



ELSEVIER

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Full length article

SNP identification and validation on genomic DNA for studying genetic diversity in *Thunnus albacares* and *Scomberomorus brasiliensis* by combining RADseq and long read high throughput sequencing



Zoila Raquel Siccha-Ramirez^{a,b,*,1}, Francesco Maroso^{c,d,1}, Belén G. Pardo^c, Carlos Fernández^c, Paulino Martínez^{c,**}, Claudio Oliveira^a

^a Departamento de Morfología, Instituto de Biociências, UNESP, Botucatu, São Paulo, Brazil

^b Instituto del Mar del Perú, Laboratorio Costero de Tumbes, Zorritos, Peru

^c Departamento de Zooloxía, Xenética e Antropoloxía Física, Facultade de Veterinaria, Universidade de Santiago de Compostela, Campus de Lugo, Lugo, 27002, Spain

^d Dipartimento di Biomedicina Comparata e Alimentazione (BCA), Università degli Studi di Padova, Italy

ARTICLE INFO

Keywords:

Scomberomorus brasiliensis

Thunnus albacares

ddRAD

Long-read high throughput sequencing

Genetic structure

ABSTRACT

A combination of a RADseq method (ddRAD) with long read high throughput sequencing (Roche 454) was tuned up in order to identify and validate a set of SNPs useful for gene diversity analysis in two important South American commercial tuna (*Thunnus albacares* and *Scomberomorus brasiliensis*). A total of 11 and 21 individuals of *T. albacares* and *S. brasiliensis*, respectively, were used for SNP identification. DNA was individually digested with two restriction enzymes (*SbfI* and *SphI*) and fragments between 300 and 600 bp selected. Combinatorial barcoding was used to identify individuals by including short sequences (5–7 bp) in the adaptors of each restriction site (P1 and P2). After adaptor ligation, samples were pooled and size-selected, amplified by PCR, and sequenced on a 454 GS-Junior sequencer. A total of 180,779 reads were produced with an average length and coverage of 287 bp and 26x, respectively. Sets of 60 and 79 SNPs were *in silico* selected for *T. albacares* and *S. brasiliensis*, respectively, and were tested and validated in 74 and 66 individuals, respectively, on a MassARRAY platform. A total of 36 and 47 SNPs were polymorphic and useful for population analysis. A preliminary study on two distant Brazilian populations of both species (~3000 km) with these SNPs suggested the absence of significant structure among local populations of both species. Our results demonstrate the possibility of combining ddRAD with long read high throughput sequencing for marker development in species with scarce genomic resources.

1. Introduction

Next Generation Sequencing (NGS) has revolutionized the field of genetics (Guo et al., 2016; Mardis, 2008; Metzker, 2010) allowing investigation on non-model species with unprecedented genomic coverage. An increasingly used application is the search, validation and large-scale genotyping of genetic markers using different methodologies (Helyar et al., 2012). SNPs are stable and usually bi-allelic polymorphisms (Mullikin et al., 2000), found in coding and non-coding regions (Bruneaux et al., 2013; Stölting et al., 2013), and

homogeneously distributed across the genome at high densities (Du et al., 2012); thus, SNPs are the most common genetic markers for genomic screening (Vera et al., 2013). In fish, SNPs are found every ~100 bp (Pardo et al., 2008; Vera et al., 2013). These properties make SNPs ideal for comparative genomics (Bouza et al., 2012), evolutionary genomics (Du et al., 2012), fine mapping of genes associated with productive traits (Sánchez et al., 2009), genomic screening in populations for conservation and management (Albaina et al., 2013; Glaubitz et al., 2003; Kuhner et al., 2000; Vilas et al., 2015), and hybridization and impact of biological invasions (Bers et al., 2010).

Abbreviations: NGS, Next Generation Sequencing; SNP, Single Nucleotide Polymorphisms; RRL, Reduced Representation Library; RADseq, Restriction-site associated DNA sequencing; ddRAD-seq, double-digestion RADseq; RE, restriction enzyme; PCR, polymerase chain reaction; H_E , expected heterozygosity; H_O , observed heterozygosity; MAF, minimum allele frequency; H-W, Hardy-Weinberg; F_{IS} , inbreeding coefficient; F_{ST} , coefficient of relative differentiation between populations; NCBI, National Center for Biotechnology Information; BLAST, Basic Local Alignment Search Tool; WGS, whole genome shotgun; nr, non redundant; ts, transition; tv, transversion

* Corresponding author at: Departamento de Morfología, Instituto de Biociências, UNESP, Botucatu, São Paulo, 18618-970, Brazil.

** Corresponding author.

E-mail addresses: rsicchar@gmail.com (Z.R. Siccha-Ramirez), paulino.martinez@usc.es (P. Martínez).

¹ These authors contributed equally to this work.

<http://dx.doi.org/10.1016/j.fishres.2017.09.002>

Received 3 March 2017; Received in revised form 30 August 2017; Accepted 4 September 2017

Available online 18 September 2017

0165-7836/ © 2017 Elsevier B.V. All rights reserved.

