"This is the **peer reviewed version** of the following article: Albaina, A., Iriondo, M., Velado, I., Laconcha, U., Zarraonaindia, I., Arrizabalaga, H., Pardo, M.A., Lutcavage, M., Grant, W.S. and Estonba, A. (2013), Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure. Anim Genet, 44: 678-692, which has been published in final form at https://doi.org/10.1111/age.12051. This article may be used for non-commercial purposes in accordance with **Wiley Terms and Conditions for Use of Self-Archived Versions**. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited."

SNP discovery in albacore and Atlantic bluefin tuna provides insights into world-wide population structure

A. Albaina, M. Iriondo, I. Velado, U. Laconcha, I. Zarraonaindia, H. Arrizabalaga, M. A. Pardo, M. Lutcavage, W. S. Grant, A. Estonba

Anim Genet, 44: 678-692.

DOI: 10.1111/age.12051

SNP discovery in albacore and Atlantic bluefin tuna provides insights into world-wide population structure

A. Albaina[1], M. Iriondo[1], I. Velado[1], U. Laconcha[1,2], I. Zarraonaindia[1], H. Arrizabalaga[2], N. Goñi[2], M. Lutcavage[3], W. S. Grant[4], A. Estonba[1]*

[1]Genetika, Antropologia Fisikoa eta Animalien Fisiologia Saila; Zientzia eta Teknologia Fakultatea, Euskal Herriko Unibertsitatea (UPV/EHU), P.O. Box 48080, Leioa, Spain.

[2] AZTI Tecnalia. Herrera Kaia Portualdea z/g, P.O. Box 20110, Pasaia, Gipuzkoa, Spain.

[3] Environmental Conservation Department, University of Massachusetts Amherst, Marine Fisheries Institute, P.O. Box 3188, Gloucester, MA 01931, USA

[4]Commercial Fisheries Division, Alaska Department of Fish and Game, Anchorage, AK 995189, USA.

* Corresponding author (also author to whom reprint requests should be addressed): Andone Estonba, (andone.estonba@ehu.es) Tel:+0034 946015517, Fax: +0034 946013145

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

24 ## Summary

25 The optimal management of the commercially important, but over-exploited,
26 pelagic tunas, albacore (*Thunnus alalunga* Bonn., 1978) and Atlantic bluefin
27 tuna (*T. thynnus* L., 1758), requires a better understanding of population
28 structure than has been provided by previous molecular methods. Despite
29 numerous studies of both species, their population structures remain
30 controversial. This study reports the development of single nucleotide
31 polymorphisms (SNPs) in albacore and Atlantic bluefin tuna (BFT) and the
32 application of these SNPs to survey genetic variability across the geographic
33 ranges of these tunas. A total of 616 SNPs were discovered in 35 albacore
34 tuna by comparing sequences of 54 nuclear DNA fragments. A panel of 53
35 SNPs yielded values of F_{ST} ranging from 0.0 to 0.050 between samples after
36 genotyping 460 albacore collected throughout the distribution of this species.
37 No significant heterogeneity was detected within oceans, but between-ocean
38 comparisons (Atlantic, Pacific and Indian oceans along with Mediterranean
39 Sea) were significant. Additionally, a 17 SNPs panel was developed in Atlantic
40 BFT by cross-species amplification in 107 fish. This limited number of SNPs
41 discriminated between samples from the two major spawning areas of Atlantic
42 BFT (F_{ST} = 0.116). The SNP markers developed in this study can be used to
43 genotype large numbers of fish without the need for standardizing alleles
44 among laboratories.

45

46 **Keywords** *Thunnus alalunga*, *Thunnus thynnus*, Single Nucleotide
47 Polymorphism (SNP), SNP discovery, population genetics, fisheries
48 management.

For Peer Review

Introduction

Molecular genetics has led to considerable progress in understanding the ecologies of marine species by providing new insights into the demographic and evolutionary dynamics of wild populations (Hauser & Carvalho 2008). Genetic markers are widely used to identify stocks, to estimate mixed stocks in a fishery, to monitor genetic diversity within populations and to measure connectivity between populations, among many other applications (e.g. Nielsen *et al.* 2009; Waples & Naish 2009). These studies have overturned the classic notion that large marine populations are genetically homogeneous with limited local adaptation by showing extensive genetic population structure in many marine species (reviewed in Hauser & Carvalho 2008). Moreover, genetic structuring has been reported even across small spatial scales (e.g. Knutsen *et al.* 2003; Jørgensen *et al.* 2005; Knutsen *et al.* 2007; Knutsen *et al.* 2011). However, data are still rare for the vast majority of highly exploited species, even though a large number of studies on genetic population structure of marine fish have been published in the past decades. Genetic studies are needed to improve the management of species for which stock structure and migration patterns are still unclear. An improper management of these fishery resources can lead to the extirpation of small independent stocks or to the under-utilization of large stocks.

This study focuses on two widely distributed pelagic tunas, albacore (*Thunnus alalunga* Bonn., 1978) and Atlantic bluefin tuna (*T. thynnus* L., 1758). Albacore is one of the smallest tunas and Atlantic bluefin tuna (BFT) one of the largest in the family Scombridae. While albacore is a widely distributed species, inhabiting both temperate and tropical pelagic waters of

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

74 all oceans, the distribution of Atlantic BFT is limited to the North Atlantic
75 Ocean and Mediterranean Sea (e.g. Nakamura 1969; Collette & Nauen 1983;
76 Fromentin & Fonteneau 2001). Both species coexist in the Mediterranean Sea.
77 Harvests of these species are large and of high economical value, especially
78 Atlantic BFT, which is sold for high prices in Japanese fish markets
79 (Magnuson *et al.* 1994). An important Atlantic BFT aquaculture industry,
80 based on the fattening of locally collected fish in floating cages, has
81 developed in the Mediterranean. Moreover, these tunas' life history traits
82 make them susceptible to collapse under continued excessive fishing
83 pressure, as their population growth rates are low (De Roos & Persson 2002).
84 Critical life-history traits include long life spans, large body sizes, late sexual
85 maturity (around 4-5 years but up to 8 years for the Western Atlantic BFT),
86 geographically restricted spawning sites, and relatively short spawning
87 periods of 1 or 2 months (Fromentin & Fonteneau 2001; Fromentin & Powers
88 2005; Rooker *et al.* 2007; Fromentin 2009; Juan-Jordá *et al.* 2011).

89 Since stocks of albacore and Atlantic BFT are currently overexploited, an
90 urgent need exists to improve conservation and management efforts,
91 including the development of alternative methods of population assessment
92 (Juan-Jordá *et al.* 2011; Collette *et al.* 2011). The management of these tunas
93 has to be at the population level, because the extent and dynamics of
94 population structuring underlies the resilience and sustainability of harvested
95 populations. Previous studies of population structure have used four classes
96 of molecular markers: allozymes, DNA blood groups, mitochondrial RFLPs
97 and microsatellite loci (Chow & Ushiama 1995; Yeh *et al.* 2007; Takagi *et al.*
98 1999; Takagi *et al.* 2001; Ely *et al.* 2002; Pujolar *et al.* 2003; Arrizabalaga *et al.*

For Peer Review

1
2
3 99 2004; Viñas *et al.* 2004; Nakadate *et al.* 2005; Arrizabalaga *et al.* 2007;
4
5 100 Carlsson *et al.* 2007; Boustany *et al.* 2008; Riccioni *et al.* 2010; Davies *et al.*
6
7 101 2011; Viñas *et al.* 2011). Despite these efforts to describe stock structure in
8
9 102 albacore and Atlantic BFT, the population structures of these species remain
10
11 103 controversial (Arrizabalaga *et al.* 2004; Walli *et al.* 2009; Galuardi *et al.* 2010).
12
13 104 Presently, albacore populations are divided into six management units
14
15 105 (namely the Mediterranean, North Atlantic, South Atlantic, Indian Ocean,
16
17 106 North Pacific and South Pacific stocks) and Atlantic BFT into two units
18
19 107 (namely the Western Atlantic stock and the Eastern Atlantic and
20
21 108 Mediterranean stock). However, the results of population surveys based on
22
23 109 microsatellite variability illustrate that these management units might not be
24
25 110 consistent with the genetic structures of both species (Riccioni *et al.* 2010;
26
27 111 Davies *et al.* 2011; Viñas *et al.* 2011).
28
29
30 112 Two factors explain the current lack of consensus on genetic structure.
31
32 113 First, none of the previous genetic studies included samples over the entire
33
34 114 distributional areas of these tunas. Second, none of the previous studies used
35
36 115 large numbers of molecular markers, such as multiple single nucleotide
37
38 116 polymorphisms (SNPs), which can be assayed rapidly in large numbers of fish
39
40 117 to yield high statistical power to test population-level genetic hypotheses (e.g.
41
42 118 Ogden 2011; Helyar *et al.* 2011).
43
44
45 119 The goals of the present study were to develop SNP markers in albacore
46
47 120 and Atlantic BFT and to use these markers to make a preliminary survey
48
49 121 geographic variability among populations over the geographic ranges of these
50
51 122 tunas (Table 1, Fig. 1). The development of SNP markers will aid in the ability
52
53
54
55
56
57
58
59
60

For Peer Review

to conduct collaborative studies among laboratories without the need for standardizing alleles. Further, the development of a large number of markers will increase the power of genetic analysis to detect the small differences among populations that are expected in high gene-flow species, such as the tunas.

Material and methods

Sample collection

Samples of muscle, fin or heart tissue from 460 albacore were collected at 8 locations (representing samples from feeding grounds and including a mixture of juveniles and adults) over the distribution of the species (Table 1; Fig. 1). Additional tissue samples from 107 Atlantic bluefin tuna (BFT) were collected from 3 locations: Western Atlantic, Bay of Biscay and Mediterranean Sea (Table 1). While the Bay of Biscay sample included a mixture of juveniles and adults from a feeding ground, the samples of Western Atlantic and Mediterranean Sea were composed of young-of-the-year (YOY) individuals incapability of trans-oceanic migration and, thus, represented reference samples for both spawning areas (Rooker *et al.* 2008). Samples were either frozen and stored at -20º C, or were preserved in 96% ethanol at 4º C. DNA was extracted from tissues using the DNeasy 96 Blood & Tissue Kit (Qiagen, Hilden, Germany) and quantified using a NanoDrop 1000™ spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) prior to storage at -20ºC for further analysis.

For Peer Review

146

147 SNP discovery in albacore (*Thunnus alalunga*)

148 Single nucleotide polymorphism (SNP) discovery via comparative sequencing
149 of nuclear DNA fragments was performed on 35 albacore from widely
150 separated areas. Specifically, five fish were selected from each of the six
151 currently hypothesized stocks, Mediterranean, Indian Ocean and Northern
152 and Southern parts of Atlantic and Pacific Oceans, except for the North
153 Atlantic, where 10 fish were used. SNPs were mined from 54 nuclear DNA
154 fragments (Table S1) and were amplified with primers designed with Primer3
155 (Rozen & Skaletsky 2000). In approach I, EPIC primers (Exon-Priming, Intron-
156 Crossing primers; Slate *et al.* 2009) for 19 DNA fragments (average length
157 318 bp) were obtained from the literature (references in Table S1). The
158 primers for the remaining 35 DNA fragments were designed from the
159 alignment of sequences from publically available databases (GenBank and
160 Ensembl). In approach II, 17 pairs of degenerate primers were designed from
161 several teleost sequences (average length 420 bp). In approach III, 18 pairs
162 of primers were designed from *Thunnus* DNA sequences (average length 487
163 bp).

