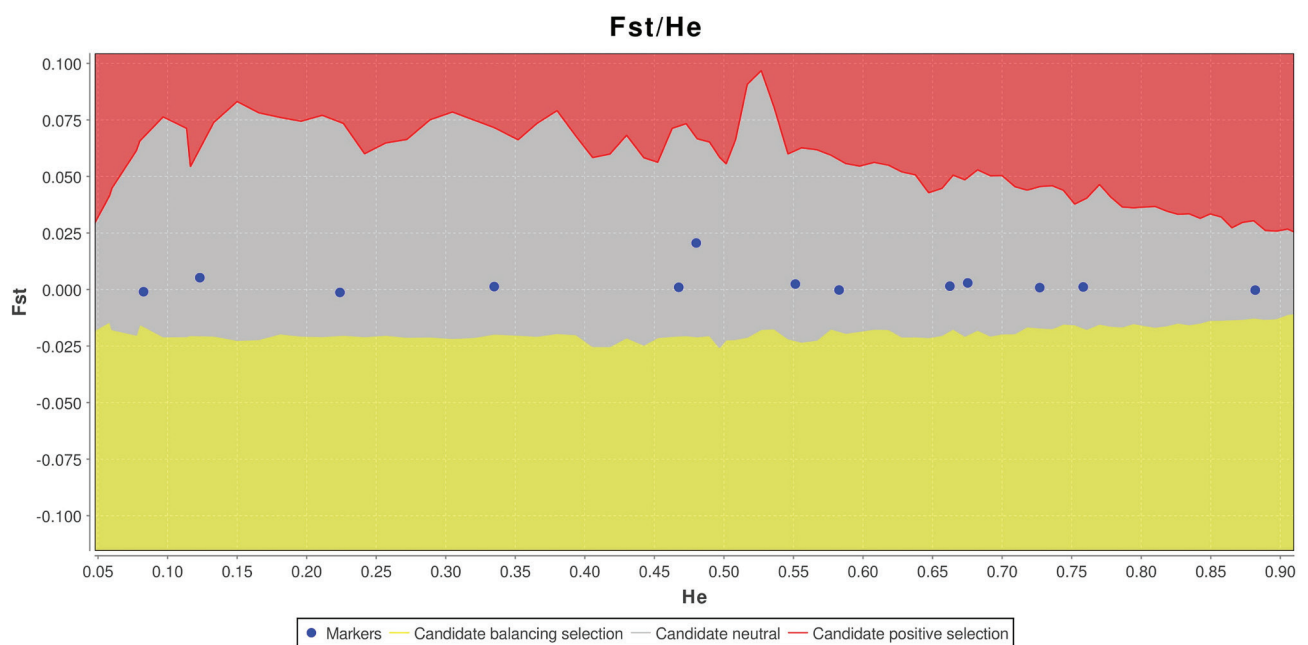


Fig. 2: Comparison of F_{st} and H_e in the 13 loci of the Atlantic bluefin tuna *T. thynnus* used to identify outliers and potential candidates for selection with the LOSITAN software. The graphical output shows the simulated confidence area for neutral loci (grey), positive selection (red) and balancing selection (yellow).

locus, with a similar allele frequency between groups in both farming cycles. Furthermore, the mean number of A_r was similar among all four groups, congruent to previous studies in bluefin tuna in the Mediterranean Sea (Carlsson *et al.*, 2004; Riccioni *et al.*, 2010a; Vella *et al.*, 2016). However, contrary to expectations, no statistical difference was observed between A_r in Harvest groups compared to Wild-caught groups. The observed similarities and the fact that there is no major allele loss between groups entering the farming cycle and at the end of the process, coupled with neutral results of LOSITAN and BAYSCAN analyses, suggest that variabilities in the test-

ed loci likely do not affect the survival rate of tuna during farming.