Reduced Representation Library (RRL) sequencing provides a high potential for efficient SNP discovery at very low cost (Sánchez et al., 2009), because sequencing is targeted to a small and adjustable genome fraction flanking the chosen RE sites (Du et al., 2012; Robledo et al., 2017). Since restriction sites are common to most individuals of a species, high throughput sequencing using DNA barcodes to label individuals, populations or species provides a set of reads at each restriction site to be explored bioinformatically. Depending on the goals of the study, SNP identification and genotyping can be jointly performed as a single step, or alternatively a subset of SNPs can be selected for further validation and genotyping using specific platforms (SNaP-shot, Sequenom, Truseq, Ampliseq, etc.) (Davey et al., 2011; Robledo et al., 2017). This last option offers the possibility of developing a panel of adjustable SNP number to be routinely used at lower cost depending on the goals of the study. RADseq is a RRL strategy which has gained popularity in non-model organisms because it allows obtaining huge SNP genotyping data at very low cost (Baird et al., 2008; Gompert et al., 2010; Vandepitte et al., 2013). ddRAD-seq is a variant of the classical RADseq (Baird et al., 2008), which eliminates random shearing and explicitly uses size selection to recover a tunable number of regions according to the objectives of the study (Peterson et al., 2012). ddRAD not only allows high-throughput multiplexed sequencing amenable for genotyping, but also provides improved efficiency and robustness compared to classical RADseq (Zhou et al., 2014). However, the short length of reads of the commonly used platforms for RADseq hampers the subsequent development of SNP panels for genotyping in PCR-based platforms, if no reference genome is available. Consequently, combining ddRAD libraries with longer-read sequencing platforms could aid to solve this drawback making affordable the selection and validation of a subset of SNPs.

The yellowfin tuna (*Thunnus albacares*, Scombridae) is a migratory species with high commercial value found in tropical and subtropical waters all over the world (Guo et al., 2016). This species is currently overfished and appears on the Red List as near threatened species (Collette et al., 2011). A variety of studies have been done to assess the population structure of *T. albacares* using different approaches, including genetic markers (Aires-da-Silva and Maunder, 2012; Alvarado Bremer et al., 1997; Appleyard et al., 2001; Dammannagoda et al., 2008; Ely et al., 2005; Kunal et al., 2013; Ward et al., 1994; Wu et al., 2009), but its genetic structure is still controversial. The Spanish mackerel (*Scomberomorus brasiliensis*, Scombridae) (Collette et al., 1978) is a neritic tuna (Fonteles Filho, 1989) distributed from Belize to south Brazil (Carvalho-Filho, 1999) of high commercial interest in Trinidad and Tobago and Venezuela (Hodgkinson-Clarke, 1990), and in Brazil, especially in the state of Maranhão (Fonteles-Filho, 1988). Few studies have been conducted in this species (Batista and da Fabr e, 2001; Maia et al., 2015), and only one aimed at evaluating its genetic structure (Gold et al., 2010). This work describes how the identification and validation of a large number of SNPs is achievable combining ddRAD and long read high throughput sequencing (e.g. 454 pyrosequencing) in *T. albacares* and *S. brasiliensis* for their routine genotyping on a Sequenom platform. Validated SNPs were further tested on a preliminary analysis of genetic diversity and structure in both species using two geographically separated Brazilian samples.

2. Material and methods

2.1. Biological material and DNA extraction

Samples of both species were collected in Brazil from areas separated by ~3000 km: *T. albacares* was collected in Natal (n = 50, 05°50'24.90"S, 34°59'24.34"W) and Santos (n = 24, 24°16'24.88"S, 45°49'51.79"W) and *S. brasiliensis* in Penha (n = 29, 26°47'24.14"S, 48°33'37.43"W) and Bragança (n = 37, 00°49'4.63"S, 46°30'0.40"W). A total of 11 *T. albacares* and 21 *S. brasiliensis* individuals including samples from both populations were used for *in silico* SNP discovery and

subsequent validation on a Sequenom platform. Seventy-four individuals of *T. albacares* and 66 of *S. brasiliensis* were genotyped using validated SNPs for a preliminary evaluation of genetic diversity and structure of species. Representative specimens and all tissues were deposited in the fish collection of Laborat rio de Biologia and Gen tica de Peixes of the Universidade Estadual Paulista (Botucatu, S o Paulo, Brazil).

Genomic DNA was obtained from ethanol-preserved tissues. Samples were lysed in 300 µl of SSTNE extraction buffer (Blanquer, 1990) plus SDS (0.1%) and 5 µl of proteinase K (20 mg/ml) for 3 h at 55 °C. After 20 min at 70 °C, samples were treated with 7.5 µl of RNase (10 mg/ml) for 1 h at 37 °C for RNA degradation. Total DNA was purified after protein precipitation (5 M NaCl) with freezing absolute ethanol (1 ml). DNA quality (high molecular weight > 20 kb) was first evaluated on agarose gels and the DNA quantity was measured using the NanoDrop[®] ND-1000 spectrophotometer (NanoDrop[®] Technologies Inc) and PicoGreen kit (Molecular Probes) according to the kit instructions. Finally, DNA concentration was accurately measured on a Qubit fluorometer (Life Technologies).

2.2. Library construction

A reduced fraction of the genome of the two species was sequenced using a modified ddRAD protocol (Peterson et al., 2012) in a single 454 GS Junior run (Fig. 1). Briefly, the same amount of DNA (78 ng) of each sample (11 *T. albacares* and 21 *S. brasiliensis*) was individually digested with *Sbf*I and *Sph*I restriction enzymes (RE). Subsequently adaptors for each RE were ligated to both ends of digested fragments. Adaptors included: i) complementary cohesive ends for the correspondent RE; ii) barcodes to identify individuals; and iii) primers for an intermediate PCR amplification. Barcoded samples were then pooled and run in agarose gels 1.1% for fragment selection (300–600 bp), followed by extraction using Qiagen MinElute Gel Extraction kit. After selection, the target DNA fragments were amplified by PCR to obtain the required concentration for sequencing in the 454 GS Junior platform: initial denaturation and enzyme activation at 98 °C for 30s; 14 cycles at 98 °C for 10 s (denaturation step), 65 °C for 30 s (annealing) and 72 °C for 30 s (extension); final extension at 72 °C for 5 min. The PCR products were purified using the Qiagen MinElute PCR Purification kit followed by a magnetic bead clean-up/size selection using an equal volume of Beckman Coulter AMPure XP beads. This protocol ensured that only those fragments including *Sbf*I and *Sph*I target sites were amplified and further sequenced.