164 Tables S1 and S2 show fragment amplification specifications. Briefly, we
165 used conventional polymerase chain reaction (PCR) to amplify 30 of 54
166 fragments and used touchdown (TD) methodology to amplify the remaining
167 fragments. Reactions were carried out in a thermo-cycler, GeneAmp®PCR
168 System 2700, GeneAmp®PCR System 9700 or Veriti 96 well Thermal Cycler
169 (Applied Biosystems, Foster City, CA), and iCycler (Biorad Laboratories,

For Peer Review

1
2
3 170 Hercules, CA). The purified PCR products were sequenced in one direction
4
5 171 with either the forward or the reverse PCR primer on an Applied Biosystems
6
7 172 (ABI) 3130X capillary electrophoresis Analyzer, with ABI BigDye Terminator
8
9 173 version 3.1 Chemistry (Applied Biosystems). Base-calling from
10
11 174 chromatograms was performed using SeqScape v2.5 (Applied Biosystems).
12
13 175 The BLASTN algorithm was used to verify that the target locus had been
14
15 176 amplified.
16
17
18 177 Nucleotide differences at a site in aligned sequences were considered to
19
20 178 be a SNP, but only when flanking sequences had high quality and the
21
22 179 alternative nucleotide was present in at least two individuals (out of 35; see
23
24 180 above). After filtering for SNPs matching the technical requirements of the
25
26 181 assays, we gave priority to selecting at least one SNP per fragment for a total
27
28 182 of 128 SNPs for genotyping the 460 albacore with TaqMan® OpenArray®
29
30 183 technology (Applied Biosystems). Moreover, a SNP was chosen if it was not
31
32 184 located near the ends of the sequence and if it was more than two bases
33
34 185 away from any other SNP (*Custom TaqMan Genomic Assays Protocol*
35
36 186 *Submission Guidelines*).
37
38
39 187
40
41
42 188 Cross-species amplifications in Atlantic bluefin tuna (*Thunnus thynnus*)
43
44
45 189 SNP discovery in Atlantic BFT was performed by cross-species amplification
46
47 190 of the 128 SNPs selected for genotyping in albacore. The 128 SNPs were
48
49 191 genotyped with TaqMan ® OpenArray ® technology in 107 Atlantic BFT
50
51 192 samples. The same criteria used to validate SNPs in albacore were used to
52
53 193 validate individual SNPs in Atlantic BFT.
54
55
56
57
58
59
60

For Peer Review

1
2
3 194
4
5
6 195 Population analysis
7
8
9 196 Figure 2 outlines the procedures used in this study, starting from SNP
10
11 197 discovery in albacore, to the selection of a subset of SNPs (panel) with origin
12
13 198 assignment (including loci under selection) or demographic analysis.
14
15 199 Genotyping call rate and minor allele frequencies (MAF) were obtained for
16
17 200 each SNP locus using AutoCaller™ 1.1 (Applied Biosystems). A SNP was
18
19 201 validated if the polymorphism remained in the genotyping results and could be
20
21 202 reliably scored. SNPs with unclear genotypes and those with a call rate below
22
23 203 70% were discarded. Deviation from Hardy-Weinberg expectations (HWE)
24
25 204 was evaluated for each locus and each sampling location (Fisher's exact test
26
27 205 in GENEPOP 4.0; Rousset 2008). The exact test for linkage disequilibrium (LD),
28
29 206 as implemented in GENEPOP, was used to detect disequilibria between SNPs
30
31 207 on the same DNA fragment and between SNPs on different fragments; $P <$
32
33 208 0.001 was used as critical probability for LD tests. SNP loci exhibiting
34
35 209 significant LD were phased into haplotype blocks using the Bayesian
36
37 210 statistical method implemented in PHASE 2.1 for each sample independently
38
39 211 (Stephens et al. 2001).
40
41 212 Expected heterozygosity (H_e), F_{IS} and F_{ST} were estimated with FSTAT 2.9.3
42
43 213 (Goudet 2001). SNPs exhibiting significant departures from HWE ($P < 0.001$)
44
45 214 in one or more samples were deemed unsuitable for estimating population
46
47 215 structure and were discarded. When two or more loci occurred on the same
48
49 216 DNA fragment, including both haplotype blocks and individual SNP loci, the
50
51 217 locus with the largest H_e was selected to ensure independence among the
52
53
54
55
56
57
58
59
60

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

218 markers. Loci with large heterozygosities provide more statistical power for
219 population structure analysis than loci with small heterozygosities (Haasl &
220 Payseur 2011; Morin *et al.* 2004; Rosenberg *et al.* 2003; Ryman *et al.* 2006).
221 The SNP panel consisted of these filtered SNPs.

222 We searched for candidate loci under selection (outlier loci) among the
223 remaining loci using the Bayesian likelihood method, as implemented in
224 BAYESCAN 2.0 (Foll & Gaggiotti 2008) and LOSITAN (Beaumont & Nichols
225 1996; Antao *et al.* 2008). Loci identified by BAYESCAN and LOSITAN as outliers
226 were removed from the SNP panel (Richter-Boix 2011). Briefly, two alternative
227 models were defined in BAYESCAN (including and excluding the effect of
228 selection) and their respective posterior odds (PO), the ratio of the posterior
229 probabilities of the two models, were calculated using 20 pilot runs of 5000
230 iterations and an additional burn-in of 50 000 iterations, for a total of 100 000
231 iterations (sample size of 5000 and thinning interval of 10). While only loci
232 with positive $\log_{10}$PO values were considered, strong support for selection
233 was limited to loci with $\log_{10}$PO > 1 ($P < 0.09$; BAYESCAN 2.0 user manual).
234 On the other hand, LOSITAN identifies outlier loci from a plot of heterozygosity
235 versus F_{ST}. To avoid an upward bias in quantiles, LOSITAN was first run for all
236 loci to estimate the mean neutral F_{ST}. Loci that were outside the 0.99
237 confidence interval were removed using only the putative neutral loci to
238 compute again the mean neutral F_{ST}. A second run was then conducted with
239 all loci using the last computed mean.

240 POWSIM 4.0 (Ryman *et al.* 2006) was used to estimate the statistical power
241 required to detect various levels of differentiation with the SNPs developed for

For Peer Review

For Peer Review

242 albacore and Atlantic BFT in this study. Burn-in consisted of 1000 steps
243 followed by 100 batches of 1000 steps. X^2 and Fisher's probabilities were
244 used to test the significance of an F_{ST} value for each replicate run. The
245 number of significant F_{ST} values in 1000 replicate simulations provided an
246 estimate of statistical power for a given level of divergence, which was
247 controlled by allowing frequencies to drift for a given number of generations.
248 Simulated effective populations sizes (N_e) equalled 2000 fish. For albacore, all
249 41 SNPs + blocks, 9 blocks only and 32 SNPs only were examined separately
250 for statistical power for divergences ranging from F_{ST} = 0.0005 to 0.016.

251 Population structure was estimated with F_{ST} between samples and with
252 Bayesian individual assignments implemented in STRUCTURE 2.3.3 (Pritchard
253 *et al.* 2000). Pairwise F_{ST} (Weir & Cockerham 1984) values were estimated
254 with FSTAT 2.9.3 (Goudet 2001) and with globally corrected *p*-values. FSTAT
255 combines individual locus *p*-values weighting them according to their
256 polymorphism level (Petit *et al*. 2001). Population groups were defined by
257 non-significant values of mean F_{ST} between samples and by significant values
258 of F_{ST} with other populations (e.g. Waples & Gaggiotti 2006). STRUCTURE
259 uses a Bayesian method to identify the number of clusters (K) of related
260 individuals using HWE and gametic disequilibria among multilocus genotypes.
261 We used the admixture model, independent allele frequencies between
262 populations and the LOCPRIOR option. We compared log-likelihood ratios in 10
263 STRUCTURE runs for values of K = 1 to 10 (Pritchard *et al.* 2000). Each run
264 consisted of 10 000 iterations with a burn-in of 10 000. CLUMPP 1.1.2
265 (Jakobsson & Rosenberg 2007) was used to determine optimal assignments
266 of individual to clusters by maximizing the similarity between pairs of

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

267 genotypes in different replicates. These groupings were visualized with
268 DISTRUCT 1.1 (Rosenberg 2004). Outlier loci that were not used to estimate
269 F_{ST}, were added for the STRUCTURE analyses as the latter does not require
270 neutral markers, unlike the F_{ST} analysis.

271

272 Results

273

274 SNP discovery and validation in albacore tuna

275 Fifty-four fragments of nuclear DNA were sequenced in 35 albacore tunas
276 (Table S1). Thirty-five of the 54 fragments showed a high degree of homology
277 with the orthologous sequences in several fishes (BLASTN; E-value < 10^{-5}),
278 including 14 best hits corresponding to other species of *Thunnus* (Table S3).
279 A total of 616 SNPs were discovered, in which an alternative allele was
280 present in at least 2 individuals, with a mean of 11.4 (SD ±10) SNPs per
281 fragment and a ratio of 1/36 bp. At least 1 SNP was present in each DNA
282 fragment, except for a fragment coding for metallothionein (*MT*). A total of 195
283 SNPs were present in fragments amplified with EPIC primers (approach I);
284 182 SNPs were found in fragments amplified with degenerate teleost primers
285 (approach II); and 239 SNPs were present in DNA fragments amplified with
286 *Thunnus* spp. primers (approach III). In addition to SNPs, 19 small indels, 1–5
287 nucleotides in length, were found in 14 fragments with a majority
288 corresponding to mono- or bi-nucleotide indels (84.2%).

For Peer Review

A total of 128 candidate SNPs were selected to genotype the albacore population samples and included 32, 47 and 49 SNPs selected from fragments obtained with approaches I, II and III, respectively. A total of 23 (18%) SNPs failed to amplify for routine genotyping, and 2 SNP loci failed to exceed call rates above 70%. Another 24 loci among the remaining 103 SNPs could not be reliably scored. The remaining 79 validated SNPs (Table S4) showed a mean call rate of 91 ±5% and an average minor allele frequency (MAF) of 0.17 ±0.14 (range 0.001–0.489). Validation success was 72%, 66% and 51% for approaches I, II and III, respectively.

SNP panel for population genetic studies in albacore

Twelve of the 79 validated SNPs departed significantly from HWE in one or more sampling locations and were discarded. The remaining 67 SNPs were tested for linkage disequilibrium (LD). No SNPs were found in LD between DNA fragments; however, 21 SNPs were in LD within fragments and were phased into 9 haplotype blocks (Table S5). After selection of only one independent locus per DNA fragment (see Methods), the final panel of markers included 41 independent markers: 32 individual SNPs, plus 9 haplotype blocks (53 SNPs in total; Table 2). Analysis of these SNPs with BAYESCAN showed no candidate loci influenced by selection. However, LOSITAN detected 3 SNP loci, *HIF1A4, MTF1* and *MYC*, with significantly larger genetic divergences than expected from neutrality. These loci were assumed to be embedded in candidate genes under divergent selection or in DNA fragments linked to genes influenced by selection. In contrast, the low

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

313 level of divergence for *PRDX2* suggested that balancing selection may be
314 influencing this locus, or a nearby locus. Interestingly, *MTF1* and *PRDX2*
315 sequences showed the highest homology with orthologs of the teleosts *Fugu*
316 *rubripes* and *Thunnus maccoyii*, respectively (Table S3), but no homologies
317 were found for *HIF1A4* and *MYC* in the teleosts tested. Overall, the average
318 expected heterozygosity over loci was H_e = 0.278 ±0.201 and the average
319 inbreeding coefficient was F_{IS} = 0.032 ±0.085 for the 41 SNP loci in the final
320 panel (Table 2).

321

322 Genetic structure in albacore tuna

323 POWSIM simulations indicated that the 41 independent markers (32 SNPs and
324 9 haplotype blocks) together were able to detect significant differences among
325 samples with F_{ST} = 0.002 in about 84% of the tests and with $F_{ST} \geq 0.004$ in
326 100% of the tests (Table 3). The nine multi-allelic haplotype blocks alone
327 provided about the same amount of statistical power for detecting differences
328 among populations as did the 32 individual SNPs. Each set of markers was
329 able to detect values of $F_{ST} \geq 0.004$ in at least 97% of the tests. Fisher's
330 method for detecting significant differences among samples provided less
331 statistical power than did the X^2 tests.