Riccioni *et al.* (2017) used a set of 12 EST-SSRs of *T. thynnus* to elucidate its population dynamics in the Mediterranean by linking genetic differentiation to adaptation to environmental factors. The authors observed a neutral behaviour of the selected markers, lacking the differentiation signal among the Atlantic bluefin tuna from the Mediterranean Sea. This is consistent with the fact that the majority of microsatellite loci associated with coding regions present lower levels of variability compared to those in non-coding regions. The reason is that

protein-coding regions are more conserved than non-coding regions, primarily because the former are translated into proteins and therefore subject to a more significant degree of selective pressure (Metzgar *et al.*, 2000; Borstnik & Pumpernik, 2002; Li, 2004; Riccioni *et al.*, 2017).

In this study, several loci (ATT3, ATT12 and ATT13 in the first farming cycle and ATT3, ATT11, ATT12, ATT15, ATT18, ATT23 and ATT25 in the second farming cycle) showed significant deviations from HWE. Such deviations have been reported previously and might indicate a departure from random mating, population mixing, the Wahlund effect and genotyping errors (Norris *et al.*, 1999; Dewoody *et al.*, 2006; Selkoe & Toonen, 2006; Morin *et al.*, 2009; Waples, 2015). In farmed tuna specifically, this could be related to the “unconscious” collection of samples from a single group that encompasses specimens of distinct generations. This is due to the carry-over of underaged specimens in the farm or cages, following a particular harvest operation (ICCAT, 2017). Namely, considering that the farming process in Croatia involves intensive feeding of juveniles in cages for periods ranging between 18 to 36 months, companies annually carry-over around 30,000 tuna into the next farming year, mixing fish from distinct generations into a single farming group.

The observed F_{ST} value in the present study was low, as expected given the fact that we focused on detecting possible individual adaptive loci during farming. Riccioni *et al.* (2017) recorded a 10 times lower F_{ST} value in wild *T. thynnus* using EST-SSRs than neutral microsatellite loci (Riccioni *et al.*, 2010; 2013), suggesting that the former are characterized by a lower mutation rate, indicative of evolutionary constraints. A loci neutrality test used by the authors confirmed that loci were neutral with the absence of selection footprints, similar to the findings reported here. Another concomitant analysis of genome-wide SNPs and microsatellites in the Atlantic bluefin collected throughout the Mediterranean Sea revealed a low level of genetic differentiation, concordant to what is expected for marine species with high gene flow potential and high dispersal capacity (Antoniou *et al.*, 2017).

Though such markers are usually characterized by low polymorphism, there is a growing number of studies employing markers originating from coding regions in the genome (Nikolic *et al.*, 2015; Riccioni *et al.*, 2017). Although they are potentially prone to selection, only a small percentage of loci show evidence of positive selection, even when closely associated with functional regions (Ellis & Burke, 2007; Riccioni *et al.*, 2017). Testing potential differences inferred by loci in physiologically relevant genes in the Atlantic bluefin tuna *T. thynnus* population in the Adriatic Sea, we failed to pinpoint an overall specific pattern that would indicate a change in fish fitness and survival at the end of the farming period. Our sampling strategy and selection of targeted loci were likely rather limited in their power to detect a trait such as farming survival, further hampered by the zootechnical practice at Croatian farms of mixing different fish generations. There is a chance that selection may be occurring, only not in association with these selected markers. However, appropriate sampling design remains a challenge for

the scientific community for this highly migratory species, even when using genome-wide screens (Riccioni *et al.*, 2017). An approach that considers sampling young of the year specimens of known geographic origins, tagging them upon entry into the rearing cycle, simplifying their identification throughout the rearing period, and using of NGS techniques like ddRAD to infer polymorphisms in thousands of reads, represents a considerable effort in aquaculture, but would potentially reveal more information in similar studies.

The loci reported here can indeed contribute to the EST-SSRs resource and expand the list of microsatellite markers available for *T. thynnus*. This is an important asset for future studies as EST-linked markers in general can be of great use in monitoring levels of genetic variation of genes of interest, and are informative in planning different zootechnic and prophylactic measures during the farming process.