2.3. Sequencing, assembly and SNP identification and selection

The final library was sequenced in a single shotgun run on a 454 GS Junior sequencer (Roche Diagnostics) available at the Sequencing and Functional Genomics Platform of the University of Santiago de Compostela (USC, Campus Lugo, Spain) (Fig. 1). Sequencing reads were filtered using default parameters, classified per individual according to barcodes and assembled with Newbler software (specifically designed for 454 GS series data). Alignments were then parsed with Tablet (Milne et al., 2010) in order to identify the most consistent contigs and detect SNPs in the assembled sequences. Only contigs containing a sequencing depth > 6 were retained for further analysis to reduce SNPs attributable to sequencing errors.

SNPs were selected according to the presence of enough flanking regions for primer design (± 100 bp) and the absence of other DNA polymorphism (SNPs and indels) in those regions that could interfere with primer annealing. Additionally, only those SNPs with at least three sequences of the least common allele were selected.

2.4. SNP genotyping

In silico selected SNPs were validated and genotyped using the

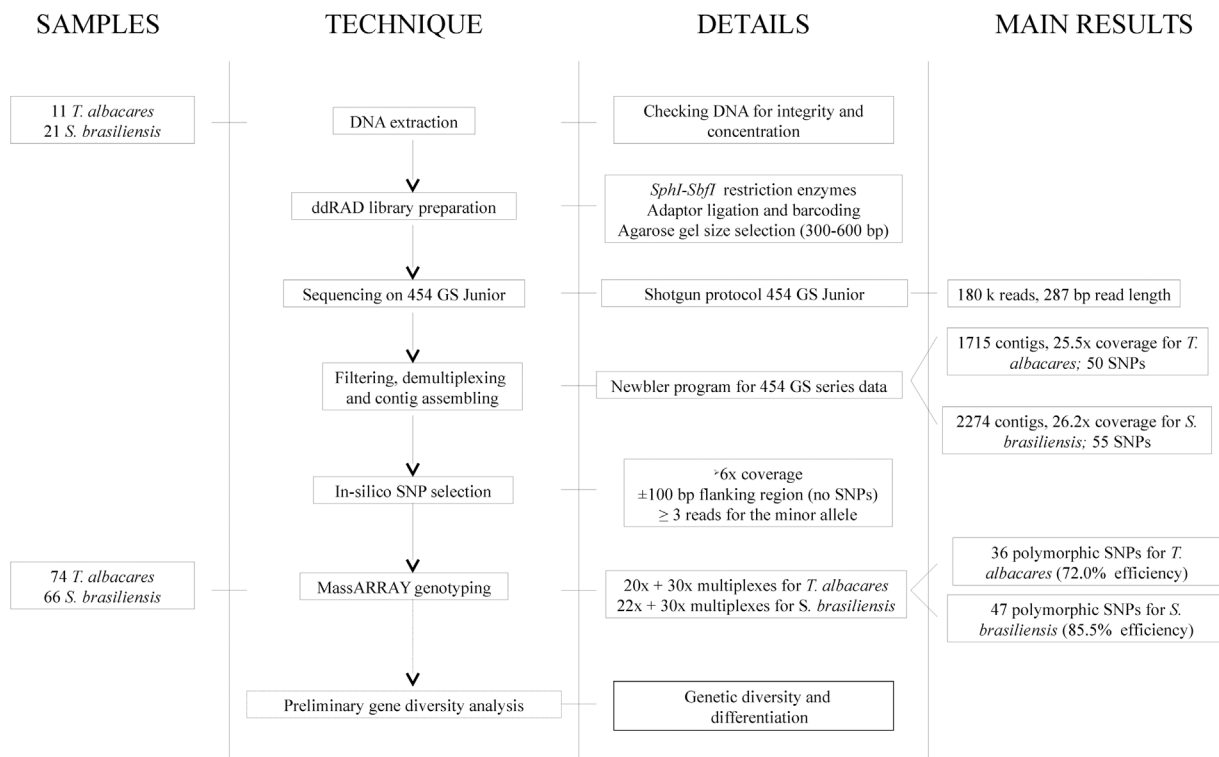


Fig. 1. Workflow of the technical procedure and main results achieved.

MassARRAY platform (Sequenom, San Diego, CA, USA) at the USC node of the Spanish National Centre of Genotyping (CeGen ISCIII) following the protocols and recommendations provided by the manufacturer (Fig. 1). Briefly, target regions were PCR amplified using primers designed on contig sequences obtained in the previous step, and target SNPs were genotyped by single-base extension using an oligonucleotide primer that anneals immediately upstream to the polymorphic site (SNP) of interest (Buetow et al., 2001; Oeth et al., 2009). MALDI-TOF mass spectrometry analysis in an Autoflex spectrometer was used for allele scoring. Two SNPs multiplexes for each species were designed *in silico* using Assay Design 3.1 program (Sequenom, San Diego, CA), which maximizes the number of SNP per multiplex and minimizes the number of multiplexes. Validated SNPs (markers with reliable genotypes) and “failed assays” (majority of genotypes not scored or difficult to cluster according to genotype) were classified by manual inspection. SNPs validated were genotyped on the 74 and 66 individuals of *T. albacares* and *S. brasiliensis*, respectively, for population analysis.

2.5. Gene diversity and annotation

Genepop 1.2 (Raymond and Rousset, 1995; Rousset, 2008) was used to estimate genetic diversity (H_E , H_O and MAF); to check for deviations from H-W equilibrium and their sense (F_{IS}), and from linkage disequilibrium; and to check for population differentiation (F_{ST}). The complete enumeration approach was used to check for H-W equilibrium (Louis and Dempster, 1987) and the Weir and Cockerham method (Weir and Cockerham, 1984) to check the sense and magnitude of the deviation (F_{IS}). Linkage (genotypic) disequilibrium was analyzed for each pair of loci using the log likelihood ratio statistic (G-test). Bonferroni correction was considered when multiple tests were performed.