332 Analyses using the 32 SNPs and nine haplotype blocks together revealed
333 an overall F_{ST} = 0.017 ±0.003 ($P < 0.05$) among the eight albacore sampling
334 locations. Levels of divergence were not significant between sampling
335 locations within oceans, but were significant between oceans (Table 4).
336 Samples from the NE Atlantic (IRE and BIS) were not significantly different

For Peer Review

1
2
3 337 from each other or from a sample from the SE Atlantic (SEA). Likewise, no
4
5 338 divergence was detected between the three samples from the Pacific (NP,
6
7 339 SEP, SWP). However, all comparisons between oceans were significant,
8
9 340 yielding four differentiated genetic entities: 1) NE Atlantic, 2) Mediterranean
10
11 341 Sea, 3) Pacific Ocean and 4) Indian Ocean). Fish from the western
12
13 342 Mediterranean (BAL) showed the highest divergence from the other locations
14
15 343 with an average F_{ST} = 0.034 (range: 0.021–0.050). Fish from the Indian
16
17 344 Ocean (IN) were most divergent from the Atlantic and Mediterranean sample
18
19 345 locations (mean: F_{ST} = 0.030), but less divergent from Pacific Ocean samples
20
21 346 (F_{ST} = 0.010). The individual Bayesian clustering (STRUCTURE) indicated the
22
23 347 largest likelihood of population structure was K = 3, placing samples into three
24
25 348 groups: Mediterranean Sea, Atlantic Ocean, and Indo-Pacific (Fig. 3).
26
27 349 Analysis with K = 4 showed that Indian Ocean albacore were differentiated to
28
29 350 a small degree from Pacific Ocean albacore, as reflected in the distribution of
30
31 351 F_{ST} values between these locations.
32
33
34 352
35
36
37 353 SNP panel for Atlantic bluefin tuna
38
39
40 354 Primers for the 128 albacore SNPs were used in cross-species reactions to
41
42 355 develop SNPs in BFT. Although 32 SNPs successfully amplified, 9 SNPs had
43
44 356 low call rates (below 70%), had unclear genotypes, or were not polymorphic,
45
46 357 and hence were discarded. This yielded 23 validated SNPs (18%) for BFT.
47
48 358 Out of these 23, 18 had been validated also when genotyping albacore while
49
50 359 the other five SNPs (*GNRH3-1-107*, *LDHB-129*, *CYCS-161*, *CS3-118*, and
51
52 360 *OPC02*-45; Table S6) were reported as reliably scored and polymorphic only
53
54
55
56
57
58
59
60

For Peer Review

1
2
3 361 in BFT. The validation success rates in BFT for SNP that had been
4
5 362 discovered in albacore by approaches I, II and III were 28% (9 of 32 SNPs),
6
7 363 15% (7 of 47 SNPs) and 14% (7 of 49 SNPs), respectively. Additionally, one
8
9 364 SNP was discarded from the final panel due to a significant deviation from
10
11 365 HWE in at least one sample (Table S6). Tests for LD between the remaining
12
13 366 22 BFT SNPs detected two cases of two linked SNP loci in a single fragment
14
15 367 (Table S5). Linked loci were phased into haplotypes. Significant LD was not
16
17 368 detected among SNPs on different DNA fragments. A final set of 15
18
19 369 independent markers, 13 individual SNPs and 2 haplotype blocks, were
20
21 370 suitable for surveys of BFT populations (17 SNPs in total; Table 5).
22
23
24 371 Average expected heterozygosity among the 15 loci was H_e = 0.272
25
26 372 ±0.178, and the average inbreeding coefficient was F_{IS} = 0.096 ±0.133. The
27
28 373 low number of markers tested in BFT precluded the use of outlier detection
29
30 374 software, as a larger number of SNPs are required to obtain a reliable
31
32 375 estimate of the neutral expectation from which the outliers are detected.
33
34 376 Therefore, all 15 loci were used to estimate population structure with F_{ST} and
35
36 377 STRUCTURE.
37
38
39 378
40
41
42
43 379 Genetic structure in Atlantic bluefin tuna
44
45
46 380 The POWSIM simulations and the average frequencies for the 13 SNPs and 2
47
48 381 blocks indicated that statistical power increased from only 0.064 for F_{ST} =
49
50 382 0.001 (Fisher's method of determining significance), to 0.579 for F_{ST} = 0.010
51
52 383 and to 1.0 for F_{ST} = 0.080 (Table 6). The 15 BFT SNPs showed significant
53
54 384 overall differentiation among populations in the NW Atlantic, NE Atlantic and
55
56
57
58
59
60

For Peer Review

Mediterranean (F_{ST} = 0.029 ±0.024, P < 0.05). Populations in the Bay of Biscay and the Mediterranean differed from the populations in the NW Atlantic (BB–NWA, F_{ST} = 0.120 ±0.091, P < 0.01; MED–NWA 0.116 ±0.078, P < 0.01). However, the Bay of Biscay and the Mediterranean did not differ significantly from each other (F_{ST} = 0.004 ±0.007). STRUCTURE indicated that the three samples most likely represented two populations (K = 2), 1) NW Atlantic and 2) Bay of Biscay and Mediterranean (Fig. 4).

Discussion

Our study outlines the development and validation of SNPs in the genomes of albacore and Atlantic bluefin tuna and provides a basis for defining discrete stocks to aid in the commercial harvests of these species. We developed *de novo* 53 SNPs for albacore and 17 cross-species SNPs for Atlantic bluefin tuna. While the focus of the study was on the development of SNP markers, the distributions of our samples allow a preliminary analysis of large-scale population structure. Several variables influence the ability of a set of molecular markers to detect genetic differences between populations. In addition to the well known effects of sample size on power, the geographical extent of a set of samples is crucial to describing population structure.

SNP discovery in Albacore tuna

Our search for variable nucleotide sites yielded 128 SNPs. Of these, 79 (62%) could be validated and were selected for routine genotyping. From

For Peer Review

1
2
3 408 these, we selected a final panel of 53 SNPs distributed over 41 loci. The 62%
4
5 409 validation success rate is similar to SNP validation success rates in other
6
7 410 studies of fishes, including *Gadus morhua* (54%) (Moen *et al.* 2008),
8
9 411 *Oncorhynchus nerka* (39%), *O. keta* (54%) and *O. tshawytscha* (64%) (Smith
10
11 412 *et al.* 2005) and *Engraulis encrasicolus* (59%) (Molecular Ecology Resources
12
13 413 Primer Development Consortium *et al.* 2012).
14
15
16 414 Our results allow an assessment of levels of polymorphism in the albacore
17
18 415 genome. A total of 616 SNPs were discovered with an overall ratio of 1 SNP
19
20 416 for each 36 base pairs. This value indicates higher levels of polymorphism in
21
22 417 albacore than has been reported, for example, in the salmonids,
23
24 418 *Oncorhynchus keta* (1/175 bp), *O. nerka* (1/242 bp) and *O. tshawytscha*
25
26 419 (1/301 bp) (Smith *et al.* 2005), *Salmo salar* (1/586 bp overall, Hayes *et al.*
27
28 420 2004) or in intronic (1/405 bp) and exonic (1/1448 bp) regions (Ryynänen &
29
30 421 Primmer 2006). The large value for albacore, however, is similar to that for
31
32 422 European anchovy (*Engraulis encrasicolus*) (1/54 bp, Zarraonaindia 2011).
33
34 423
35
36
37
38 424 SNP development in Atlantic bluefin tuna
39
40
41 425 We used primers for the 128 SNPs detected in albacore to search for SNPs in
42
43 426 Atlantic BFT and achieved an overall validation success rate of 18% (23
44
45 427 SNPs). This success rate indicates that the regions flanking these SNPs
46
47 428 contain highly conserved sequences so that the SNP primers developed for
48
49 429 albacore also work in Atlantic BFT. When successful, cross-species
50
51 430 amplifications are a cost-effective method of SNP discovery (Malhi *et al.*
52
53 431 2011). Successful cross-species amplifications are thought to be free of
54
55
56
57
58
59
60

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

432 ascertainment bias, especially for SNPs embedded in conserved sequences
433 near or within coding regions (Malhi et al. 2011). Cross-species amplifications
434 have generally not been used to develop SNP markers in fish, because of
435 generally low success rates in other vertebrates (e.g. Seeb *et al.* 2011; Miller
436 *et al.* 2011). For example, only about 1% of the nearly 50 000 SNP loci
437 developed for domestic sheep were polymorphic in two related ungulates
438 (Miller *et al.* 2011). In a panel of a similarly large number of SNPs designed
439 for cattle, only about 2.5 and 3% of the cross-species amplifications were
440 successful in two lines of European bison and two species of antelopes,
441 respectively (Kaminski *et al.* 2012; Ogden *et al.* 2012). However, the species
442 used in these cross-amplification attempts were distantly related to one
443 another, unlike albacore and Atlantic BFT, which are phylogenetically closely
444 related (Chow & Kishino 1995; Chow *et al*. 2006) and are known to hybridize
445 (Viñas & Tudela 2009). Our results are similar to those of Mahli *et al.* (2011),
446 who reported up to 30% cross amplification successes between related
447 species of Old World monkeys.

448

449 Stock structure

450 The statistical power of the SNPs is influenced by several factors. First,
451 the power of the markers to detect population structure increases substantially
452 when locus heterozygosities are larger than $H_e \geq 0.2$ (Haasl & Payseur 2011).
453 In our study, about 60% of the SNP loci in both species showed $H_e > 0.2$.
454 Second, while individual SNPs show less power than multi-allelic
455 microsatellite loci (Ryman *et al*. 2006; Haasl & Payseur 2011), combining

For Peer Review

1
2
3 456 physically linked SNPs into haplotype blocks increases statistical power
4
5 457 (Gattepaille & Jakobsson 2012). For example, Mesnick *et al.* (2011) retrieved
6
7 458 the same population structure for sperm whales (*Physeter macrocephalus*)
8
9 459 with 6 microsatellites, or with 36 SNPs representing 24 independent markers,
10
11 460 including 8 haplotype blocks. In our study, POWSIM simulations showed that
12
13 461 haplotype 9 blocks of linked SNPs in albacore had about the same statistical
14
15 462 power as 32 individual SNPs. In any case, haplotype blocks and SNPs
16
17 463 together yielded a type II error rate (failure to detect a real difference) for both
18
19 464 species of 0% for divergences of F_{ST} = 0.004 or greater.
20
21 465
22
23 466 Stock structure in Atlantic bluefin tuna
24
25 467 The International Commission for the Conservation of Atlantic Tuna (ICCAT)
26
27 468 currently manages Atlantic BFT as two stock units that are divided by the mid-
28
29 469 ocean longitude at 45°W (Fromentin & Powers 2005). This geographically
30
31 470 defined stock concept is supported by two spatially separated spawning areas,
32
33 471 one in the Mediterranean and the other in the Gulf of Mexico (Rooker *et al.*
34
35 472 2007). However, tagging and microchemical studies suggest a more complex
36
37 473 stock distribution pattern (Rooker *et al.* 2008; Walli *et al*. 2009; Galuardi *et al*.
38
39 474 2010). Foraging aggregations in the NE Atlantic and elsewhere may
40
41 475 potentially originate from both main spawning areas, because trans-Atlantic
42
43 476 migrations in both directions have been documented (Magnuson *et al*. 1994;
44
45 477 Mather *et al*. 1995; Lutcavage *et al*. 1999; Block *et al*. 2001; Block *et al*. 2005;
46
47 478 Rooker *et al*. 2006). However, homing to natal spawning areas (Boustany *et*
48
49 479 *al*. 2008; Block *et al*. 2005; Teo *et al*. 2007; Dickhut *et al*. 2009) appear to
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3 480 isolate the two major groups. Recent studies with mtDNA and microsatellite
4
5 481 markers show that populations in the Mediterranean Sea may also be
6
7 482 structured into partially isolated subpopulations (Riccioni *et al*. 2010; Viñas *et*
8
9 483 *al*. 2011).
10
11
12
13 484 The Atlantic BFT SNP panel developed in the present study, although
14
15 485 limited in number due to the SNP discovery approach followed, is the
16
17 486 beginning of a valuable tool to improve the management of this overexploited
18
19 487 species. The advent of Next Generation Sequencing (NGS) technologies such
20
21 488 as Roche´s 454 or Illumina´s HiSeq platforms (see Garvin et al. 2010 and
22
23 489 Seeb et al. 2011 for a review) is making possible to discover hundreds-
24
25 490 thousands of SNP type markers in non-model organisms and, further
26
27 491 application of these techniques on BFT (and also in albacore) will improve the
28
29 492 relevance of the present reported tool. The preliminary analysis in the present
30
31 493 study confirms the genetic distinction between the two major spawning areas
32
33 494 on the western and eastern margins of the Atlantic (both F_{ST} and STRUCTURE
34
35 495 results) and is largely consistent with the ICCAT management plan. While
36
37 496 trans oceanic migrations of adults have been documented with tags, finer-
38
39 497 scale population structure may also exist. Our SNP analysis showed that the
40
41 498 foraging area sample from the Bay of Biscay clustered with the Mediterranean
42
43 499 samples indicating the Mediterranean origins of these fish. To our knowledge,
44
45 500 this is the first time a mixed age classes of Atlantic BFT in the NE Atlantic
46
47 501 have been assigned to a spawning area based on DNA data. Previous
48
49 502 insights have come from relatively expensive and laborious tagging
50
51 503 experiments (Rooker *et al.* 2007) and from an allozyme study comparing
52
53
54
55
56
57
58
59
60

For Peer Review

samples from Azores (Mid-Atlantic) and the Mediterranean (Pujolar et al. 2003).