Acknowledgements

We are especially grateful to Iva Žužul for assisting with statistical data analysis and the staff of the Laboratory for Aquaculture, Institute of Oceanography and Fisheries, for their support during laboratory work. No formal funding was obtained for this particular study; however, the research received secondary support from the Croatian Scientific Foundation, grants AnGEL #5576 and AquaPop #9050, and from the Ministry of Agriculture with the grant PRUT.

References

- Albaina, A., Iriondo, M., Velado, I., Laconcha, U., Zarraonaindia, 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.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.
- Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., Luikart, G., 2008. LOSITAN, A workbench to detect molecular adaptation based on a F_{ST} -outlier method. *BMC Bioinformatics*, 9, 1-5.
- Antoniou, A., Kasapidis, P., Kotoulas, G., Mylonas, C. C., Magoulas, A., 2017a. Genetic diversity of Atlantic Bluefin tuna in the Mediterranean Sea, insights from genome-wide SNPs and microsatellites. *Journal of Biological Research (Thessalonike, Greece)*, 24 (3), 1-14.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H. *et al.*, 2000. Gene Ontology, tool for the unification of biology. *Nature Genetics*, 25, 25-29.
- Block, B.A., Teo, S. L. H., Walli, A., Boustany, A., Stokesbury, M.J.W. *et al.*, 2005. Electronic tagging and population structure of Atlantic bluefin tuna. *Nature*, 434, 1121-1127.
- Borstnik, B., Pumpernik, D., 2002. Tandem Repeats in Protein Coding Regions of Primate Genes. *Genome Research*, 12,

909-915.