BLASTn was used to look for significant hits in whole genome sequencing (WGS) databases of six fish species selected among those phylogenetically closer and with best characterized genomes: Pacific bluefin tuna (*Thunnus orientalis*), fugu (*Takifugu rubripes*), stickleback (*Gasterosteus aculeatus*), zebrafish (*Danio rerio*), medaka (*Oryzias latipes*) and tetraodon (*Tetraodon nigroviridis*). Threshold significance was set at

$10e-5$. In addition to BLASTn searching, SNP containing sequences in both species were blasted against nr protein database from NCBI, using BLASTx.

3. Results and discussion

3.1. SNP discovery and validation

The Roche 454 GS Junior run rendered a total of 246,663 reads, 180,779 of them passing the quality filter (73.3%) (Fig. 1). Average read length was 287.0 bp and the average Phred quality per read was 30.2. High quality reads were separated per species using barcodes and then assembled. A total of 1715 contigs were obtained for *T. albacares*, with an average length of 374.9 bp and average coverage of 25.5 reads per contig, while for *S. brasiliensis*, the number of contigs was 2274, with an average length of 374.6 bp and an average coverage of 26.2 reads (Table 1). The coverage achieved in the 454 GS Junior platform was enough to detect a high amount of SNPs in the assembled contigs distributed across the genome, representing a reservoir of markers for future studies in these species depending on the goals pursued. An alternative approach to reduce the genome looking for SNPs is to target coding regions for sequencing. This has been carried out in different fish species like Atlantic herring (*Clupea harengus*, Helyar et al., 2012) and pacú (*Piaractus mesopotamicus*, Mastrochirico-Filho et al., 2016) starting from cDNA of specific tissues sequenced on a Roche 454 platform. However, this approach requires very fresh samples for RNA extraction and storage to construct cDNA libraries, and this is not always achievable, especially in deep sea fishing species like many tuna.

A total of 60 SNPs for *T. albacares* and 79 for *S. brasiliensis* were initially selected for subsequent validation according to the criteria outlined before. Contigs containing selected SNPs averaged 290.1 bp for *T. albacares* and 298.7 bp for *S. brasiliensis*. After manual inspection, 50 SNPs (83.3% of the 60 SNPs selected) for *T. albacares* and 55 (69.6% of the 79 SNPs selected) for *S. brasiliensis* were finally chosen for validation (Fig. 1). The design of flanking primers for subsequent amplification and genotyping of selected loci was feasible thanks to the length

Table 1
Characteristics of 454 GS Junior run and genotyping SNPs.

Sequencing results	Roche 454 GS-Junior stats	
Number of HQ reads	180,779	
Total megabases (Mb)	51,879,629	
Average length of reads	287	
N° individuals sequenced	32	
Assembly results	<i>T. albacares</i>	<i>S. brasiliensis</i>
N° individuals sequenced	11	21
Number of aligned reads	42,875	59,124
Total n° of contigs	1715	2274
Average contig length	374.9	374.6
Average coverage per contig	25.3	26.2
Genotyping	<i>T. albacares</i>	<i>S. brasiliensis</i>
N° individuals	74	66
N° markers	50	55
N° variable markers	36	47

of the contigs obtained relying on the long read sequencing platform used. Markers were combined in two multiplex reactions for each species including 30 and 20 SNPs each for *T. albacares* and 30 and 25 for *S. brasiliensis*, respectively (Supporting information Tables S1 and S2). A total of 36 and 47 SNPs *T.* were polymorphic in *T. albacares* and *S. brasiliensis*, respectively. This represents a validation success of 72.0% and 85.5%, respectively, in the range previously documented (Vera et al., 2013). The number of transitions (ts) and transversions (tv) for SNPs was calculated in both species. Out of 36 variable markers in *T. albacares* dataset, 30 (83.3%) were transitions and 6 (16.7%) transversions, while in *S. brasiliensis* 17 SNPs (36.2%) represented transitions out of 47 markers scored and the remaining 30 (63.8%), transversions (Fig. 2). Accordingly, ts/tv ratio was greatly different between both species (5.000 vs 0.567, respectively), and further, in the extremes of the range regarding previous ts:tv ratios in other fish species (~1.5, Cenadelli et al., 2007; Vera et al., 2013; Zhu et al., 2012). We cannot discard some bias produced during the process of SNP selection and further data will be required to confirm the observed trend in both species. Raw sequencing reads for both species have been deposited in NCBI Sequence Read Archive (SRA) (accession number SRR5998416).

The consensus sequences of the selected 105 SNP-containing contigs were used for searching in NCBI public databases to look for annotation. As expected, the best significant BLASTn hits were obtained against the available *T. orientalis* genome (Supporting information Table S3). Average E-values were much higher for *T. albacares* (8 E-63) than for *S. brasiliensis* (3.71 E-28; Supporting information Table S3), an expected outcome considering the closer phylogenetic relationship between both congeneric species. Twelve and 14 sequences out of 105 showed a significant hit (E-value < 10⁻⁵) against NCBI's nr protein database for *T. albacares* and for *S. brasiliensis*, respectively (Supporting

information Table S1), but mostly showing a not consistent annotation (predicted, hypothetical or unnamed proteins). The proportion of significant hits to protein databases (13%) is in the range expected considering the ~800 Mb of the Pacific bluefin tuna genome (*Thunnus orientalis*; Nakamura et al., 2013), where exons represent around 10% of the genome. The low consistency of annotations could be in part due to the high diversification and annotation success of fish, but a bias in the amplicon distribution due to the ddRAD protocol cannot be discarded. Globally, our results are very similar to those obtained in a parallel study in stingrays from South America using the same methodology (Cruz et al., 2017), confirming the consistency of the method. A similar approach was used by Jansson et al. (2016) to detect and genotype SNPs in goldsinny wrasse (*Ctenolabrus rupestris*), but using Illumina MiSeq for sequencing fragments obtained from ddRAD methodology. The length of the reads in this study was smaller than in ours (160 bp vs 287 bp), thus limiting the SNPs useful for validation due to the smaller flanking regions for primer design. However, the higher coverage of Illumina vs Roche 454, compensated this drawback and a high amount of contigs with SNPs were available for validation, which was still higher than in our study (92.3% vs 85.5%–72%). The longer read technology recently released by Illumina MiSeq (~600 bp using pair-end sequencing) should ensure longer contigs and high coverage, thus representing the best option for this approach, but still needs for validation.

