Molecular Phylogenetics and Evolution 102 (2016) 202–207

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas

Natalia Díaz-Arce [a], Haritz Arrizabalaga [a], Hilario Murua [a], Xabier Irigoien [b], Naiara Rodríguez-Ezpeleta [a,*]

[a] AZTI, Marine Research Division, Sukarrieta, Bizkaia, Spain
[b] King Abdullah University of Science and Technology, Red Sea Research Center, Biological and Environmental Sciences & Engineering Division, Thuwal, Saudi Arabia

ARTICLE INFO

Article history:
Received 9 May 2016
Accepted 2 June 2016
Available online 7 June 2016

Keywords:
Phylogeny
Restriction-site associated DNA sequencing
Genome-wide markers
Phylogenomics
Tuna

ABSTRACT

Although species from the genus *Thunnus* include some of the most commercially important and most severely overexploited fishes, the phylogeny of this genus is still unresolved, hampering evolutionary and traceability studies that could help improve conservation and management strategies for these species. Previous attempts based on mitochondrial and nuclear markers were unsuccessful in inferring a congruent and reliable phylogeny, probably due to mitochondrial introgression events and lack of enough phylogenetically informative markers. Here we infer the first genome-wide nuclear marker-based phylogeny of tunas using restriction site associated DNA sequencing (RAD-seq) data. Our results, derived from phylogenomic inferences obtained from 128 nucleotide matrices constructed using alternative data assembly procedures, support a single *Thunnus* evolutionary history that challenges previous assumptions based on morphological and molecular data.

1. Introduction

Species of the genus *Thunnus* include some of the most economically important, but also most severely overexploited fish on the planet. Yet, despite its relevance for enabling more efficient management plans and avoiding masked trade of mislabeled tuna species by providing tools for DNA-based species identification, the phylogeny of the genus *Thunnus* remains unresolved. According to morphological features, this genus was divided into two subgenera: the temperate *Thunnus* (bluefin tuna group), comprising the albacore (*Thunnus alalunga*) and the Atlantic (*Thunnus thynnus*), Pacific (*Thunnus orientalis*) and Southern (*Thunnus maccoyii*) bluefin tunas, and the tropical *Neothunnus* (yellowfin tuna group), comprising the blackfin (*Thunnus atlanticus*), longtail (*Thunnus tonggol*) and yellowfin (*Thunnus albacares*) tunas. Although sharing a similar number of morphological features with both groups, the bigeye tuna (*Thunnus obesus*) has been included into the subgenus *Thunnus* due to its adaptation to cooler waters (Collette et al., 2001).

Most of molecular phylogenies aimed at solving the relationships among tuna species are based on mitochondrial markers (Alvarado Bremer et al., 1997; Chow and Kishino, 1995; Chow et al., 2006; Tseng et al., 2012; Viñas and Tudela, 2009) and/or do not include all the species of the genus *Thunnus* (Miya et al., 2013; Orrell et al., 2006; Santini et al., 2013). These studies consistently recover the *Neothunnus* subgenus and a close relationship between the albacore and the Pacific bluefin tuna. Yet, evidence of mitochondrial introgression in *T. orientalis* with *T. thynnus* or *T. alalunga* (Chow and Inoue, 1993; Chow and Kishino, 1995; Chow et al., 2006), and in *T. thynnus* with *T. alalunga* or *T. orientalis* (Alvarado Bremer et al., 2005; Chow et al., 2006; Viñas et al., 2003; Takeyama et al., 2001; Viñas and Tudela, 2009) makes mitochondrial-based inferences of the relationships between these three species ambiguous. The only nuclear based phylogenetic studies that include all eight species of *Thunnus* (Chow et al., 2006; Viñas and Tudela, 2009) group the Atlantic and Pacific bluefin tunas in a well-supported clade. Thus, overall, the analyses published to date support the monophyly of the *Neothunnus* subgenus, but do not provide resolution for the relationships within this clade nor congruence or support for the relationships within *Neothunnus* or among *T. maccoyii, T. obesus, T. alalunga* and the *T. orientalis* + *T. thynnus* group. Besides preventing accurate inferences of their evolutionary history, this lack of congruence or resolution is translated into a lack of suitable DNA-based tools for tuna species discrimination. The few genetic markers in use for this purpose (e.g. Bartlett and Davidson, 1991; Chow et al., 2003; Pardo and Pérez-Villareal, 2004; Takeyama et al., 2001; Viñas and Tudela, 2009) are mitochondrial, and the sole contrasted nuclear marker, the ribosomal

* Corresponding author at: AZTI, Marine Research Division, Txatxarramendi ugartea z/g, Sukarrieta 48395, Bizkaia, Spain.
E-mail address: nrodriguez@azti.es (N. Rodríguez-Ezpeleta).

http://dx.doi.org/10.1016/j.ympev.2016.06.002
1055-7903/

internal transcribed spacer (ITS1), can only be used as a complement to mitochondrial markers when introgression is suspected (Viñas and Tudela, 2009), as it does not discriminate between all eight *Thunnus* species (Chow et al., 2006; Viñas and Tudela, 2009). Thus, overcoming the lack of congruence and/or resolution associated to the published *Thunnus* evolutionary relationships requires evolutionary inferences based on genome-wide phylogenetically informative positions of nuclear origin.