- Boustany, A.M., Reeb, C.A., Block, B. A. 2008. Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*). *Marine Biology*, 156, 13-24.
- Broughton, R.E., Gold, J.R., 1997. Microsatellite development and survey of variation in northern bluefin tuna (*Thunnus thynnus*). *Molecular Marine Biology and Biotechnology*, 6, 308-314.
- Buentello, A., Seoka, M., Kato, K., Partridge, G.J., 2016. Tuna Farming in Japan and Mexico. p. 189-215. In, *Advances in Tuna Aquaculture*, Benetti, D. D., Partridge, G.J., Buentello, A., (Eds). Academic Press is an imprint of Elsevier. Oxford.
- Carlson, M. 2018., org.Dr.eg.db, Genome wide annotation for Zebrafish. R package version 3.6.0.
- Carlsson, J., McDowell, J.A.N., Diaz- Jaimes, P., Carlsson, J.E., Boles, S.B. *et al.*, 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, J.R., Carlsson, J.E.L., Graves, J.E., 2007. Genetic identity of YOY bluefin tuna from the Eastern and Western Atlantic spawning areas. *Journal of Heredity*, 98, 23-28.
- Carlsson, J., Gauthier, D.T., Carlsson, J.E.L., Coughlan, J.P., Dillane, E. *et al.*, 2013. Rapid, economical single-nucleotide polymorphism and microsatellite discovery based on *de novo* assembly of a reduced representation genome in a non-model organism, a case study of Atlantic cod *Gadus morhua*, novel SNPs and microsatellites in *Gadus morhua*. *Journal of Fish Biology*, 82, 944-958.
- Cerdà, J., Manchado, M., 2013. Advances in genomics for flat-fish aquaculture. *Genes & Nutrition*, 8, 5-17.
- Chapuis, M.P., Estoup, A., 2007. Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution*, 24, 621-631.
- Chistiakov, D.A., Helleman, B., Volckaert, F.A.M., 2006a. Microsatellites and their genomic distribution, evolution, function and applications, A review with special reference to fish genetics. *Aquaculture*, 255, 1-29.
- Clark, T.B., Ma, L., Saillant, E., Gold, J.R., 2004. Microsatellite DNA markers for population-genetic studies of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) and other species of genus *Thunnus*. *Molecular Ecology Resources*, 4, 70-73.
- Cuéllar-Pinzón, J., Presa, P., Hawkins, S.J., Pita, A., 2016. Genetic markers in marine fisheries, Types, tasks and trends. *Fisheries Research*, 173, 194-205.
- Dewoody, J., Nason, J.D., Hipkins, V.D., 2006 Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*, 6, 951-957.
- Earl, D.A., vonHoldt, B.M., 2012. Structure Harvester, a web-site and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359-361.
- Ellis, J.R., Burke, J.M., 2007. EST-SSRs as a resource for population genetic analyses. *Heredity*, 99, 125-132.
- Evans, D.L. 2015. *Stress and immune responses of farmed southern bluefin tuna (Thunnus maccoyii castlneau)*. PhD Thesis. University of Tasmania. 124 pp.
- Excoffier, L., Lischer H.L.E., 2010. Arlequin suite ver. 3.5, a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10 (3), 564-567.
- Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data, linked loci and correlated allele frequencies. *Genetics*, 164, 1567-1587.
- Foll, M., 2012. BayeScan v2. 1 user manual. *Ecology*, 20, 1450-1462.
- Gemayel, R., Vences, M.D., Legendre, M., Verstrepen, K.J., 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual Review of Genetics*, 44, 445-477.
- Goudet, J. 2002. FSTAT version 2.9.3.2, a program to estimate and test gene diversities and fixation indices. Available online, <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- ICCAT, 2017. International Commission for the Conservation of Atlantic Tunas. Available online, <https://www.iccat.int/en/#>
- Iseli, C., Jongeneel, C.V., Bucher, P., 1999. ESTScan, A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *International Conference on Intelligent Systems for Molecular Biology*, 138-148.
- Jeffreys, H., 1998. *Theory of probability*. Oxford classic texts in the physical sciences. Clarendon Press; Oxford University Press, Oxford [Oxfordshire], New York. 459 pp.
- Ji, P., Zhang, Y., Li, C., Zhao, Z., Wang, J., Li, J., Xu, P., *et al.* 2012. High throughput mining and characterization of microsatellites from common carp genome. *International Journal of Molecular Sciences*, 13, 9798-9807.
- Jombart, T. 2008. Adegenet, a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405.
- Ju, Z., Wells, M.C., Martinez, A., Hazlewood, L., Walter, R.B., 2005. An in silico mining for simple sequence repeats from expressed sequence tags of zebrafish, medaka, *Fundulus*, and *Xiphophorus*. *In Silico Biology*, 5, 439-463.
- Katavic, I., Franicevic, V., Ticina, V., 2003. Rearing of small bluefin tunas (*Thunnus thynnus* L.) in the Adriatic Sea Preliminary study. *Cashiers Options Mediterranennes*, 60, 95-99.
- Kirchhoff, N.T., 2012. *Health and performance of farmed southern bluefin tuna*. PhD Thesis. University of Tasmania, 183 pp.
- Laird, P.W., Zijderveld, A., Linders, K., Rudnicki, M.A., Jaenisch, R. *et al.*, 1991. Simplified mammalian DNA isolation procedure. *Nucleic Acids Research*, 19, 4293.
- Li, Y.C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E., 2002. Microsatellites, genomic distribution, putative functions and mutational mechanisms, a review. *Molecular Ecology*, 11, 2453-2465.
- Li, Y.C., 2004. Microsatellites within genes, structure, function, and evolution. *Molecular Biology and Evolution*, 21, 991-1007.
- Liu, Z.J., Cordes, J.F., 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238, 1-37.
- Louro, B., Passos, A.L.S., Souche, E.L., Tsigenopoulos, C., Beck, A. *et al.*, 2010. Gilthead sea bream (*Sparus auratus*) and European sea bass (*Dicentrarchus labrax*) expressed sequence tags, characterization, tissue-specific expression