3.2. Genetic diversity and structure

A preliminary evaluation of genetic diversity and structure was addressed with the validated SNPs in both species using two populations separated by around 3000 km. The MAF for *T. albacares* ranged from 0.007 to 0.464, with an average of 0.177, and for *S. brasiliensis* from 0.008 to 0.485, averaging 0.209 (Supplementary Table 2). H_E ranged from 0.024 to 0.495 (average 0.180) for *T. albacares* and from 0.020 to 0.500 (average 0.240) for *S. brasiliensis*. The average heterozygosity for *S. brasiliensis* (0.240) was similar to that detected in other tuna species (*Thunnus alalunga*: 0.278 and *T. thynnus*: 0.272, Albaina et al., 2013) and within the range described in other marine species like European sea bass (*Dicentrarchus labrax*; Souche et al., 2015), Atlantic herring (*Clupea harengus*), turbot (*Scophthalmus maximus*; Vera et al., 2013) and Asian sea bass (*Lates calcarifer*; Wang et al., 2015). In the case of *T. albacares*, genetic diversity was somewhat lower than the aforementioned figures, and further work will be needed to confirm if this is a genomic feature of the species, but it is possibly bias due to the number of markers used. Hardy-Weinberg exact tests only detected deviations due to heterozygote deficit ($F_{IS} > 0$): at one locus in *T. albacares* ($p < 0.05$), but none after Bonferroni correction, and at eight loci in *S. brasiliensis* ($p < 0.05$), six of them after Bonferroni correction (Supporting information Table S2). The consistency of the deviations across populations and its magnitude strongly suggests the presence of null alleles at these loci in *S. brasiliensis*. However, we cannot discard some degree of population admixture in this species (e.g. Walhund effect), considering the structuring suggested for this species (Gold et al.,

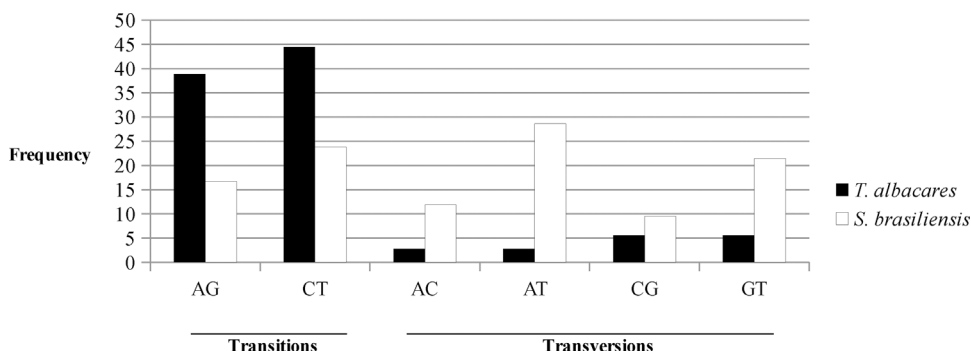


Fig. 2. Frequency of transitions and transversions in the SNPs validated for *Thunnus albacares* and *Scomberomorus brasiliensis*.

2010). No pair of loci showed significant deviation from linkage equilibrium after Bonferroni correction in both species (630 and 859 tests, respectively).

Our analysis suggested the absence of significant genetic structure in both *T. albacares* and *S. brasiliensis*. For *T. albacares*, our results are in accordance with previous reports carried out with microsatellites and SNPs that could not find significant population structure within oceanic basins likely related to its high migration capacity (Ely et al., 2005; Pecoraro et al., 2016a). Grewe et al. (2015) suggested the existence of differentiation in this species within the Pacific Ocean using 215 SNPs potentially under selection, but no differentiation was found at the remaining 5054 presumably neutral loci. Pecoraro et al. (2016b) found discrete genetic structure only between different ocean basins employing 6772 SNPs. Conversely, a previous study of the Spanish mackerel detected significant genetic differentiation between geographic areas (Gold et al., 2010). In our study, despite the overall absence of significant differentiation (F_{ST}), five loci showed significant F_{ST} values ranging between 6.2% and 10.3% (Supporting information Table S2). Furthermore, we cannot discard that the heterozygote deficit detected at several loci in this species could be related to population admixture suggestive of underlying structuring, which suggests the need for further work on the species using a representative sample across its distribution area.

4. Conclusions

A combination of RADseq (ddRAD) and long read high throughput sequencing (454 pyrosequencing) was successfully used to identify a high amount of SNPs in two non-model fish species lacking a reference genome. The longer sequences obtained with this sequencing approach facilitated primer design for validating and genotyping a subset of SNPs accommodated to the goals of the study in a MassARRAY platform. This method enabled repeatable and tunable recovery of thousands of sampled regions from *T. albacares* and *S. brasiliensis* genomes demonstrating its utility to identify hundreds of SNPs for population studies applicable to other non-model species. This approach is faster and cheaper than other high throughput SNP genotyping methods when the number of SNPs required is not too high (i.e. ~50–100), which is though enough for estimating population parameters with reasonable accuracy. SNPs developed with this approach demonstrated to be useful for a preliminary evaluation of genetic diversity and structure on *T. albacares* and *S. brasiliensis*.

