Accepted Manuscript

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

Natalia Díaz-Arce, Haritz Arrizabalaga, Hilario Murua, Xabier Irigoien, Naiara Rodríguez-Ezpeleta

PII: S1055-7903(16)30138-5

DOI: <http://dx.doi.org/10.1016/j.ympev.2016.06.002>

Reference: YMPEV 5552

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 9 May 2016

Accepted Date: 2 June 2016

Please cite this article as: Díaz-Arce, N., Arrizabalaga, H., Murua, H., Irigoien, X., Rodríguez-Ezpeleta, N., RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas, *Molecular Phylogenetics and Evolution* (2016), doi: <http://dx.doi.org/10.1016/j.ympev.2016.06.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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*}

^aAZTI, Marine Research Division, Sukarrieta, Bizkaia, Spain

^bKing Abdullah University of Science and Technology (KAUST), Biological and Environmental Science and Engineering Division, Red Sea Research Center, Thuwal. Saudi Arabia.

Corresponding author: Naiara Rodriguez-Ezpeleta, AZTI, Marine Research Division, Txatxarramendi ugarte 2/g, Sukarrieta, 48395, Bizkaia, Spain; nrodriguez@azti.es

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.

Keywords: phylogeny; Restriction-site Associated DNA sequencing; genome-wide markers; phylogenomics; tuna

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* (Bremer et al., 2005; Chow et al., 2006; Pla 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; Tseng et al., 2011; Viñas and Tudela, 2009) are mitochondrial, and the sole contrasted nuclear marker, the ribosomal 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 Atlantic 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 cm³ 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 5bp 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 (53x vs 50x) and produce fewer RAD loci per individual (Figure 1).

The number of estimated *SbfI* cut sites (about 30,000 to 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 2,625 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 present in all individuals within the same species are scarcer. However, when no missing data are allowed, the number of markers is similar whether using individuals or species, particularly when IUPAC encoded sites are selected. When IUPAC encoded sites are used, $M=2$ always results in larger matrices than $M=1$, and the observed differences are higher the lower the missing data threshold (factors of 1.34, 1.45, 1.62 and 2.29 for missing data of 75, 50, 25 and 0% respectively); when fixed positions are used, differences of matrix sizes produced with different M values are lower, but still more pronounced when less missing data are allowed. When building each individual locus, higher values of M will allow higher heterozygosity within loci. Thus, putative loci presenting more than one heterozygous positions that would otherwise be split when $M=1$, cluster together when $M=2$. This increases the probability of alignment of those loci among individuals when building the catalog, and thus, the probability of finding these loci in a higher number of individuals also increases. This added variability information will be represented only when IUPAC encoded sites are used, explaining why the differences between matrix sizes using $M=1$ and $M=2$ are bigger. When all species are considered, $n=8$ results in larger matrices than $n=4$ (by an average factor of 1.3 and 1.7 for fixed and IUPAC encoded sites respectively), whereas almost no difference is observed when using different n values when only *Thunnus* species are considered. The outgroup species are expected to be more divergent with respect to the rest the species. Higher values of n allow recovering more divergent orthologous loci, providing more informative positions resulting from the relatively high genetic distance of the outgroup species. When fixed positions are considered, matrices including all species are always larger than those including only *Thunnus* individuals (by an average factor of 1.6 and 2.8 when using individuals or species respectively); when IUPAC encoded sites are used, including all or only *Thunnus* species results in similar size matrices when high thresholds of missing data are used (75 and 50%), but on smaller

matrices when all species are considered when small missing data thresholds are used (factors of 0.6 and 0.5 when 25 and 0% of missing data are allowed). More positions fixed within but variable between species are found if the outgroup is included, which is expected given that these positions are product of fixation process during speciation, and the greater the evolutionary time, the greater the allele fixation probability. IUPAC encoding of polymorphic sites allows for retention of heterozygous positions within individuals and species, which may be valuable in resolving recent divergences. However, when low percentages of missing data are allowed and the outgroup is included, positions resulting from recent speciation events may not be included in the analysis and thus prevent resolution of close relationships. On the other hand, high percentages of missing data allow positions present in only a subset of individuals to be included in the analysis.