The larger levels of differentiation than expected under neutrality may indicate that some SNP loci are embedded in genes under selection. Although the small number of SNPs for Atlantic BFT precluded tests of neutrality, the remarkably large value of F_{ST} = 0.116 indicating a high level of differentiation between populations of Atlantic BFT may result from directional selection on some SNPs. Markers influenced by directional selection often show higher F_{ST} values among populations of fishes than do neutral markers (e.g. André *et al*. 2011; Ackerman *et al*. 2011; Poulsen *et al*. 2011). Additional SNPs for Atlantic BFT are needed to test hypotheses of selection, to estimate population structure and to identify mixed-stock components in fishery areas.

Stock structure in Albacore tuna

ICCAT, the Indian Ocean Tuna Commission (IOTC), the Western and Central Pacific Fisheries Commission (WCPFC) and the Inter-American Tropical Tuna Commission (IATTC) manage albacore with a six-stock model, which includes 1) Mediterranean Sea, 2) North Atlantic, 3) South Atlantic, 4) Indian Ocean, 5) North Pacific Ocean and 6) South Pacific Ocean. These stocks are based on a limited understanding of spawning areas, the geographical distribution of fisheries, life-history variables and the results of tagging studies (Arrizabalaga *et al*. 2004).

For Peer Review

1
2
3 526 Our analyses resolved the genetic relationships among oceanic
4
5 527 populations. Pairwise F_{ST} values distinguish four albacore groups, 1)
6
7 528 Mediterranean Sea, 2) Atlantic Ocean, 3) Pacific Ocean and 4) Indian Ocean,
8
9 529 with no within-ocean heterogeneity (Table1, Fig. 1). The tests for
10
11 530 heterogeneity did not detect differences between northern and southern
12
13 531 populations within the Pacific or within the Atlantic. The Mediterranean group
14
15 532 appears to be most differentiated from other global populations, and this is in
16
17 533 agreement with previous results for microsatellites (Davies *et al.* 2011). Both
18
19 534 SNPs and microsatellites show that Mediterranean and North Atlantic
20
21 535 populations are partially isolated from each other, and this genetic separation
22
23 536 is consistent with tag-recapture analysis showing limited movement between
24
25 537 the North Atlantic and Mediterranean (Arrizabalaga *et al.* 2004). The SNP
26
27 538 data also showed that the Indian Ocean populations were genetically closer to
28
29 539 Pacific populations than to Atlantic populations. In contrast, a closer
30
31 540 relationship between Indian Ocean and Atlantic fish was reported from blood-
32
33 541 group frequencies (Arrizabalaga *et al.* 2004).
34
35
36 542 Overall, the genetic results together support the existence of at least 4
37
38 543 genetic entities, Mediterranean Sea, Atlantic, Pacific and Indian oceans that
39
40 544 are isolated from one another to some degree. These results support
41
42 545 management plans in which oceanic populations, including the Mediterranean,
43
44 546 are managed separately, but differ from the present use of six management
45
46 547 units (stocks). However, due to our relatively limited sampling of locations and
47
48 548 individuals within ocean basins, the current six-stock management model
49
50 549 represents a conservative approach that reduces the risk of inadvertently
51
52 550 overfishing some populations.
53
54
55
56
57
58
59
60

For Peer Review

Conclusions

We developed 128 SNP markers *de novo* in albacore tuna and used a final panel 53 SNPs (41 SNPs, including 32 individual SNPs and 9 haplotype blocks) to genotyped over 400 individuals collected over the distributional range of the species. Although the coarse scale of sampling limits our inferences about population structure, the results for albacore are largely consistent with previous molecular studies in indicating the existence of at least four albacore populations: Mediterranean Sea, Atlantic, Pacific and Indian oceans. We then used SNP assays developed for albacore to develop 17 validated SNPs in Atlantic bluefin tuna (15 SNPs, including 13 individual SNPs and 2 haplotype blocks). This small number of SNPs discriminated the two major spawning areas of Atlantic BFT in the Gulf of Mexico and Mediterranean Sea, and identified the Mediterranean origin of juveniles foraging in the Bay of Biscay. The additional development of new SNPs will increase the statistical power needed to resolve the population structures of these two overfished tunas.

SNPs have two advantages over other markers, such as allozymes and microsatellites, for the international management of far ranging tunas. First, a large number of SNPs, individually or in haplotype blocks, provides a large amount of statistical power to detect biologically meaningful genetic differences between stocks of highly mobile tunas. High mobility is expected to produce only small differences between stocks. Second, and most importantly, SNP assays can be used in any laboratory without the need for standardizing allelic variants among research or management groups.

For Peer Review

International cooperation in establishing a universal database is essential for the conservation and management of these vulnerable species with stocks that straddle international boundaries. The SNPs developed for both species can provide valuable tools for population management and can additionally be used as markers to trace fishery products.

Acknowledgements

The authors thank I. Miguel and F. Rendo (Sequencing and Genotyping Service of the UPV/EHU, SGIKer) for their technical assistance as well as C. Manzano (UPV/EHU), M. A. Pardo (AZTI-Tecnalia), I. Cancio (UPV/EHU) and G. Heinisch (University of New Hampshire) for their collaboration. We are grateful to J. Areso, D. Fuller, K. Schaefer, V. Allain, S.Y. Yeh, I. Fraile and I. Arregi for providing samples. This work was supported by ATM2010Hegaluze (351BI20090047) ATM2009Hegalabur (351BI20090034) and ATM2008Bonorte (ACM2008BONORTE) projects funded by the Basque Government, and the ACEITUNA (CTM2011-27505) project funded by the Spanish *Ministerio de Economía y Competitividad*. Urtzi Laconcha´s work was supported by a PhD grant by the *Fundación Centros Tecnologicos Iñaki Goenaga*. Sequencing was performed at the General Research Services Unit, SGIKer- University of the Basque Country (UPV/EHU).

Conflict of interest

The authors have declared no potential conflicts.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

599 ## References

600 Ackerman M.W., Habicht C. & Seeb L.W. (2011) Single-Nucleotide
601 Polymorphisms (SNPs) under diversifying selection provide Increased
602 accuracy and precision in mixed-stock analyses of sockeye salmon from
603 the Copper River, Alaska. *Transactions of the American Fisheries Society*
604 **140**, 865–81.

605 André C., Larsson L.C., Laikre L., Bekkevold D., Brigham J., Carvalho G.R.,
606 Dahlgren T.G., Hutchinson W.F., Mariani S., Mudde K., Ruzzante D.E. &
607 Ryman N.(2011) Detecting population structure in a high gene-flow
608 species, Atlantic herring (*Clupea harengus*): direct, simultaneous
609 evaluation of neutral vs putatively selected loci. *Heredity* **106**, 270–80.

610 Antao T., Lopes A., Lopes R.J., Beja-Pereira A. & Luikart G. (2008) LOSITAN:
611 A workbench to detect molecular adaptation based on a FST-outlier
612 method. *BMC Bioinformatics* **9**, 323–7.

613 Arrizabalaga H., Costas E., Juste J., González-Garcés A., Nieto B. & López-
614 Rodas V. (2004) Population structure of albacore *Thunnus alalunga*
615 inferred from blood groups and tag-recapture analyses. *Marine Ecology*
616 *Progress Series* **282**, 245–52.

617 Arrizabalaga H., López-Rodas V., Costas E. & González-Garcés A. (2007)
618 Use of Genetic data to assess the uncertainty in stock assessments due to
619 the assumed stock structure: The case of albacore (*Thunnus Alalunga*)
620 from the Atlantic Ocean. Fisheries Bulletin **105**, 140–46.

621 Beaumont M.A. & Nichols R.A. (1996) Evaluating loci for use in the genetic
622 analysis of population structure. *Proceedings of the Royal Society of*
623 *London. Series B: Biological Sciences* **263**, 1619–26.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

624 Block B.A., Dewar H., Blackwell S.B., Williams T.D., Prince E.D., Farwell C.J.,
625 Boustany A., Teo S.L.H., Seitz A., Walli A. & Fudge D. (2001) Migratory
626 movements, depth preferences, and thermal biology of Atlantic bluefin
627 tuna. *Science* **293**, 1310–4.
628 Block B.A., Teo S.L.H., Walli A., Boustany A., Stokesbury M.J.W., Farwell
629 C.J., Weng K.C., Dewar H. & Williams T.D. (2005) Electronic tagging and
630 population structure of Atlantic bluefin tuna. *Nature* **434**, 1121–7.
631 Boustany A.M., Reeb C.A. & Block B.A. (2008) Mitochondrial DNA and
632 electronic tracking reveal population structure of Atlantic bluefin tuna
633 (*Thunnus thynnus*). *Marine Biology* **156**, 13–24.
634 Carlsson J., McDowell J.R., Carlsson J.E. & Graves J.E. (2007) Genetic
635 identity of YOY bluefin tuna from the eastern and western Atlantic
636 spawning areas. *Journal of Heredity* **98**, 23–8.
637 Chow S. & Kishino H. (1995) Phylogenetic relationships between tuna species
638 of the genus *Thunnus* (Scombridae: Teleostei): Inconsistent implications
639 from morphology, nuclear and mitochondrial genomes. *Journal of*
640 *Molecular Evolution* **41**, 741–8.
641 Chow S. & Ushiama H. (1995) Global population structure of albacore
642 (*Thunnus alalunga*) inferred by RFLP analysis of the mitochondrial
643 ATPase gene. *Marine Biology* **123**, 39–45.
644 Chow S., Nakagawa T., Suzuki N., Takeyama H. & Matsunaga T. (2006)
645 Phylogenetic relationships among *Thunnus* species inferred from rDNA
646 ITS1 sequence. *Journal of Fish Biology* **68** (Suppl. A), 24–35.
647 Collette B. & Nauen C. (1983) *FAO species catalogue. Vol. 2 Scombrids of*
648 *the world. An annotated and illustrated catalogue of tunas, mackerels,*