Acknowledgements

We are grateful to the following people, who helped during the sampling process: Leonardo Machado (UNESP), Jonathan Ready (UFGA), Ricardo Britzke (UNESP) and all fishery owners who kindly help us to get the samples from their boats. We thank Susana Sánchez, Lucía Insua and María Villar for technical support. We also thank CeGen for the genotyping of our samples. This work was supported by São Paulo Research Foundation (grant 2011/00881-1 and grant 2013/07560-1) and the investigation funding provided by Xunta de Galicia local Government (GRC2014/010).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fishres.2017.09.002>.

References

Aires-da-Silva, A., Maunder, M.N., 2012. Status of bigeye tuna in the eastern Pacific Ocean in 2010 and outlook for the future. IATTC Stock Assess. Rep. 13.
Albaina, A., Iriondo, M., Velado, I., Laconcha, U., Zarraonaindia, I., Arrizabalaga, H., Pardo, M.A., Lutcavage, M., Grant, W.S., Estonba, A., 2013. Single nucleotide

polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure. *Anim. Genet.* 44, 678–692. <http://dx.doi.org/10.1111/age.12051>.

Alvarado Bremer, J.R., Naseri, I., Ely, B., 1997. Orthodox and unorthodox phylogenetic relationships among tunas revealed by the nucleotide sequence analysis of the mitochondrial DNA control region. *J. Fish Biol.* 50, 540–554. <http://dx.doi.org/10.1111/j.1095-8649.1997.tb01948.x>.

Appleyard, S.A., Renwick, J.M., Mather, P.B., 2001. Individual heterozygosity levels and relative growth performance in *Oreochromis niloticus* (L.) cultured under Fijian conditions. *Aquac. Res.* 32, 287–296. <http://dx.doi.org/10.1046/j.1365-2109.2001.00557.x>.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3, e3376. <http://dx.doi.org/10.1371/journal.pone.0003376>.

Batista, V., da Fabrã, N.N., 2001. Temporal and spatial patterns on serra, *Scomberomorus brasiliensis* (Teleostei, Scombridae), catches from the fisheries on the Maranhão coast, Brazil. *Braz. J. Biol.* 61, 541–546. <http://dx.doi.org/10.1590/S1519-69842001000400003>.

Bers, N.E.V., Oers, K., van Kerstens, H.H., Dibbitts, B.W., Croijmans, R.P., Visser, M.E., Groenen, M.A., 2010. Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing. *Mol. Ecol.* 19, 89–99. <http://dx.doi.org/10.1111/j.1365-294X.2009.04486.x>.

Blanquer, A., 1990. *Phylogéographie intraspécifique d'un poisson marin, le flet *Platichthys flesus* L. (Heterosomata): polymorphisme des marqueurs nucléaires et mitochondriaux*. Montpellier 2.

Bouza, C., Hermida, M., Pardo, B.G., Vera, M., Fernández, C., de la Herrán, R., Navajas-Pérez, R., Álvarez-Dios, J.A., Gómez-Tato, A., Martínez, P., 2012. An Expressed Sequence Tag (EST)-enriched genetic map of turbot (*Scophthalmus maximus*): a useful framework for comparative genomics across model and farmed teleosts. *BMC Genet.* 13, 54. <http://dx.doi.org/10.1186/1471-2156-13-54>.

Bruneaux, M., Johnston, S.E., Herczeg, G., Merilä, J., Primmer, C.R., Vasemägi, A., 2013. Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. *Mol. Ecol.* 22, 565–582. <http://dx.doi.org/10.1111/j.1365-294X.2012.05749.x>.

Buetow, K.H., Edmonson, M., MacDonald, R., Clifford, R., Yip, P., Kelley, J., Little, D.P., Strausberg, R., Koester, H., Cantor, C.R., 2001. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci.* 98, 581–584. <http://dx.doi.org/10.1073/pnas.98.2.581>.

Carvalho-Filho, A., 1999. *Peixes: costa brasileira*.

Cenadelli, S., Maran, V., Bongioni, G., Fusetti, L., Parma, P., Aleandri, R., 2007. Identification of nuclear SNPs in gilthead seabream. *J. Fish Biol.* 70, 399–405. <http://dx.doi.org/10.1111/j.1095-8649.2007.01454.x>.

Collette, B.B., Russo, J.L., Zavala-Camin, L.A., 1978. *Scomberomorus brasiliensis*, a new species of Spanish mackerel from the western Atlantic. *Fish. Bull.* 76, 273–280.

Collette, B.B., Carpenter, K.E., Polidoro, B.A., Juan-Jordá, M.J., Boustany, A., Die, D.J., Elfes, C., Fox, W., Graves, J., Harrison, L.R., 2011. High value and long life—double jeopardy for tunas and billfishes. *Science* 333, 291–292. <http://dx.doi.org/10.1126/science.1208730>.

Cruz, V.P., Vera, M., Pardo, B.G., Taggart, J., Martínez, P., Oliveira, C., Foresti, F., 2017. Identification and validation of single nucleotide polymorphisms as tools to detect hybridization and population structure in freshwater stingrays. *Mol. Ecol. Resour.* 17 (3), 550–556.

Dammannagoda, S.T., Hurwood, D.A., Mather, P.B., 2008. Evidence for fine geographical scale heterogeneity in gene frequencies in yellowfin tuna (*Thunnus albacares*) from the north Indian Ocean around Sri Lanka. *Fish. Res.* 90, 147–157. <http://dx.doi.org/10.1016/j.fishres.2007.10.006>.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. <http://dx.doi.org/10.1038/nrg3012>.

Du, Y., Jiang, H., Chen, Y., Li, C., Zhao, M., Wu, J., Qiu, Y., Li, Q., Zhang, X., 2012. Comprehensive evaluation of SNP identification with the restriction enzyme-based reduced representation library (RRL) method. *BMC Genom.* 13, 1. <http://dx.doi.org/10.1186/1471-2164-13-77>.