3.3. Phylogeny of the genus *Thunnus*

Taking all analyses into account, a single topology (Figure 2) is supported (all branches with BV>95%) by 90 out of the 128 analyses, whereas no alternative topology is supported at that BV level (Table 1). Only when lowering the BV threshold down to 80%, two new topologies, supported by a few analyses based on the smallest matrices, arose. As for the position of the root, 9 or 22 (for BV>95% or 80% respectively) of the 64 analyses that include all species support the position of the root on the branch of *T. alalunga*, whereas an alternative rooting in the *T. orientalis*/*T. thynnus* branch is supported by 3 analyses only when lowering the BV to 80% (Table 1). As reported by other studies (Cruaud et al., 2014; Hipp et al., 2014; Hou et al., 2015; Jones et al., 2013; Leaché et al., 2015; Rubin et al., 2012; Wagner et al., 2013), the use of IUPAC encoded sites and allowing higher missing data percentages not only produces larger matrices, but also results in better resolved trees (Table 1). Yet, although high bootstrap alone is not evidence of tree accuracy (Rodríguez-Ezpeleta et al., 2007), the congruence obtained among our different analyses is considered a good indicator of true evolutionary signal in the data (Hillis, 1995; Rubin et al., 2012). Additionally, the fact that the trees with the highest support are the ones that have relatively the longest internal and shortest external branches suggests that the increased support is obtained by the use of more common and less independently acquired substitutions among lineages. In sum, our analyses support an alternative evolutionary history of *Thunnus* species that

challenges previous definitions of the *Neothunnus* subgenus (Alvarado Bremer et al., 1997) and sets the root of the *Thunnus* genus within the *T. alalunga* branch. According to our results, *T. obesus* should be included within the *Neothunnus* subgenus. This implies that either this species lost the central heat exchanger and adapted to colder waters independently from the rest of temperate tunas by developing the lateral heat exchanger, as well as visceral and cranial retia mirabilia as it has been reported before for different lineages of Scombridae (Block et al., 1993) or that tropical tunas developed their central heat exchanger *a posteriori* independently from other lineages of Scombridae (Collette et al., 2001).

3.4. Relationships between the Atlantic and Pacific bluefin tunas

All species- and individual-based phylogenetic trees (see Appendix I) strongly support, with bootstrap values (BV) of 100%, the sister group of *T. thynnus* and *T. orientalis*, and contradict previous mtDNA-based studies where Atlantic and Pacific bluefin tunas did not cluster as sister species (Bremer et al., 2005; Chow and Inoue, 1993; Chow and Kishino, 1995; Takeyama et al., 2001; Viñas and Tudela, 2009). Some authors have hypothesized mitochondrial introgression events between different *Thunnus* species as the main reason for discordances between mtDNA and nuclear DNA-based inferred relationships (Chow et al., 2006; Viñas and Tudela, 2009). This is supported by our analyses as the comparative study of the COI sequences revealed two *T. thynnus* individuals with introgressed mitochondrial DNA (Supplementary Figure 1) that had no phylogenetic relationships with *T. alalunga* or *T. orientalis* in the nuclear DNA-based phylogenetic trees. Attending to morphological features (Gibbs Jr and Collette, 1967) and allozyme and mitochondrial molecular markers (Collette et al., 2001), the Atlantic and Pacific bluefin tunas have been considered as two different subspecies; yet, this has been questioned after their ITS1 nuclear marker sequences were found nearly identical (Chow et al., 2006). In our study, most analyses support the respective monophylies of *T. thynnus* and *T. orientalis*, with only some trees based on the smallest matrices not providing support for either of the two monophylies (Table 1). The fact that these analyses do not provide support for any other alternative grouping and that these two species constitute closely related lineages (Chow et al., 2006; Viñas and Tudela, 2009), suggest that finding enough phylogenetic signal to resolve their relationship requires not including distant groups, such as the outgroup, allowing high percentages of missing

data in order to allow markers not present in other taxa, and using positions that are heterozygous within individuals and/or variable among species. Additionally, events such as incomplete lineage sorting, introgression or hybridization may hamper phylogenetic inferences (Hou et al., 2015). Resolving the recent evolutionary history of Pacific and Atlantic Bluefin tuna will require further analyses based on more individuals.