- and gene markers. *Marine Genomics*, 3, 179-191.
- Magnadóttir, B., 2006. Innate immunity of fish (overview). *Fish & Shellfish Immunology*, 20, 137-151.
- Metzgar, D., Bytof, J., Wills, C. 2000. Selection against frame-shift mutations limits microsatellite expansion in coding DNA. *Genome Research*, 10, 72-80.
- Meyer, F. P. 1991. Aquaculture disease and health management. *Journal of Animal Science*, 69, 4201-4208.
- Miyake, P., De la Serna, J., Di Natale, F., 2003. General review of bluefin tuna farming in the Mediterranean area. *Collective Volumes of Scientific Papers ICCAT*, 55, 114-124.
- Miyashita, S., Sawada, Y., Hattori, N., Nakatsukasa, H., Okada, T. *et al.*, 2000. Mortality of northern bluefin tuna *Thunnus thynnus* due to trauma caused by collision during growout culture. *Journal of the World Aquaculture Society*, 31, 632-639.
- Mladineo, I., Miletić, I., Bočina, I., 2006. *Photobacterium damselae* subsp. *piscicida* Outbreak in Cage-Reared Atlantic Bluefin Tuna *Thunnus thynnus*. *Journal of Aquatic Animal Health*, 18, 51-54.
- Morin, P.A., Leduc, R.G., Archer, F.I., Martien, K.K., Huebinger, R. *et al.*, 2009. Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors: technical advances. *Molecular Ecology Resources*, 9, 498-504.
- Mylonas, C.C., De La Gándara, F., Corriero, A., Ríos, A.B., 2010. Atlantic bluefin tuna (*Thunnus thynnus*) farming and fattening in the Mediterranean Sea. *Reviews in Fisheries Science*, 18, 266-280.
- Nakada, M., 2008. Capture-based aquaculture, global overview. FAO fisheries technical paper. Food and Agriculture Organization of the United Nations, Rome. 298 pp.
- Neff, B.D., Fu, P., Gross, M.R. 2000. Microsatellite multiplexing in fish. *Transactions of the American Fisheries Society*, 129, 584-593.
- Nikolic, N., Duthoy, S., Destombes, A., Bodin, N., West, W. *et al.*, 2015. Discovery of genome-wide microsatellite markers in Scombridae, a pilot study on albacore tuna. *PloS One*, 10, 1-22.
- Norris, A., Bradley, D., Cunningham, E., 1999. Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. *Aquaculture*, 180, 247-264.
- Ottolenghi, F., 2008. Capture-based aquaculture, global overview. FAO fisheries technical paper. Food and Agriculture Organization of the United Nations, Rome, 298 pp.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.
- Rauta, P.R., Nayak, B., Das, S., 2012. Immune system and immune responses in fish and their role in comparative immunity study. A model for higher organisms. *Immunology Letters*, 148, 23-33.
- Riccioni, G., Landi, M., Ferrara, G., Milano, I., Cariani, A., *et al.*, 2010a. Spatio-temporal population structuring and genetic diversity retention in depleted Atlantic bluefin tuna of the Mediterranean Sea. *Proceedings of the National Academy of Sciences*, 107, 2102-2107.
- Riccioni, G., Ferrara, G., Landi, M., Sella, M., 2010b. Spatio-temporal genetic patterns in Mediterranean bluefin tuna, population structuring and retention of genetic diversity. *Collective Volumes of Scientific Papers ICCAT*, 65, 996-1003.
- Riccioni, G., Stagioni, M., Landi, M., Ferrara, G., Barbujani, G. *et al.*, 2013. Genetic structure of bluefin tuna in the Mediterranean Sea correlates with environmental variables. *Plos One*, 8, 1-9.
- Riccioni, G., Cariani, A., Ferrara, G., Cannas, R., Melis, R. *et al.* 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.
- Rice, P., Longden, I., Bleasby, A., 2000. EMBOS, the European Molecular Biology Open Software Suite. *Trends in genetics*, 16, 276-277.
- Rice, W.R., 1989. Analyzing Tables of Statistical Tests. *Evolution*, 43, 223.
- Rooker, J.R., Alvarado Bremer, J.R., Block, B.A., Dewar, H., De Metrio, G. *et al.*, 2007. Life history and stock structure of Atlantic bluefin tuna (*Thunnus thynnus*). *Reviews in Fisheries Science*, 15, 265-310.
- Rooker, J., Arrizabalaga, H., Fraile, I., Secor, D., Dettman, D. *et al.*, 2014. Crossing the line, migratory and homing behaviors of Atlantic bluefin tuna. *Marine Ecology Progress Series*, 504, 265-276.
- Ryman, N., Palm, S., 2006. POWSIM, a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, 6, 600-602.
- Selkoe, K.A., Toonen, R.J., 2006 Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, 9, 615-629.
- Tautz, D., Renz, M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, 12, 4127-4138.
- Taylor, N.G., McAllister, M.K., Lawson, G.L., Carruthers, T., Block, B.A., 2011. Atlantic bluefin tuna, a novel multistock spatial model for assessing population biomass. *PloS One*, 6, 1-10.
- Tóth, G., Gáspári, Z., Jurka, J. 2000. Microsatellites in different eukaryotic genomes, survey and analysis. *Genome Research*, 10, 967-981.
- Trumbić, Ž., Bekaert, M., Taggart, J.B., Bron, J.E., Gharbi, K. *et al.*, 2015. Development and validation of a mixed-tissue oligonucleotide DNA microarray for Atlantic bluefin tuna, *Thunnus thynnus* (Linnaeus, 1758). *BMC Genomics*, 16, 1-17.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C. *et al.*, 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40, 2-12.
- Van Oosterhout, C., Hutchinson, W.F., Wills, D.P. M., Shipley, P., 2004. micro-checker, software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4, 535-538.
- Vella, A., Vella, N., Karakulak, F.S., Oray, I., Garcia-Tiscar, S. *et al.*, 2016. Population genetics of Atlantic bluefin tuna, *Thunnus thynnus* (Linnaeus, 1758), in the Mediterranean, implications for its conservation management. *Journal of Applied Ichthyology*, 32, 523-531.
- Vinas, J., Gordo, A., Fernández-Cebrián, R., Pla, C., Vahdet, Ü. *et al.*, 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.
- Waples, R.S., 2015. Testing for Hardy-Weinberg Proportions, Have We Lost the Plot? *Journal of Heredity*, 106, 1-19.
- Ward, R.D., 1995. Population genetics of tunas. *Journal of Fish Biology*, 47, 259-280.
- Wren, J.D., Forgacs, E., Fondon, J.W., Pertsemlidis, A., Cheng, S.Y. *et al.*, 2000. Repeat Polymorphisms within Gene Regions, Phenotypic and Evolutionary Implications. *The American Journal of Human Genetics*, 67, 345-356.