Ely, B., Viñas, J., Bremer, J.R.A., Black, D., Lucas, L., Covelto, K., Labrie, A.V., Thelen, E., 2005. Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (*Thunnus albacares*) and the skipjack tuna (*Katsuwonus pelamis*). *BMC Evol. Biol.* 5, 19. <http://dx.doi.org/10.1186/1471-2148-5-19>.

Fonteles Filho, A.A., 1989. *Recursos pesqueiros: biologia e dinâmica populacional*.

Fonteles-Filho, A.A., 1988. Sinopse de informacoes sobre a cavala, *Scomberomorus cavalla* (Cuvier) ea serra, *Scomberomorus brasiliensis* Collette, Russo and Zaval-Camin (Pisces: Scombridae), no estado do Ceara, Brasil. *Arq. Defic. Mar Fortaleza* 27, 21–48.

Glaubit, J.C., Rhodes, O.E., DeWoody, J.A., 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol. Ecol.* 12, 1039–1047. <http://dx.doi.org/10.1046/j.1365-294X.2003.01790.x>.

Gold, J.R., Jobity, A.M.C., Saillant, E., Renshaw, M.A., 2010. Population structure of carite (*Scomberomorus brasiliensis*) in waters offshore of Trinidad and northern Venezuela. *Fish. Res.* 103, 30–39. <http://dx.doi.org/10.1016/j.fishres.2010.01.009>.

Gompert, Z., Forister, M.L., Fordyce, J.A., Nice, C.C., Williamson, R.J., Alex Buerkle, C., 2010. Bayesian analysis of molecular variance in pyrosequencing quantifies population genetic structure across the genome of *Lycaeides* butterflies. *Mol. Ecol.* 19, 2455–2473. <http://dx.doi.org/10.1111/j.1365-294X.2010.04666.x>.

Grewe, P.M., Feutry, P., Hill, P.L., Gunasekera, R.M., Schaefer, K.M., Itano, D.G., Fuller,