Recently, the restriction site-associated DNA sequencing (RAD-seq) method, which, allows to rapidly and cost-efficiently sequence thousands of homologous regions in hundreds of individuals both, with and without available reference genomes, has been applied to resolve phylogenetic relationships (e.g. Cruaud et al., 2014; Herrera and Shank, 2015; Leaché et al., 2015). Yet some studies have shown that gathering a suitable set of phylogenetically informative markers from RAD-seq data relies on the ability to discover enough orthologous loci among the species under study, which largely depends on divergence times between lineages and filtering and assembly parameters applied for orthology inference (Jones et al., 2013; Leaché et al., 2015; Rubin et al., 2012; Wagner et al., 2013). It is therefore recommended to explore the results applying different parameter combinations (Ree and Hipp, 2015; Rubin et al., 2012). Here, we infer the evolutionary history of tunas based on phylogenomic analyses of RAD-seq derived nuclear markers. With the final aim of building a robust phylogenetic tree, we have explored alternative procedures for selecting phylogenetically informative sites and built 128 RAD-seq derived nucleotide matrices obtained by (i) using different parameters for putative orthologous loci identification, (ii) including different sets of species, (iii) selecting variable or fixed sites (iv) within individuals or within species and (v) allowing different thresholds of missing individuals or species to select a locus. Our analyses highlight the influence of RAD-seq data analyses procedures in derived nucleotide matrices and phylogenetic inferences, and provide the first genome-wide resolved evolutionary tree of the *Thunnus* genus. The inferred relationships restructure the *Neothunnus* subgenus including *T. obesus* within this group, and clarify the relationships between the Altantic and Pacific bluefin tunas and albacore, setting the root of the genus within the latter. Our results establish the basics for future evolutionary studies of these species and provide valuable data for developing species identification and traceability tools that will assist better management and conservation of tunas.

2. Materials and methods

2.1. Sampling and genomic DNA extraction

Samples from *T. thynnus* (nine individuals), *T. albacares*, *T. atlanticus*, *T. orientalis, T. tonggol* (five individuals), *T. alalunga*, *T. maccoyii* and *T. obesus* (four individuals) and from three other Scombridae species included as outgroup (*Katsuwonus pelamis* – four individuals, *Euthynnus alletteratus* – five individuals, and *Auxis rochei* – three individuals) were obtained from scientific surveys and commercial fisheries. From each fish, about 1 cm^3 of muscle tissue was resected and immediately stored in 96% molecular grade ethanol at −20 °C. Genomic DNA was extracted from about 20 mg of muscle tissue using the Wizard® Genomic DNA Purification kit (Promega, WI, USA) following manufacturer's instructions for "Isolating Genomic DNA from Tissue Culture Cells and Animal Tissue". Extracted DNA was suspended in Milli-Q water and concentration was determined with the Quant-iT dsDNA HS assay kit using a Qubit® 2.0 Fluorometer (Life Technologies). DNA integrity was assessed by electrophoresis, migrating about 100 ng of GelRed™-stained DNA on an agarose 1.0% gel.

2.2. Restriction site associated DNA sequencing library preparation and analysis

Restriction-site-associated DNA libraries were prepared following the methods of Etter et al. (2011). Briefly, about 300 ng of genomic DNA were digested with the *SbfI* restriction enzyme and ligated to modified Illumina P1 adapters containing 5 bp unique barcodes. Pools of 33 individuals were sheared using the Covaris® M220 Focused-ultrasonicator™ Instrument (Life Technologies) and size selected to 300–500 pb by cutting agarose migrated DNA. After Illumina P2 adaptor ligation, library was amplified using 14 PCR cycles. Each pool was sequenced (100 pb) on an Illumina HiSeq2000 lane. RAD sequencing data were processed with Stacks version 1.27 (Catchen et al., 2013) with default parameters unless otherwise specified. Raw sequences were demultiplexed and filtered for low quality using the *process_radtags* module; for each individual, putative loci were identified using *ustacks* allowing a minimum stack depth parameter of 5 (parameter *m*) and 1 or 2 mismatches (parameter *M*). Catalogs of loci were built based on two different subsets of individuals (all individuals and only *Thunnus* individuals) using *cstacks* allowing 4 or 8 mismatches (parameter *n*). A total of 8 different catalogs were produced. Individual sets of loci were matched against the catalog using *sstacks*. From the 8 catalogs built, the *populations* program was used to select phylogenetically informative markers based on all possible combinations of (i) considering individuals or species for informative marker selection, (ii) considering fixed (within individuals or species) or IUPAC encoded variable sites (can include heterozygous sites when using individuals), and (iii) allowing different thresholds (0%, 25%, 50% or 75%) of missing individuals or species for marker selection. A total of 128 matrices (16 per catalog) including only the phylogenetically informative sites within each locus were produced for phylogenetic analyses. Maximum Likelihood phylogenetic trees were built using the unpartitioned GTRCAT model as implemented in RAxML version 8.1.21 (Stamatakis, 2014) and branch support was assessed by a 100 replicate rapid-bootstrap analysis.

3. Results and discussion

3.1. RAD-seq data preprocessing

The number of reads per individual that met the quality requirements ranges from 1,529,640 to 6,645,450, with an average of 3,266,093, of which from 92% to 99% per individual were used for stacks (RAD loci) formation. As expected, higher M values increase coverage (53× vs 50×) and produce fewer RAD loci per individual (Fig. 1). The number of estimated *SbfI* cut sites (about 30,000–32,000 restriction sites for *Thunnus* species and slightly higher for outgroup species) show that the RAD-seq approach using this restriction enzyme allows to cost-effectively produce high coverage orthologous markers in these species, which is relevant for future population genetics studies on these taxa.

3.2. Assembly of phylogenetically informative nucleotide matrices

The number of nucleotide positions included in the matrices ranges from 2625 to 426,052 varying substantially depending on the catalog building and filtering parameters applied (Table 1). In general, allowing more missing positions, using IUPAC encoded sites instead of fixed positions and using individuals instead of species to select positions results in larger matrices. When missing data are allowed, individual-based site selection results in larger matrices both for IUPAC encoded and fixed positions, which is due to the fact that even when missing data are allowed, markers