For Peer Review

649 *bonitos and related species known to date. FAO Fisheries Synopsis No.*
650 *125, Volume 2.* FAO, Rome.
651 Collette B.B., Carpenter K.E., Polidoro B.A., Juan-Jordá M.J., Boustany A.,
652 Die D.J., Elfes C., Fox W., Graves J., Harrison L.R., McManus R., Minte-
653 Vera C.V., Nelson R., Restrepo V., Schratwieser J., Sun C.-L., Amorim A.,
654 Peres M.B., Canales C., Cardenas G., Chang S.-K., Chiang W.-C., de
655 Oliveira Leite Jr. N., Harwell H., Lessa R., Fredou F.L., Oxenford H.A.,
656 Serra R., Shao K.-T., Sumaila R., Wang S.-P., Watson R. & Yáñez E.
657 (2011) High Value and Long Life—Double Jeopardy for Tunas and
658 Billfishes. *Science* **333**, 291–2.
659 Davies C.A., Gosling E.M., Was A., Brophy D. & Tysklind N. (2011)
660 Microsatellite analysis of albacore tuna (*Thunnus alalunga*): population
661 genetic structure in the North-East Atlantic Ocean and Mediterranean Sea.
662 *Marine Biology* **158**, 2727–40.
663 De Roos A.M. & Persson L. (2002) Size-dependent life-history traits promote
664 catastrophic collapses of top predators. *Proceedings of the National*
665 *Academy of Sciences of the USA* **99**, 12907–12.
666 Dickhut R.M., Deshpande A.D., Cincinelli A., Cochran M.A., Corsolini S., Brill
667 R.W., Secor D.H. & Graves J.E. (2009) Atlantic Bluefin Tuna (*Thunnus*
668 *thynnus*) population dynamics delineated by organochlorine tracers.
669 *Environmental Science and Technology* **43**, 8522–7.
670 Ely B., Stoner D.S., Bremer A.J., Dean J.M., Addis P., Cau A., Thelen E.J.,
671 Jones W.J., Black D.E., Smith L., Scott K., Naseri I. & Quattro J.M. (2002)
672 Analyses of nuclear ldhA gene and mtDNA control region sequences of
673 Atlantic Northern bluefin tuna populations. *Marine Biotechnology* **4**, 583–8.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

674 Foll M. & Gaggiotti O.E. (2008) A genome scan method to identify selected
675 loci appropriate for both dominant and codominant markers: A Bayesian
676 perspective. *Genetics* **180**, 977–93.
677 Fromentin J.M. & Fonteneau A. (2001) Fishing effects and life history traits: A
678 case study comparing tropical versus temperate tunas. *Fisheries*
679 *Research* **53**, 133–50.
680 Fromentin J.M., Powers J.E. (2005) Atlantic bluefin tuna: Population dynamics,
681 ecology, fisheries, and management. *Fish and Fisheries* **6**, 281–306.
682 Fromentin J.M. (2009) Lessons from the past: Investigating historical data
683 from bluefin tuna fisheries. *Fish and Fisheries* **10**, 197–216.
684 Galuardi B., Royer F., Golet W., Logan J., Neilson J. & Lutcavage M. (2010)
685 Complex migration routes of Atlantic bluefin tuna (*Thunnus thynnus*)
686 question current population structure paradigm. *Canadian Journal of*
687 *Fisheries and Aquatic Sciences* **67**, 966–76.
688 Gattepaille L.M. & Jokobsson M. (2012) Combining markers into haplotypes
689 can improve population structure inference. *Genetics* **190**, 159–174.
690 Garvin M.R., Saitoh K. & Gharrett A.J. (2010) Application of single nucleotide
691 polymorphisms to non-model species: a technical review. *Molecular*
692 *Ecology Resources*, **10**, 915–934.
693 Goudet J. (2001) FSTAT, a program to estimate and test gene diversities and
694 fixation indices (version 2.9.3). Available:
695 http://www2.unil.ch/popgen/softwares/fstat.htm
696 Haasl R.J. & Payseur B.A. (2011) Multi-locus inference of population
697 structure: a comparison between single nucleotide polymorphisms and
698 microsatellites. *Heredity* **106**, 158–71.

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

699 Hauser L. & Carvalho G.R. (2008) Paradigm shifts in marine fisheries
700 genetics: Ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**,
701 333–62.
702 Hayes B., Lærdahl J., Lien S., Berg P., Davidson W., Koop B., Adzhubei A. &
703 Høyheim B. (2004) Detection of single nucleotide polymorphisms (SNPs)
704 from Atlantic salmon expressed sequence tags (ESTs). In: *Proceedings of*
705 *the 55th Annual Meeting of the European Association for Animal*
706 *Production: 5–9 September 2004; Bled, Slovenia.* European Federation of
707 Animal Science.
708 Helyar S.J., Hemmer-Hansen J., Bekkevold D., Taylor M.I., Ogden R.,
709 Limborg M.T., Cariani A., Maes G.E. , Diopere E., Carvalho G.R. &
710 Nielsen E.E. (2011) Application of SNPs for populations genetics of non-
711 model organism: new opportunities and challenges. *Molecular Ecology*
712 *Resources* **11** (Suppl. 1), 123–36.
713 Jakobsson M. & Rosenberg N.A. (2007) CLUMPP: a cluster matching and
714 permutation program for dealing with label switching and multimodality in
715 analysis of population structure. *Bioinformatics* **23**, 1801–6.
716 Jørgensen H.B.H., Hansen M.M., Bekkevold D., Ruzzante D.E. & Loeschcke
717 V. (2005) Marine landscapes and population genetic structure of herring
718 (*Clupea harengus* L.) in the Baltic Sea. *Molecular Ecology* **14**, 3219–34.
719 Juan-Jordá M.J., Mosqueira I., Cooper A.B., Freire J. & Dulvy N.K. (2011)
720 Global population trajectories of tunas and their relatives. *Proceedings of*
721 *the National Academy of Sciences of the USA* **108**, 20650–5.
722 Kaminski S., Olech W., Olenski K., Nowak Z. & Rusc A. (2012) Single
723 nucleotide polymorphisms between two lines of European bison (*Bison*

For Peer Review

724 *bonasus*) detected by the use of Illumina Bovine 50 K BeadChip.
725 *Conservation Genetics Resources* **4**, 311–14.
726 Knutsen H., Jorde P.E., André C. & Stenseth N.C. (2003) Fine-scaled
727 geographical population structuring in a highly mobile marine species: the
728 Atlantic cod. *Molecular Ecology* **12**, 385–94.
729 Knutsen H., Jorde P.E., Albert O.T., Hoelzel A.R. & Stenseth N.C. (2007)
730 Population genetic structure in the North Atlantic Greenland halibut
731 (*Reinhardtius hippoglossoides*): influenced by oceanic current systems?
732 *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 857–66.
733 Knutsen H., Olsen E.M., Jorde P.E., Espeland S.H., André C. & Stenseth N.C.
734 (2011) Are low but statistically significant levels of genetic differentiation in
735 marine fishes 'biologically meaningful'? A case study of coastal Atlantic
736 cod. *Molecular Ecology* **20**, 768–83.
737 Lutcavage M.E., Brill R.W., Skomal G.B., Chase B.C. & Howey P.W. (1999)
738 Results of pop-up satellite tagging on spawning size class fish in the Gulf
739 of Maine: Do North Atlantic bluefin tuna spawn in the mid-Atlantic?
740 *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 173–7.
741 Magnuson J.J., Block B.A., Deriso R.B., Gold J.R., Grant W.S., Quinn T.J.,
742 Saila S.B., Shapiro L. & Stevens E.D. (1994) *An assessment of Atlantic*
743 *bluefin tuna.* National Academy Press, Washington, DC.
744 Malhi R.S., Trask J.S., Shattuck M., Johnson J., Chakraborty D.,
745 Kanthaswamy S., Ramakrishnan U. & Smith D.G. (2011) Genotyping
746 single nucleotide polymorphisms (SNPs) across species in Old World
747 monkeys. *American Journal of Primatology* **73**, 1031–40.

For Peer Review

Mather F.J., Mason J.M. & Jones A.C. (1995) Historical document: life history and fisheries of Atlantic bluefin tuna. *NOAA Technical Memorandum* 370, NOAA, Miami.

Mesnick S.L., Taylor B.L., Archer F.I., Martien K.K., Treviño S.E., Hancock-Hanser B.L., Moreno Medina S.C., Pease V.L., Robertson K.M., Straley J.M., Baird R.W., Calambokidis J., Schorr G.S., Wade P., Burkanov V., Lunsford C.R., Rendell L. & Morin P.A. (2011) Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. *Molecular Ecology Resources* **11** (Suppl. 1), 278–98.

Miller J.M., Poissant J., Kijas J.W. & Coltman D.W. (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314–22.

Moen T., Hayes B., Nilsen F., Delghandi M., Fjalestad K.T., Fevolden S.E., Berg P.R. & Lien S. (2008) Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genetics* **9**, 18.

Molecular Ecology Resources Primer Development Consortium, Abreu A.G., Albaina A., Alpermann T.J., Apkenas V.E., Bankhead-Dronnet S., Bergek S., Berumen M.L., Cho C.H., Clobert J., Coulon A., De Feraudy D., Estonba A., Hankeln T., Hochkirch A., Hsu T.W., Huang T.J., Irigoien X., Iriondo M., Kay K.M., Kinitz T., Kothera L., Le Hénanff M., Lieutier F., Lourdais O., Macrini C.M., Manzano C., Martin C., Morris V.R., Nanninga G., Pardo M.A., Plieske J., Pointeau S., Prestegaard T., Quack M., Richard M., Savage H.M., Schwarcz K.D., Shade J., Simms E.L., Solferini

For Peer Review

773 V.N., Stevens V.M., Veith M., Wen M.J., Wicker F., Yost J.M. &
774 Zarraonaindia I. (2012) Permanent genetic resources added to Molecular
775 Ecology Resources Database 1 October 2011-30 November 2011
776 *Molecular Ecology Resources* **12**, 374–6.
777 Morin P.A., Luikart G. & Wayne R.K. (2004) SNPs in ecology, evolution, and
778 conservation. *Trends in Ecology and Evolution* **19**, 208–16.
779 Nakadate M., Viñas J., Corriero A., Clarke S., Suzuki N. & Chow S. (2005)
780 Genetic isolation between Atlantic and Mediterranean albacore
781 populations inferred from mitochondrial and nuclear DNA markers. *Journal*
782 *of Fish Biology* **66**, 1545–57.
783 Nakamura H. (1969) *Tuna Distribution and Migration.* Fishing News (Books)
784 Ltd, London.
785 Nielsen E.E., Hemmer-Hansen J., Larsen P.F. & Bekkevold D. (2009)
786 Population genomics of marine fishes: identifying adaptative variation in
787 space and time. *Molecular Ecology* **18**, 3128–50.
788 Ogden R. (2011) Unlocking the potential of genomic technologies for wildlife
789 forensics. *Molecular Ecology Resources* **11** (Suppl. 1), 109–16.
790 Ogden R., Baird J., Senn H. & McEwing R. (2012) The use of cross-species
791 genome-wide arrays to discover SNP markers for conservation genetics: a
792 case study from Arabian and scimitar-horned oryx. *Conservation Genetics*
793 *Resources* **4**, 471–3.
794 Petit E., Balloux F. & Goudet J. (2001) Sex-biased dispersal in a migratory
795 bat: a characterization using sex-specific demographic parameters.
796 Evolution **55**, 635–40.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

797 Poulsen N.A., Hemmer-Hansen J., Loeschcke V., Carvalho G.R. & Nielsen
798 E.E. (2011) Microgeographical population structure and adaptation in
799 Atlantic cod *Gadus morhua*: spatio-temporal insights from gene-associated
800 DNA markers. *Marine Ecology Progress Series* **436**, 231–43.
801 Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population
802 structure using multilocus genotype data. *Genetics* **155**, 945–59.
803 Pujolar J.M., Roldán M.I. & Pla C. (2003) Genetic analysis of tuna populations,
804 *Thunnus thynnus thynnus* and T*. alalunga*. *Marine Biology* **143**, 613–21.
805 Riccioni G., Landi M., Ferrara G., Milano I., Cariani A., Zane L., Sella M.,
806 Barbujani G. & Tinti F. (2010) Spatio-temporal population structuring and
807 genetic diversity retention in depleted Atlantic bluefin tuna of the
808 Mediterranean Sea. *Proceedings of the National Academy of Sciences of*
809 *the USA* **107**, 2102–7.
810 Richter-Boix A., Quintela M., Segelbacher G. & Laurila A. (2011) Genetic
811 analysis of differentiation among breeding ponds reveals a candidate gene
812 for local adaptation in *Rana arvalis*. *Molecular Ecology* **20**, 1582–600.
813 Rooker J.R., Secor D.H., De Metrio G. & Rodríguez-Marín E. (2006)
814 Evaluation of population structure and mixing rates of Atlantic bluefin tuna
815 from chemical signatures in otoliths. *The International Commission for the*
816 *Conservation of Atlantic Tuna Collective Volume of Scientific Papers* **59**,
817 813–8.
818 Rooker J.R., Alvarado Bremer J.R., Block B.A., Dewar H., de Metrio G.,
819 Corriero A., Krause R.T., Prince E.D., Rodríguez-Marín E. & Secor D.H.
820 (2007) Life history and stock structure of Atlantic Bluefin Tuna (*Thunnus*
821 *thynnus*). *Reviews in Fisheries Science* **15**, 265–310.