- D.W., Foster, S.D., Davies, C.R., 2015. Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Sci. Rep.* 5. <http://dx.doi.org/10.1038/srep16916>.
- Guo, L., Li, M., Zhang, H., Yang, S., Chen, X., Meng, Z., Lin, H., 2016. Next-generation sequencing of the yellowfin tuna mitochondrial genome reveals novel phylogenetic relationships within the genus *Thunnus*. *Mitochondrial DNA Part A* 27, 2089–2090. <http://dx.doi.org/10.3109/19401736.2014.982570>.
- Helyar, S.J., Limborg, M.T., Bekkevold, D., Babbucci, M., van Houdt, J., Maes, G.E., Bargelloni, L., Nielsen, R.O., Taylor, M.I., Ogden, R., Cariani, A., Carvalho, G.R., Panitz, F., 2012. SNP discovery using next generation transcriptomic sequencing in Atlantic herring (*Clupea harengus*). *PLoS One* 7, e42089. <http://dx.doi.org/10.1371/journal.pone.0042089>.
- Hodgkinson-Clarke, F.M., 1990. The Carite (*Scomberomorus brasiliensis*) Fishery of South Trinidad: A Comparison of Catch Rates, Catch Composition, Use and Operation of the Monofilament and Multifilament Gillnets. Centre for Resource Management and Environmental Studies, University of the West Indies, Cave Hill Campus.
- Jansson, Eva, et al., 2016. Development of SNP and microsatellite markers for goldsinny wrasse (*Ctenolabrus rupestris*) from ddRAD sequencing data. *Conserv. Genet. Resour.* 8 (3), 201–206. <http://dx.doi.org/10.1007/s12686-016-0532-0>.
- Kuhner, M.K., Beerli, P., Yamato, J., Felsenstein, J., 2000. Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics* 156, 439–447. PMID: PMC1461258.
- Kunal, S.P., Kumar, G., Menezes, M.R., Meena, R.M., 2013. Mitochondrial DNA analysis reveals three stocks of yellowfin tuna *Thunnus albacares* (Bonnaterre, 1788) in Indian waters. *Conserv. Genet.* 14, 205–213. <http://dx.doi.org/10.1007/s10592-013-0445-3>.
- Louis, E.J., Dempster, E.R., 1987. An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* 805–811. <http://dx.doi.org/10.2307/2531534>.
- Maia, R.C., Da Silva, B.B., Pereira, L. de J.G., Holanda, F.C.A.F., 2015. Pesca comercial e estrutura populacional da serra, *Scomberomorus brasiliensis* (Collette, Russo & Zavala, 1978), desembarcada em um pólo pesqueiro na Costa Norte do Brasil. *Biota Amaz. Biote Amaz. Biota Amazon. Amaz. Biota* 5, 99–106. <http://dx.doi.org/10.18561/2179-5746/biotaamazonia.v5n2p99-106>.
- Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24, 133–141. <http://dx.doi.org/10.1016/j.tig.2007.12.007>.
- Mastrochirico-Filho, V.A., Hata, M.E., Sato, L.S., Jorge, P.H., Foresti, F., Vera, M., Martínez, P., Porto-Foresti, F., Hashimoto, D.T., 2016. SNP discovery from liver transcriptome in the fish *Piaractus mesopotamicus*. *Conserv. Genet. Res.* 8, 109–114. <http://dx.doi.org/10.1007/s12686-016-0521-3>.
- Metzker, M.L., 2010. Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11, 31–46. <http://dx.doi.org/10.1038/nrg2626>.
- Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., Marshall, D., 2010. Tablet—next generation sequence assembly visualization. *Bioinformatics* 26, 401–402. <http://dx.doi.org/10.1093/bioinformatics/btp666>.
- Mullikin, J.C., Hunt, S.E., Cole, C.G., Mortimore, B.J., Rice, C.M., Burton, J., Matthews, L.H., Pavitt, R., Plumb, R.W., Sims, S.K., 2000. An SNP map of human chromosome 22. *Nature* 407, 516–520. <http://dx.doi.org/10.1038/35035089>.
- Nakamura, Y., Mori, K., Saitoh, K., Oshima, K., Mekuchi, M., Sugaya, T., Yasuike, M., 2013. Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. *Proc. Natl. Acad. Sci.* 110 (27), 11061–11066.
- Oeth, P., del Mistro, G., Marnellos, G., Shi, T., van den Boom, D., 2009. Qualitative and quantitative genotyping using single base primer extension coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MassARRAY®). *Single Nucleotide Polymorphisms*. Springer, pp. 307–343. http://dx.doi.org/10.1007/978-1-60327-411-1_20.
- Pardo, B.G., Fernández, C., Millán, A., Bouza, C., Vázquez-López, A., Vera, M., Alvarez-Dios, J.A., Calaza, M., Gómez-Tato, A., Vázquez, M., 2008. Expressed sequence tags (ESTs) from immune tissues of turbot (*Scophthalmus maximus*) challenged with pathogens. *BMC Vet. Res.* 4, 1. <http://dx.doi.org/10.1186/1746-6148-4-37>.
- Pecoraro, C., Babbucci, M., Villamor, A., Franch, R., Papetti, C., Leroy, B., Ortega-García, S., Muir, J., Rooker, J., Arocha, F., 2016a. Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*). *Mar. Genom.* <http://dx.doi.org/10.1016/j.margen.2015.12.002>.
- Pecoraro, C., Zudaire, I., Bodin, N., Murua, H., Taconet, P., Díaz-Jaimes, P., Cariani, A., Tinti, F., Chassot, E., 2016b. n. d. Putting all the pieces together: integrating current knowledge of the biology, ecology, fisheries status, stock structure and management of yellowfin tuna (*Thunnus albacares*). *Rev. Fish Biol. Fish.* 1–31. <http://dx.doi.org/10.1007/s11160-016-9460-z>.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7, e37135. <http://dx.doi.org/10.1371/journal.pone.0037135>.
- Raymond, M., Rousset, F., 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248–249.
- Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., Houston, R., 2017. Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* <http://dx.doi.org/10.1111/raq.12193>.
- Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. <http://dx.doi.org/10.1111/j.1471-8286.2007.01931.x>.
- Sánchez, C.C., Smith, T.P., Wiedmann, R.T., Vallejo, R.L., Salem, M., Yao, J., Rexroad, C.E., 2009. Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genom.* 10, 559. <http://dx.doi.org/10.1186/1471-2164-10-559>.
- Souche, E.L., Hellemans, B., Babbucci, M., MacAoidh, E., Guinand, B., Bargelloni, L., Chistiakov, D.A., Patarnello, T., Bonhomme, F., Martinsohn, J.T., 2015. Range-wide population structure of European sea bass *Dicentrarchus labrax*. *Biol. J. Linn. Soc.* 116, 86–105. <http://dx.doi.org/10.1111/bij.12572>.
- Stöltig, K.N., Nipper, R., Lindtke, D., Caseys, C., Waerber, S., Castiglione, S., Lexer, C., 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Mol. Ecol.* 22, 842–855. <http://dx.doi.org/10.1111/mec.12011>.
- Vandepitte, K., Gristina, A.S., De Raedt, R., Roldán-Ruiz, I., Marcenò, C., Sciandrello, S., Honnay, O., 2013. Conservation genetics of an endemic from the Mediterranean Basin: high genetic differentiation but no genetic diversity loss from the last populations of the Sicilian Grape *Hyacinth Leopoldia gussonei*. *Conserv. Genet.* 14, 963–972. <http://dx.doi.org/10.1007/s10592-013-0487-6>.
- Vera, M., Alvarez-Dios, J.-A., Fernandez, C., Bouza, C., Vilas, R., Martinez, P., 2013. Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. *Int. J. Mol. Sci.* 14, 5694–5711. <http://dx.doi.org/10.3390/ijms14035694>.
- Vilas, R., Vandamme, S.G., Vera, M., Bouza, C., Maes, G.E., Volckaert, F.A., Martinez, P., 2015. A genome scan for candidate genes involved in the adaptation of turbot (*Scophthalmus maximus*). *Mar. Genom.* 23, 77–86. <http://dx.doi.org/10.1016/j.margen.2015.04.011>.
- Wang, L., Wan, Z.Y., Lim, H.S., Yue, G.H., 2015. Genetic heterogeneity and local adaptation of Asian seabass across Indonesian Archipelago revealed with gene-associated SNP markers. *Fish. Res.* 170, 205–211. <http://dx.doi.org/10.1016/j.fishres.2015.06.012>.
- Ward, R.D., Elliott, N.G., Grewe, P.M., Smolenski, A.J., 1994. Allozyme and mitochondrial DNA variation in yellowfin tuna (*Thunnus albacares*) from the Pacific Ocean. *Mar. Biol.* 118, 531–539. <http://dx.doi.org/10.1007/BF00347499>.
- Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. <http://dx.doi.org/10.2307/2408641>.
- Wu, G.C.-C., Chiang, H.-C., Chen, K.-S., Hsu, C.-C., Yang, H.-Y., 2009. Population structure of albacore (*Thunnus alalunga*) in the Northwestern Pacific Ocean inferred from mitochondrial DNA. *Fish. Res.* 95, 125–131. <http://dx.doi.org/10.1016/j.fishres.2008.07.014>.
- Zhou, X., Xia, Y., Ren, X., Chen, Y., Huang, L., Huang, S., Liao, B., Lei, Y., Yan, L., Jiang, H., 2014. Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). *BMC Genom.* 15, 351. <http://dx.doi.org/10.1186/1471-2164-15-351>.
- Zhu, C., Cheng, L., Tong, J., Yu, X., 2012. Development and characterization of new single nucleotide polymorphism markers from expressed sequence tags in common carp (*Cyprinus carpio*). *Int. J. Mol. Sci.* 13, 7343–7353. <http://dx.doi.org/10.3390/ijms13067343>.