4. Conclusions

Our study corroborates RAD-seq as an efficient tool to gather enough phylogenetically informative nuclear markers to resolve relationships among closely related species. Testing different strategies to select markers to be used for phylogenetic inference, we have observed that, generally, despite including more missing data, bigger matrixes based on variable sites provide higher phylogenetic resolution. Applied to the resolution of the phylogeny of the genus *Thunnus*, our study based on RAD-seq derived genome-wide nuclear markers, redefine the *Neothunnus* group, which should now include *T. obesus*, raising new questions about the evolutionary history and adaptation processes to temperate or tropical waters within the genus *Thunnus*. Additionally, we found markers that support the differentiation of Atlantic and Pacific bluefin tunas, although their classification as two distinct species remains unclear and needs further review. Importantly, the analyses performed and data generated within this study represent a valuable resource for the development of nuclear genetic markers for species identification that, unlike previously developed markers, are valid for both, closely related (*T. thynnus* and *T. orientalis*) or more distantly related species and are not affected by mitochondrial introgression events. Yet, in order to be applicable for fish product traceability and thus assist implementation and enforcement of conservation and management plans these markers should be validated in an alternative set of samples of known origin.

Appendix A. Supplementary Material

Supplementary material associated with this article can be found, in the online version at X.

Acknowledgments

We wish to thank Iñaki Mendibil (AZTI) for excellent technical assistance, Jon Ruiz, Mikel Basterretxea and Iñaki Oyarzabal (AZTI) for collecting samples, and Pete Grewe (CSIRO) for sharing samples. Natalia Díaz-Arce is supported by “Ayudas de formación a jóvenes investigadores y tecnólogos en el entorno científico-tecnológico y empresarial del sector agropesquero y alimentario vasco” (Basque Government). This manuscript is contribution X from the Marine Research Division of AZTI.

Tables

Table 1 Matrix sizes (also shown as proportional grey bars in the “matrix size” column) and supported topologies according to bootstrap value thresholds of 95% (BV>95) and 80% (BV>80) for all the 128 analyses produced from the 8 catalogs, using both fixed and IUPAC encoded positions filtered per individual or per species and allowing different percentages of missing data (represented in black in the “missing” column). Topology 1 stands for the one supported by the majority of the analyses and topologies 2 and 3 for the alternative ones (corresponding to trees 60, 31 and 105 in Appendix I, respectively). When the outgroup is included, the genus was either rooted by *T. alalunga* (ala) or the *T. thynnus* and *T. orientalis* (thyori) clade. The asterisks indicate analyses for which the either monophyly of *T. thynnus* or of *T. orientalis* is not supported.

Figures

Fig. 1 Boxplots depicting median, first and third quartile and standard deviation of number of inferred RAD loci per species when allowing a maximum of 1 (dark grey) or 2 (light grey) mismatches between stacks to create a locus (parameter M).

Fig. 2 Tree obtained from the analysis corresponding to M=2, n=8, IUPAC encoded sites selected per species allowing 75% of missing data that represents topology 1, supported by the majority of the analyses (outgroup has been removed to improve visibility; see Tree 60 in Appendix I). Values above branches indicate bootstrap values and below, percentage of analyses that support that branch at BV > 95 % (above) or BV > 80% (below). The scale bar indicates number of estimated substitutions per site. Images are courtesy of FAO.