For Peer Review

822 Rooker J.R., Secor D.H., De Metrio G., Schloesser R., Block B.A. & Neilson
823 J.D. (2008) Natal homing and connectivity in Atlantic bluefin tuna
824 populations. *Science* **322**, 742–4.
825 Rosenberg N.A., Li L.M., Ward R. & Pritchard J.K. (2003) Informativeness of
826 genetic markers for inference of ancestry. *American Journal of Human*
827 *Genetics* **73**, 1402–22.
828 Rosenberg N.A. (2004) Distruct: a program for the graphical display of
829 population structure. *Molecular Ecology Notes* **4**, 137–8.
830 Rousset F. (2008) Genepop'007: a complete reimplementation of the
831 Genepop software for Windows and Linux. *Molecular Ecology Resources*
832 **8**, 103–6.
833 Rozen S. & Skaletsky H.J. (2000) Primer3 on the WWW for general users and
834 for biologist programmers. *Methods in Molecular Biology* **132,** 365–86.
835 Humana Press.
836 Ryman N., Palm S., André C., Carvalho G.R., Dahlgren T.G., Jorde P.E.,
837 Laikre L., Larsson L.C., Palmé A. & Ruzzante D.E. (2006) Power for
838 detecting genetic divergence: differences between statistical methods and
839 marker loci. *Molecular Ecology* **15**, 2031–45.
840 Ryynänen H.J. & Primmer CR (2006) Single nucleotide polymorphism (SNP)
841 discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a
842 strategy for avoiding amplification of duplicated loci in Atlantic salmon
843 (*Salmo salar*) and other salmonid fishes. *BMC Genomics* **7**, 192.
844 Seeb J.E., Carvalho G., Hauser L., Naish K., Roberts S. & Seeb L.W. (2011)
845 Single-nucleotide polymorphism (SNP) discovery and applications of SNP

For Peer Review

846 genotyping in nonmodel organisms. *Molecular Ecology Resources* **11**
847 (Suppl. 1) 1–8.
848 Slate J., Gratten J., Beraldi D., Stapley J., Hale M. & Pemberton J.M. (2009)
849 Gene mapping in the wild with SNPs: guidelines and future directions.
850 *Genetica* 136, 97–107.
851 Smith C.T., Elfstrom C.M., Seeb L. & Seeb J.E. (2005) Use of sequence data
852 from rainbow trout and Atlantic salmon for SNP detection on Pacific
853 salmon. *Molecular Ecology* 14, 4193–203.
854 Stephens M., Smith N. & Donnelly P. (2001) A new statistical method for
855 haplotype reconstruction from population data. *American Journal of*
856 *Human Genetics* **68**, 978–89.
857 Takagi M., Okamura T., Chow S. & Taniguchi N. (1999) PCR primers for
858 microsatellite loci in tuna species of the genus *Thunnus* and its application
859 for population genetic study. *Fisheries Science* **65**, 571–6.
860 Takagi M., Okamura T., Chow S. & Taniguchi N. (2001) Preliminary study of
861 albacore (*Thunnus alalunga*) stock differentiation inferred from
862 microsatellite DNA analysis. *Fisheries Bulletin* **99**, 697–701.
863 Teo S.L.H., Boustany A., Dewar H., Stokesbury M.J.W., Weng K.C., Beemer
864 S., Seitz A.C., Farwell C.J., Prince E.D. & Block B.A. (2007) Annual
865 migrations, diving behavior, and thermal biology of Atlantic bluefin tuna,
866 *Thunnus thynnus*, on their Gulf of Mexico breeding grounds. *Marine*
867 *Biology* **151**, 1–18.
868 Viñas J., Alvarado Bremer J.R. & Pla C. (2004) Inter-oceanic genetic
869 differentiation among albacore (*Thunnus alalunga*) populations. *Marine*
870 *Biology* **145**, 225–32.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

871 Viñas J. & Tudela S. (2009) A validated methodology for genetic identification
872 of tuna species (genus *Thunnus*). *PLoS ONE* **4**, e7606.
873 Viñas J., Gordoa A., Fernández-Cebrián R., Pla C., Vahdet U. & Araguas R.M.
874 (2011) Facts and uncertainties about the genetic population structure of
875 Atlantic bluefin tuna (*Thunnus thynnus*) in the Mediterranean. Implications
876 for fishery management. *Reviews in Fish Biology and Fisheries* **21**, 527–
877 41.
878 Walli A., Teo S.L.H., Boustany A., Farwell C.J., Williams T., Dewar H., Prince
879 E. & Block B.A. (2009) Seasonal Movements, Aggregations and Diving
880 Behavior of Atlantic Bluefin Tuna (*Thunnus thynnus*) Revealed with
881 Archival Tags. *PLoS ONE* **4**, e6151.
882 Waples R.S. & Gaggiotti O. (2006) What is a population? An empirical
883 evaluation of some genetic methods for identifying the number of gene
884 pools and their degree of connectivity. *Molecular Ecology* **15**, 1419–39.
885 Waples R.S. & Naish K.A. (2009) Genetic and evolutionary considerations in
886 fishery management: Research needs for the future. In: *The Future of*
887 *Fisheries Science in North America*. Fish and Fisheries Series Volume 31
888 (ed. by R. Beamish & B. Rothschild) pp. 427–51. Springer, Netherlands.
889 Weir B.S. & Cockerham C.C. (1984) Estimating F statistics for the analysis of
890 population structure. *Evolution* **38**, 1358–70.
891 Yeh S.Y., Treng T.D., Hui C.F. & Penney A.J. (2007) Mitochondrial DNA
892 sequence analysis on albacore, *Thunnus alalunga*, meat samples
893 collected from the waters off western South Africa and the eastern Indian
894 Ocean. *Collective Volume of Scientific Papers. International Commission*
895 *for the Conservation of Atlantic Tunas* **46**, 152–9.

For Peer Review

896 Zarraonaindia I. (2011) Estudio genético de la estructura poblacional de la
897 anchoa europea (*Engraulis encrasicolus*). *PhD thesis.* University of the
898 Basque Country (UPV/EHU), Genetics, Physical Anthropology & Animal
899 Physiology Department.

For Peer Review

1
2
3
4
5
6
7 900
8 901
9 902 **Table 1 Sampling details.** Sample code, number of individuals per sample (*N*), sample location, current management stock, FAO
10
11 903 major fishing area and geographical coordinates along with year of capture.
12
13 904
14
15

Sample	Abbreviation	*N*	Location	Latitude	Longitude	Year	Current stock	FAO
Albacore tuna								
1	BAL	50	Balearic Sea	40.00	1.58	2005	Mediterranean	37
2	BIS	52	Bay of Biscay	45.10	-4.35	2009	North Atlantic	27
3	IRE	57	Ireland	54.17	-12.89	2008	North Atlantic	27
4	SEA	91	South Africa	-24.25	4.42	2009	South Atlantic	47
5	IN	24	Seychelles	-7.11	54.65	2008–2009	Indian	51
6	NP	101	California	43.50	-127.00	2008	North Pacific	77
7	SWP	30	New Caledonia	-18.53	165.97	2003–2008	South Pacific	71
8	SEP	55	French Polynesia	-19.01	-152.84	2003–2008	South Pacific	71
Atlantic Bluefin tuna								
9	NEA	46	Bay of Biscay	45.10	-4.35	2009	East Atlantic	27
10	MED	46	Balearic Sea	40.58	1.21	2009	East Atlantic	37
11	NWA	15	Northwest Atlantic	36.24	74.49	2008	West Atlantic	21

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31 905
32 906
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

For Peer Review

1
2
3 907 **Table 2 Selected set of SNPs for population genetic studies in *T.***
4
5 908 ***alalunga*.** SNP code (in bold), the nuclear DNA fragment were it was
6
7 909 discovered and the discovery approach (I, II and III; see Methods) are shown,
8
9 910 for the panel of 53 SNPs, representing 41 loci including 32 individual SNPs
10
11 911 and 9 haplotype blocks (shaded), along with the number of alleles or
12
13 912 haplotypes and mean values, across samples, for both expected
14
15 913 heterozygosity (H_e) and inbreeding coefficient (F_{IS}).
16
17
18 914

Fragment	Method	SNP	no. alleles/ haplotypes	Mean H_e (±SE)	Mean F_{IS} (±SE)
ADRB2	I	***ADRB2-97***	2	0.025±0.016	-0.006±0.006
ALDOB1	I	***ALDOB1-47*** ***ALDOB1-95***	4	0.373±0.058	0.150±0.104
CALM4	I	***CALM4-124***	2	0.432±0.040	0.247±0.165
GNRH3-1	I	***GNRH3-1-124***	2	0.365±0.040	0.030±0.047
GNRH3-3	I	***GNRH3-3-219***	2	0.077±0.035	-0.033±0.018
LDHB	I	***LDHB-287***	2	0.362±0.061	0.176±0.205
LYZ	I	***LYZ-128*** ***LYZ-138*** ***LYZ-340***	8	0.579±0.079	-0.005±0.155
MLL25a	I	***MLL25a-144*** ***MLL25a-183***	4	0.526±0.057	0.092±0.076
MYC	I	***MYC-91***	2	0.043±0.082	-0.024±0.050
MYL3	I	***MYL3-97***	2	0.024±0.025	-0.009±0.009
RHO	I	***RHO-111***	2	0.033±0.039	-0.014±0.017
RPS7-2	I	***RPS7-2-69***	2	0.270±0.081	0.180±0.138
APOE	II	***APOE-148***	2	0.269±0.039	0.014±0.116
CYCS	II	***CYCS-132*** ***CYCS-218***	4	0.468±0.062	-0.050±0.096
DAD1	II	***DAD1-444***	2	0.088±0.021	-0.039±0.011
FGB	II	***FGB-257***	2	0.053±0.036	-0.024±0.020
FOS	II	***FOS-107***	2	0.154±0.084	-0.089±0.060
HGF	II	***HGF-375***	2	0.039±0.022	-0.015±0.011
HMOX1	II	***HMOX1-416***	2	0.316±0.102	0.181±0.214
MMP9	II	***MMP9-68*** ***MMP9-111***	4	0.220±0.107	0.051±0.074
MTF1	II	***MTF1-263***	2	0.316±0.110	0.077±0.104

19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

PSM	II	***PSM-33*** ***PSM-117*** ***PSM-138***	8	0.737±0.037	-0.045±0.102
PTGS2	II	***PTGS2-56*** ***PTGS2-317***	4	0.708±0.022	-0.017±0.154
RASA3	II	***RASA3-188***	2	0.063±0.048	-0.028±0.022
RHOC	II	***RHOC-55***	2	0.054±0.056	-0.021±0.027
RPL12	II	***RPL12-213*** ***RPL12-423***	4	0.444±0.169	0.042±0.042
c-mos	III	***c-mos-242***	2	0.313±0.059	-0.055±0.139
CS1	III	***CS1-197*** ***CS1-442*** ***CS1-512***	6	0.533±0.042	-0.030±0.124
CS3	III	***CS3-394***	2	0.449±0.039	0.006±0.220
CS5	III	***CS5-44***	2	0.494±0.013	0.209±0.161
ELOVL2	III	***ELOVL2-519***	2	0.253±0.052	0.095±0.239
ELOVL3	III	***ELOVL3-365***	2	0.129±0.081	-0.071±0.057
FGG	III	***FGG-242***	2	0.371±0.102	-0.009±0.148
GPX-458	III	***GPX-458***	2	0.495±0.009	0.107±0.115
HIF1A2-3	III	***HIF1A2-3-350***	2	0.449±0.027	0.073±0.083
HIF1A4	III	***HIF1A4-219***	2	0.145±0.122	-0.022±0.043
MB	III	***MB-188***	2	0.174±0.068	0.095±0.225
OPC02	III	***OPC02-249***	2	0.167±0.046	-0.032±0.111
PRDX2	III	***PRDX2-452***	2	0.013±0.023	-0.004±0.009
RAG2	III	***RAG2-114***	2	0.312±0.088	0.147±0.195
Tmo-4C4	III	***Tmo-4C4-188***	2	0.063±0.046	-0.026±0.027