Supplementary data

The following supplementary information is available on line for the article:

Fig. S1: Geographical location of the tuna facility in the Adriatic Sea (Croatia).

Fig. S2: Most common EST-SSR motifs found for the Atlantic bluefin tuna *T. thynnus* per each repeat category. Red bars represent motif counts in perfect repeats only, while light grey bars represent counts of specific motifs in perfect and compound formation combined.

Fig. S3: Estimated population structure for the Atlantic bluefin *T. thynnus* sampled during two consecutive farming cycles at the tuna facility in the Adriatic Sea. a) Bar-plot showing the results of STRUCTURE analysis. Each individual of *T. thynnus* is represented by a vertical line, and the colour indicates the coefficient of the relationship (scale at left) to each cluster (K); b) Plot of the Log posterior probability of data for each value of K.

Table S1. Parameters of genetic diversity of four Atlantic bluefin tuna *T. thynnus* groups sampled during two consecutive farming cycles at the tuna facility in the Adriatic Sea.

Table S2. Pairwise F_{ST} values among two consecutive farming cycles of the Atlantic bluefin tuna *T. thynnus* sampled at the tuna facility in the Adriatic Sea.

Table S3. Results of BayeScan analysis performed to identify the newly developed EST-SSR loci under natural selection.

Table S4. Gene ontology and the description of gene function of each newly developed microsatellite marker derived from EST libraries of the Atlantic bluefin tuna *T. thynnus*, sampled during two consecutive farming cycles at the tuna facility in the Adriatic Sea.