References

- Alvarado Bremer, J., Naseri, I., Ely, B., 1997. Orthodox and unorthodox phylogenetic relationships among tunas revealed by the nucleotide sequence analysis of the mitochondrial DNA control region. *Journal of Fish Biology* 50, 540-554.
- Bartlett, S.E., Davidson, W.S., 1991. Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome b genes. *Canadian Journal of Fisheries and Aquatic Sciences* 48, 309-317.
- Block, B.A., Finnerty, J.R., Stewart, A.F., Kidd, J., 1993. Evolution of endothermy in fish: mapping physiological traits on a molecular phylogeny. *SCIENCE-NEW YORK THEN WASHINGTON* 260, 210-210.
- Bremer, J.R.A., Viñas, J., Mejuto, J., Ely, B., Pla, C., 2005. Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Molecular phylogenetics and evolution* 36, 169-187.
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. *Molecular ecology* 22, 3124-3140.
- Collette, B.B., Reeb, C., Block, B.A., 2001. Systematics of the tunas and mackerels (Scombridae). *Fish Physiology* 19.
- Cruaud, A., Gautier, M., Galan, M., Foucaud, J., Sauné, L., Genson, G., Dubois, E., Nidelet, S., Deuve, T., Rasplus, J.-Y., 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. *Molecular biology and evolution* 31, 1272-1274.
- Chow, S., Inoue, S., 1993. Intra- and interspecific restriction fragment length polymorphism in mitochondrial genes of *Thunnus* tuna species. *Bull Natl Res Inst Far Seas Fish* 30, 207-225.
- Chow, S., Kishino, H., 1995. Phylogenetic relationships between tuna species of the genus *Thunnus* (Scombridae: Teleostei): inconsistent implications from morphology, nuclear and mitochondrial genomes. *Journal of Molecular Evolution* 41, 741-748.
- Chow, S., Nakagawa, T., Suzuki, N., Takeyama, H., Matsunaga, T., 2006. Phylogenetic relationships among *Thunnus* species inferred from rDNA ITS1 sequence. *Journal of Fish Biology* 68, 24-35.
- Chow, S., Nohara, K., Tanabe, T., Itoh, T., Tsuji, S., Nishikawa, Y., Uyeyanagi, S., Uchikawa, K., 2003. Genetic and morphological identification of larval and small juvenile tunas (Pisces: Scombridae) caught by a mid-water trawl in the western Pacific. *BULLETIN-FISHERIES RESEARCH AGENCY JAPAN*, 1-14.
- Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., Cresko, W.A., 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in molecular biology (Clifton, N.J.)* 772, 157-178.
- Gibbs Jr, R.H., Collette, B.B., 1967. Comparative anatomy and systematics of the tunas, genus *Thunnus*.
- Herrera, S., Shank, T.M., 2015. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *bioRxiv*, 019745.
- Hillis, D.M., 1995. Approaches for assessing phylogenetic accuracy. *Systematic Biology* 44, 3-16.
- Hipp, A.L., Eaton, D.A., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. *PLOS one* 9, e93975.

- Hou, Y., Nowak, M.D., Mirré, V., Bjorå, C.S., Brochmann, C., Popp, M., 2015. Thousands of RAD-seq Loci Fully Resolve the Phylogeny of the Highly Disjunct Arctic-Alpine Genus *Diapensia* (Diapensiaceae). *PLOS one* 10, e0140175.
- Jones, J.C., Fan, S., Franchini, P., Scharl, M., Meyer, A., 2013. The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular ecology* 22, 2986-3001.
- Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D., Linkem, C.W., 2015. Phylogenomics of Phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome biology and evolution* 7, 706-719.
- Miya, M., Friedman, M., Satoh, T.P., Takeshima, H., Sado, T., Iwasaki, W., Yamanoue, Y., Nakatani, M., Mabuchi, K., Inoue, J.G., 2013. Evolutionary origin of the Scombridae (tunas and mackerels): members of a paleogene adaptive radiation with 14 other pelagic fish families. *PLOS one* 8, e73535.
- Orrell, T.M., Collette, B.B., Johnson, G.D., 2006. Molecular data support separate scombroid and xiphioid clades. *Bulletin of Marine Science* 79, 505-519.
- Pardo, M.A., Pérez-Villareal, B., 2004. Identification of commercial canned tuna species by restriction site analysis of mitochondrial DNA products obtained by nested primer PCR. *Food Chemistry* 86, 143-150.
- Pla, C., Hattour, A., Tawil, M., Farrugia, A., de la Serna, J., Viñas, J., 2003. Mitochondrial Genetic Characterization Of Bluefin Tuna (*Thunnus Thynnus*) From Three Mediterranean (Libya, Malta, Tunisia); And One Atlantic Locations (Gulf Of Cadiz).
- Ree, R.H., Hipp, A.L., 2015. Inferring phylogenetic history from restriction site associated DNA (RADseq).
- Rodríguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., Philippe, H., 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Systematic Biology* 56, 389-399.
- Rubin, B.E., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. *PLOS one* 7, e33394.
- Santini, F., Carnevale, G., Sorenson, L., 2013. First molecular scombrid timetree (Percomorpha: Scombridae) shows recent radiation of tunas following invasion of pelagic habitat. *Italian Journal of Zoology* 80, 210-221.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.
- Takeyama, H., Chow, S., Tsuzuki, H., Matunaga, T., 2001. Mitochondrial DNA sequence variation within and between tuna *Thunnus* species and its application to species identification. *Journal of Fish Biology* 58, 1646-1657.
- Tseng, M.-C., Jean, C.-T., Smith, P.J., Hung, Y.-H., 2012. Interspecific and intraspecific genetic diversity of *Thunnus* species. INTECH Open Access Publisher.
- Viñas, J., Tudela, S., 2009. A validated methodology for genetic identification of tuna species (genus *Thunnus*). *PLOS one* 4, e7606.
- Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular ecology* 22, 787-798.