36 915
37 916
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3 917 **Table 3 Probability of detecting a particular level of differentiation (F_{ST})**
4
5 918 **among populations of albacore.** Power analysis conducted with POWSIM
6
7 919 (1000 replicates) with mean frequencies and sample sizes used for albacore
8
9 920 tuna.
10
11
12 921

	41 SNPs and blocks		**9 Blocks**		**32 SNPs**	
F_{ST}	***P*** $\mathbf{X^2}$	***P* Fisher**	***P*** $\mathbf{X^2}$	***P* Fisher**	***P*** $\mathbf{X^2}$	***P* Fisher**
0.0005	0.197	0.186	0.125	0.127	0.156	0.150
0.001	0.439	0.404	0.276	0.278	0.286	0.289
0.002	0.886	0.844	0.662	0.638	0.628	0.625
0.004	1.0	0.999	0.981	0.970	0.980	0.976
0.008	1.0	1.0	1.0	1.0	1.0	1.0
0.016	1.0	1.0	1.0	1.0	1.0	1.0

13
14
15
16
17
18
19
20
21
22 922
23 923
24 924
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3 925 **Table 4. Pairwise F_{ST} values between samples of albacore tuna (*Thunnus*
4
5 926 ***alalunga*).** F_{ST} values appear below the diagonal and standard errors above
6
7 927 the diagonal. Sample abbreviations as in Table 1, F_{ST} values significantly
8
9 928 larger than 0.0 are in bold ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$)
10
11 929
12
13

	BAL	BIS	IRE	SEA	IN	NP	SEP	SWP
BAL		0.008	0.008	0.008	0.013	0.010	0.011	0.008
BIS	**0.026*****		0.002	0.005	0.008	0.006	0.007	0.008
IRE	**0.030*****	0.000		0.002	0.007	0.004	0.005	0.008
SEA	**0.033*****	0.004	-0.001		0.008	0.005	0.004	0.007
IN	**0.050*****	**0.020*****	**0.017*****	**0.020*****		0.004	0.005	0.006
NP	**0.033*****	**0.014*****	**0.019*****	**0.016*****	**0.008***		0.002	0.002
SEP	**0.043*****	**0.017*****	**0.016*****	**0.012*****	**0.007***	-0.000		0.003
SWP	**0.021*****	**0.014****	**0.011****	**0.012****	**0.011***	0.001	0.001	

930

14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

For Peer Review

1
2
3 931 **Table 5 Selected set of SNPs for population genetic studies in *T.***
4
5 932 ***thynnus.*** SNP code (in bold), the identity of the nuclear DNA fragment where
6
7 933 it was discovered in *T. alalunga* along with the applied SNP discovery
8
9 934 approach (see Methods) are shown for the panel of 17 SNPs, representing 15
10
11 935 independent loci including 13 SNPs and 2 haplotype blocks (shaded), along
12
13 936 with the number of alleles or haplotypes and mean values, pooling all
14
15 937 sampling locations, for both expected heterozygosity (H_e) and inbreeding
16
17 938 coefficient (F_{IS}).

Fragment	Method	SNP	no. alleles/ haplotypes	Mean H_e (±SE)	Mean F_{IS} (±SE)
ADRB2	I	***ADRB2-97***	2	0.046±0.040	-0.016±0.014
GNRH3-1	I	***GNRH3-1-107*** ***GNRH3-1-124***	4	0.676±0.023	-0.056±0.178
LDHB	I	***LDHB-129***	2	0.468±0.068	0.057±0.268
LYZ	I	***LYZ-128***	2	0.062±0.054	0.439±0.380
RPS7-2-313	I	***RPS7-2-313***	2	0.084±0.079	0.083±0.176
TPMA	I	***TPMA-53***	2	0.355±0.100	0.000±0.351
CYCS	II	***CYCS-161***	2	0.175±0.195	0.216±0.240
HGF	II	***HGF-375***	2	0.211±0.090	0.106±0.140
MTF1-263	II	***MTF1-263***	2	0.273±0.055	0.010±0.162
RPL12	II	***RPL12-423***	2	0.486±0.012	0.006±0.157
CS3	III	***CS3-118***	2	0.291±0.253	0.023±0.070
CS5	III	***CS5-395*** ***CS5-425***	4	0.243±0.211	0.137±0.183
FGG	III	***FGG-242***	2	0.401±0.108	0.215±0.198
HIF1A2-3	III	***HIF1A2-3-417***	2	0.145±0.041	0.216±0.500
OPC02	III	***OPC02-45***	2	0.167±0.289	0.000±0.000

939
940

18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

For Peer Review

Table 6 Probability of detecting a particular level of differentiation (F_{ST}) among populations of Atlantic bluefin tuna. Power analysis conducted with POWSIM (1000 replicates) with mean frequencies and sample sizes used for Atlantic bluefin tuna.

	13 SNPs and 2 blocks	
F_{ST}	*P* X^2	*P* Fisher
0.001	0.091	0.064
0.0025	0.159	0.104
0.005	0.314	0.240
0.010	0.654	0.579
0.020	0.961	0.931
0.040	1.000	0.998
0.080	1.000	1.000

For Peer Review

Figure legends

Figure 1. Sampling locations and approximate locations of spawning areas in *Thunnus alalunga* and *T. thynnus*. Respectively, black circles and right-oriented hatching for *T. alalunga*, and black squares and left-oriented hatching for *T. thynnus*; both species spawn in Mediterranean waters.

Figure 2: SNP selection design. Design of the filtering steps used to select SNP panels for individual assignment and for genetic population surveys in *T. alalunga* and *T. thynnus* (see Methods for further information)

Figure 3. *Thunnus alalunga* STRUCTURE results. Individual clustering analysis with STRUCTURE (respectively, $K = 2$, $K = 3$ and $K = 4$) of 460 *T. alalunga* individuals for 53 SNPs located in 41 independent fragments. Each vertical bar represents an individual, and sampling locations are separated by vertical black lines. The colour proportions of each bar correspond to the individual's estimated membership fraction to each of the clusters (cluster membership coefficient).

Figure 4. *Thunnus thynnus* STRUCTURE results. Individual clustering analysis with STRUCTURE ($K = 2$) analysis of 107 *T. thynnus* individuals for 17 SNPs on 15 independent DNA fragments. Each vertical bar represents an individual, and sampling locations are separated by vertical black lines. The colour proportions of each bar correspond to the individual's estimated membership fractions to each of the clusters (cluster membership coefficient).

For Peer Review

1
2
3 974 ## Supplementary material
4
5
6
7 975 **Table S1.** Information on the 54 nuclear DNA fragments selected for SNP
8
9 976 discovery in *Thunnus alalunga* (see Methods): fragment name, gene region,
10
11 977 SNP discovery approach (1, 2 and 3; see Methods), source, forward and
12
13 978 reverse primer sequences, PCR annealing temperatures, sequenced
14
15 979 fragment lengths and recorded number of SNPs and indels. EPIC primers
16
17 980 (Exon-Priming, Intron-Crossing primers; SNP discovery approach I) were
18
19
20 981 obtained from the literature while primers for the remaining DNA fragments
21
22 982 were designed from the alignment of sequences from publically available
23
24 983 databases (GenBank and Ensembl), respectively, 17 pairs of degenerate
25
26
27 984 primers from several teleost species sequences (approach II) and 18 pairs of
28
29 985 primers from genus *Thunnus* DNA sequences (approach III).
30
31 986
32
33 987 **Table S2.** Information for the polymerase chain reactions (PCR) and cycling
34
35
36 988 conditions used to amplify the 54 nuclear DNA fragments selected for SNP
37
38 989 discovery in *T. alalunga*. While 30 fragments were amplified with a
39
40 990 conventional PCR, touchdown (TD) methodology, involving a decreasing
41
42 991 annealing Tº, was applied for the remaining fragments. [(1)] Annealing
43
44 992 temperature started in the first cycle with that registered in Table S1
45
46
47 993 ("Touchdown temperature" column) for each fragment. Then, annealing
48
49 994 temperature decreased 0.5ºC in each subsequent cycle. [(2)] Fixed annealing
50
51 995 temperature for each fragment. See "Fixed temperature" column in Table S1.
52
53 996
54
55 997 **Table S3.** BLASTN homology search results for the *T. alalunga* sequenced
56
57
58 998 fragments. Only *best hits* showing identity ≥ 70% and E-value < $10e^{-5}$ were
59
60