RAD loci per specie

					only <i>Thunnus</i>			all species				
						TOPOLOGY			TOPOLOGY		ROOT	
M	n	positions	sp/ind	missing	matrix size	BV>95	BV>80	matrix size	BV>95	BV>80	BV>95	BV>80
1	4	IUPAC	ind		279084	1	1	303886	1	1	-	ala
					218133	1	1	223175	1*	1	-	-
					151770	1	1	98594	1*	1	-	-
					25107	1*	1*	10957	-*	-*	-	-
			sps		146462	1	1	173859	1	1	-	ala
					114492	1	1	110185	1	1	-	ala
					75737	1	1	37361	1	1	-	-
					25112	-	-	10957	-	-	-	-
		fixed	ind		106321	1	1	125843	1	1	-	-
					64379	1	1	75907	1*	1*	-	-
					32334	1*	1	36210	1*	1*	-	thyori
					3435	-*	-*	5006	-*	-*	-	-
			sps		19618	1	1	46740	1	1	ala	ala
					14132	1	1	31424	-	-	-	-
					8480	1	1	16344	-	-	-	-
					2653	-	-	4793	-	-	-	-
	8	IUPAC	ind		297574	1	1	370270	1	1	-	ala
					228350	1	1	274984	1*	1*	-	ala
					154610	1	1	146512	1*	1	ala	ala
					24079	1*	1*	14483	-*	-*	-	-
			sps		147745	1	1	208048	1	1	-	-
					114556	1	1	139795	1	1	-	ala
					75067	-	-	57151	1	1	-	-
					24083	-	-	14483	-	-	-	-
		fixed	ind		127019	1	1	191582	1	1	-	-
					75110	1	1	124622	1*	1*	-	-
					35738	1*	1	71968	1*	1	-	-
					3404	-*	-*	7508	-*	2*	-	-
			sps		21759	1	1	78371	1	1	ala	ala
					15254	1	1	56389	-	1	-	-
					8944	-	1	29274	-	2	-	-
					2625	-	-	7230	-	-	-	-
2	4	IUPAC	ind		328966	1	1	357395	1	1	-	-
					271503	1	1	283198	1	1	-	-
					214363	1	1	153249	1*	1	-	-
					64492	1*	1	20808	-*	-*	-	-
			sps		223772	1	1	263424	1	1	-	ala
					190082	1	1	185416	1	1	ala	ala
					144305	1	1	58529	1	1	-	ala
					64497	-	1	20808	-	-	-	-
		fixed	ind		83906	1	1	104393	1	1	-	-
					51918	1	1	68202	1*	1*	-	-
					30471	1*	1*	39151	1*	1*	-	thyori
					6478	-*	3*	8142	-*	2*	-	-
			sps		18482	1	1	50093	1	1	ala	ala
					14705	1	1	35243	-	1	-	-
					10350	-	-	21090	-	-	-	-
					4441	-	3	7694	-	-	-	-
	8	IUPAC	ind		346583	1	1	426052	1	1	-	ala
					283161	1	1	339603	1*	1	-	ala
					218996	1	1	207223	1*	1	ala	ala
					61359	1*	1	30851	-*	-*	-	ala
			sps		224582	1	1	305538	1	1	ala	ala
					189662	1	1	224042	1	1	-	ala
					143035	1	1	98675	1	1	ala	ala
					61363	1	1	30851	-	-	-	-
		fixed	ind		103893	1	1	168757	1	1	-	-
					63489	1	1	117399	1*	1	-	-
					35285	1*	1*	77852	1*	1*	-	thyori
					6444	-*	-*	13950	-*	2*	-	-
			sps		21092	1	1	87287	1	1	ala	ala
					16230	1	1	66299	1	1	-	ala
					11139	-	-	42688	-	2	-	-
					4438	-	-	13304	-	-	-	-

Highlights

- RAD-seq derived data successfully resolved phylogenetic relationships within the genus *Thunnus*.
- Bigeye tuna is included in the lineage of tropical tunas, challenging previous definitions of the *Neothunnus* subgenus.
- The theory of a mitochondrial introgression event not involving nuclear genome between some tuna species is reinforced.
- Atlantic and Pacific Bluefin tunas are genetically differentiated, but deciphering their recent evolutionary history requires more data.