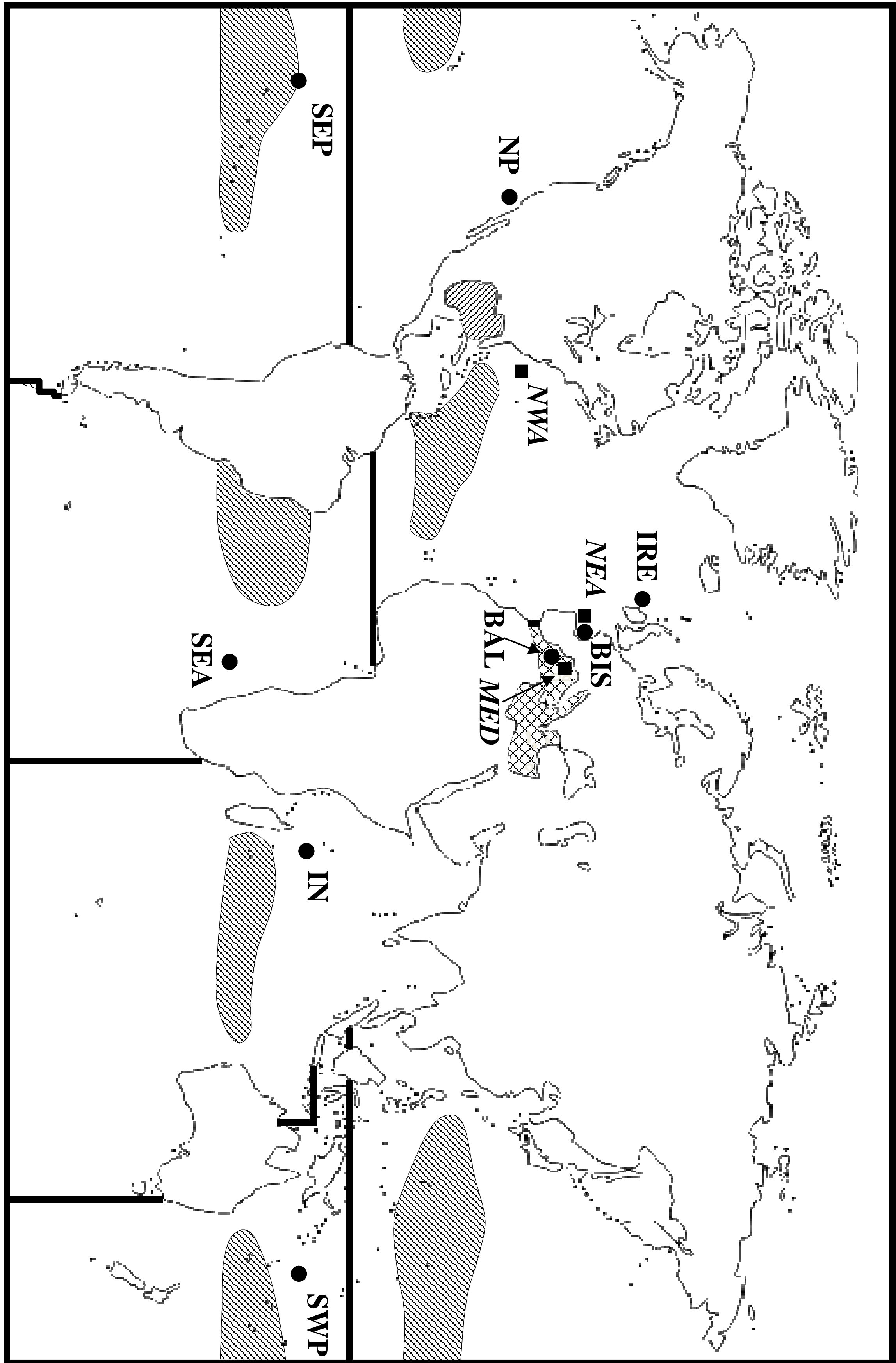
For Peer Review

1
2
3 999 considered. Out of 54 fragments of nuclear DNA that were sequenced in *T.*
4
5 1000 *alalunga*, 35 fulfilled the required conditions.
6
7 1001
8
9
10 1002 **Table S4.** Characteristics of the 79 validated SNPs, meaning reliably scored
11
12 1003 and polymorphic, in *Thunnus alalunga*. Marker name, SNP discovery
13
14 1004 approach (1, 2 and 3; see Methods), call rate (%), alleles and minor allele
15
16 1005 frequency (MAF) values are shown along with the global and per location
17
18 1006 deviation (*P*) from the Hardy-Weinberg equilibrium (HWE). “SIGN.” means
19
20
21 1007 that at least 1 location was not in HWE for those markers (N.S. corresponds
22
23 1008 to not significant); as multiple comparisons were involved, a probability of $P <$
24
25 1009 0.001 was considered significant. Last column shows NCBI Assay ID (rs#) for
26
27 1010 every individual SNP (NCBI´s dbSNP database).
28
29
30 1011
31
32 1012 **Table S5.** Results from tests of Linkage disequilibrium (LD) between all
33
34 1013 possible SNP pairs (validated SNPs only) within each DNA fragment, for both
35
36 1014 *T. alalunga* and *T. thynnus* (*$P < 0.001$). No SNPs were found in LD when
37
38 1015 comparing among DNA fragments. Linked SNPs were phased into haplotypes
39
40
41 1016 using the Bayesian statistical method implemented in PHASE 2.1. Haplotypes
42
43 1017 were reconstructed by location to avoid biases from population structuring.
44
45 1018 Missing genotypes were classified as null genotypes to avoid haplotype
46
47 1019 reconstruction errors.
48
49
50 1020
51
52 1021 **Table S6.** Characteristics of the 23 validated SNPs in *Thunnus thynnus*.
53
54 1022 Marker name, SNP discovery approach (1,2 and 3), call rate (%), alleles and
55
56 1023 minor allele frequency (MAF) values are shown along with the global and per
57
58
59
60

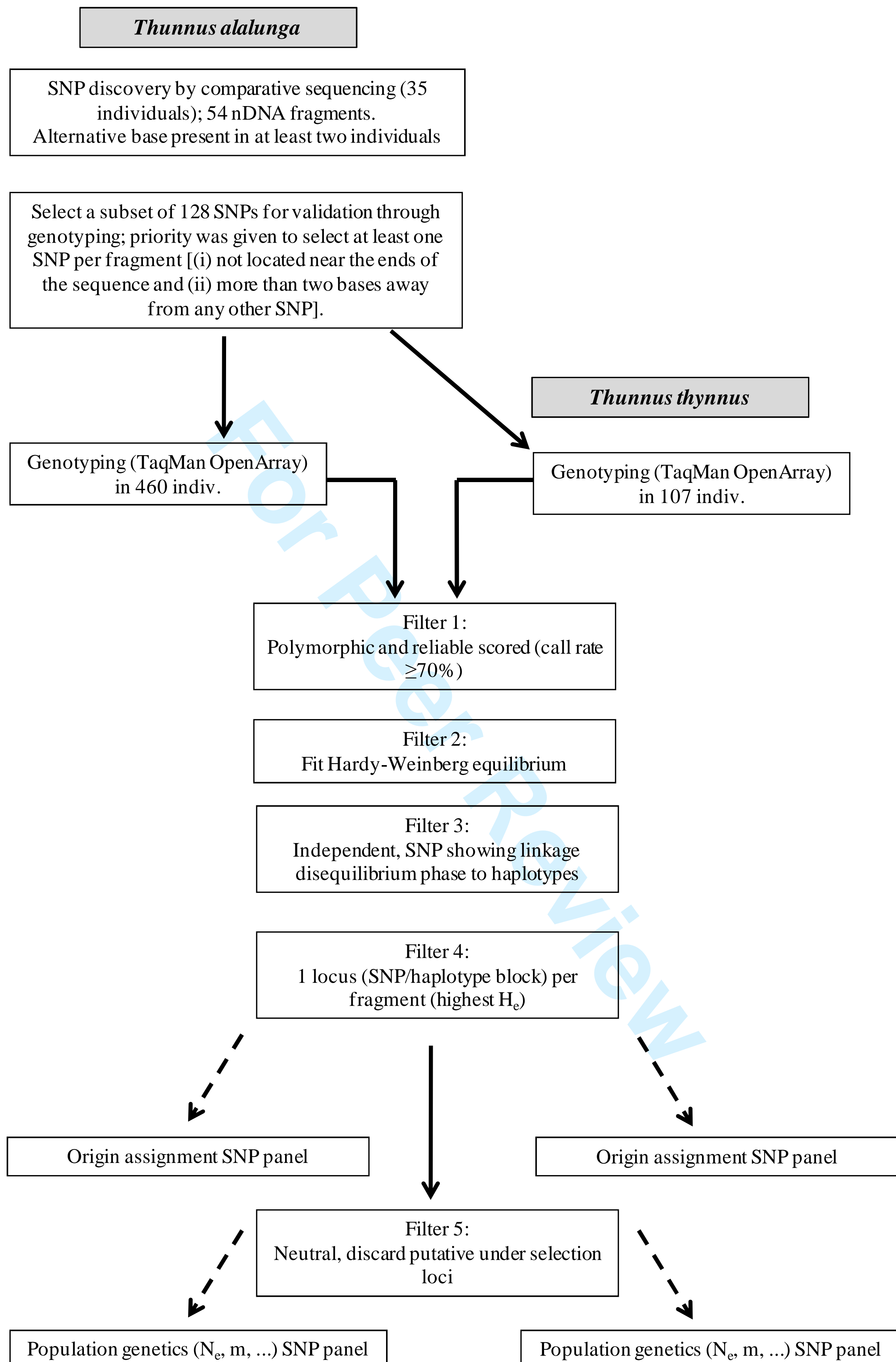
For Peer Review

1024 location deviation (*P*) from the Hardy-Weinberg equilibrium (HWE). "SIGN."
1025 means that at least 1 location was not in HWE for those markers (N.S.
1026 corresponds to not significant); as multiple comparisons were involved, a
1027 probability of $P < 0.001$ was considered significant. Out of the 23 SNPs
1028 validated in BFT by cross-species amplification of 128 SNPs discovered in
1029 albacore (see Methods and Fig. 2), 18 had been validated also when
1030 genotyping albacore while the other five SNPs (*GNRH3-1-107, LDHB-129,*
1031 *CYCS-161, CS3-118,* and *OPC02*-45) were reported as reliably scored and
1032 polymorphic only in BFT (last column shows their individual NCBI Assay IDs
1033 (ss#) in the NCBI´s dbSNP database).

For Peer Review

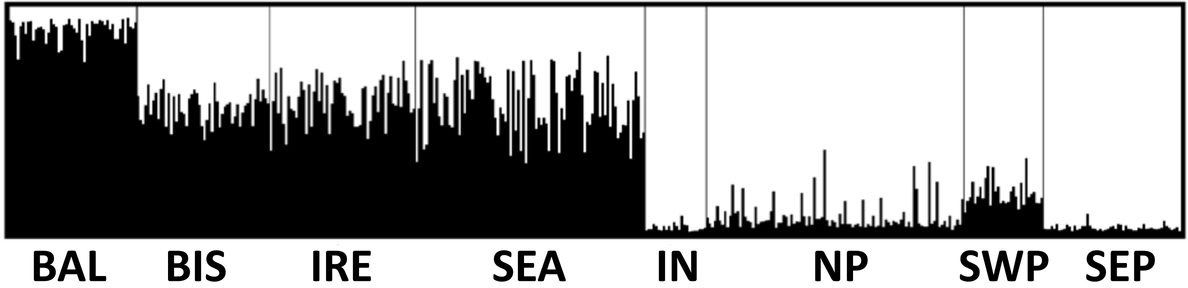

SEP
NP
NWA
IRE
NEA
BIS
BAL
MED
SEA
IN
SWP

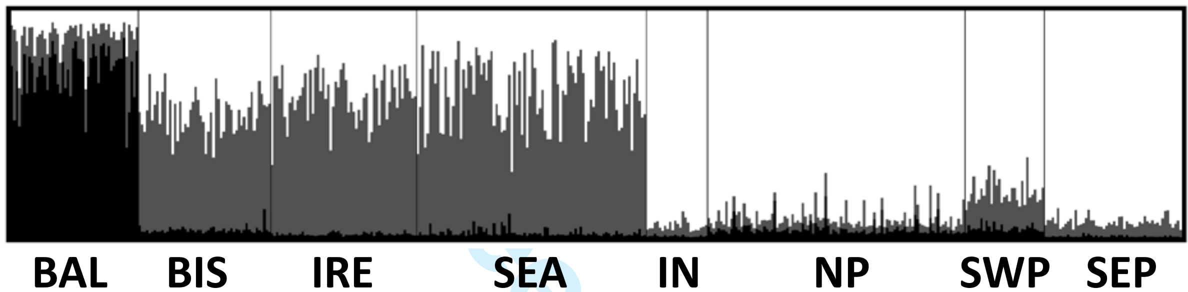
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

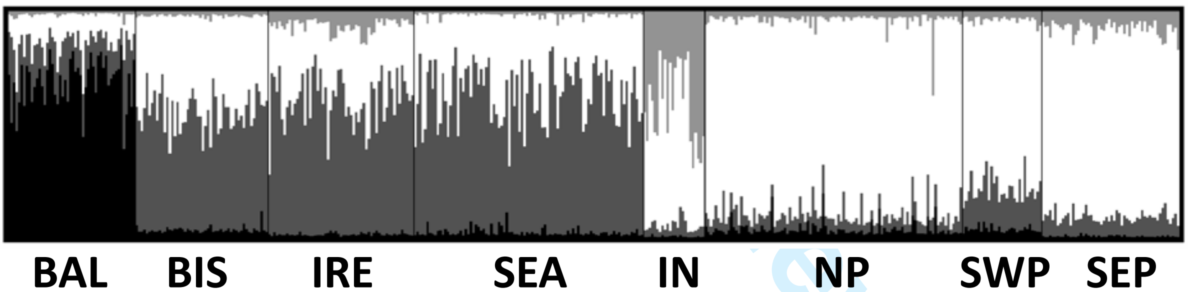


For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58

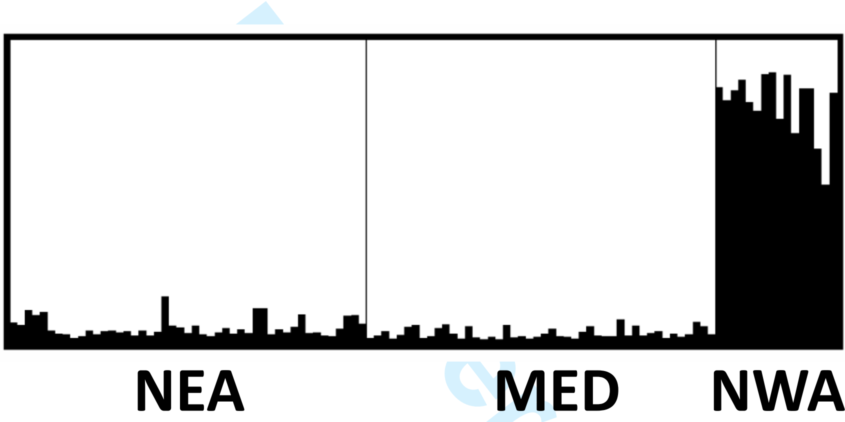






For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58



For Peer Review