Marine Genomics 25 (2016) 43–48

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*)

Carlo Pecoraro [a,b,*,1], Massimiliano Babbucci [c,1], Adriana Villamor [a], Rafaella Franch [c], Chiara Papetti [k], Bruno Leroy [d], Sofia Ortega-Garcia [e], Jeff Muir [f], Jay Rooker [g], Freddy Arocha [h], Hilario Murua [i], Iker Zudaire [j,b], Emmanuel Chassot [b], Nathalie Bodin [b], Fausto Tinti [a], Luca Bargelloni [c], Alessia Cariani [a]

[a] Dept. Biological, Geological and Environmental Sciences (BIGEA), University of Bologna, Via Selmi 3, 40126 Bologna, Italy
[b] Institut de Recherche pour le Développement (IRD), UMR MARBEC (IRD/Ifremer/UM2/CNRS) SFA, Fishing Port, BP570 Victoria, Seychelles
[c] Comparative Biomedicine and Food Science, University of Padova, viale dell'Università 16, 35020 Legnaro, PD, Italy
[d] Secretariat of the Pacific Community, Oceanic Fishery Programme, BP D5, 98848 Noumea, New Caledonia
[e] Departamento de Pesquerias, Instituto Politécnico Nacional-CICIMAR, Avenida IPN s/n, La Paz, BCS, Mexico
[f] Pelagic Fisheries Research Program, University of Hawaii, Marine Science Building 312, Honolulu, HI 96822, USA
[g] Department of Marine Biology, Texas A&M University, 1001 Texas Clipper Road, Galveston, TX 77553, USA
[h] Instituto Oceanográfico de Venezuela, Universidad de Oriente, Avda. Universidad Cerro Colorado, Cumana 6101, Venezuela
[i] Marine Research Division, AZTI, Herrera Kaia-Portualdea z/g, Pasaia, 20110, Gipuzkoa, Spain
[j] Ikerbasque Fundazioa, Maria Diaz de Haro, 3-6°, Bilbao, 48013, Bizkaia, Spain
[k] Section of Integrative Ecophysiology, Alfred-Wegener-Institute for Polar and Marine Research, Am Handelshafen 12, Bremerhaven 27570, Germany

ARTICLE INFO

Article history:
Received 8 September 2015
Received in revised form 1 December 2015
Accepted 8 December 2015
Available online 23 December 2015

Keywords:
Marine fish
Population genomics
RAD sequencing
SNP
Tropical tuna
Tuna fishery

ABSTRACT

Global population genetic structure of yellowfin tuna (*Thunnus albacares*) is still poorly understood despite its relevance for the tuna fishery industry. Low levels of genetic differentiation among oceans speak in favour of the existence of a single panmictic population worldwide of this highly migratory fish. However, recent studies indicated genetic structuring at a much smaller geographic scales than previously considered, pointing out that YFT population genetic structure has not been properly assessed so far. In this study, we demonstrated for the first time, the utility of 2b-RAD genotyping technique for investigating population genetic diversity and differentiation in high gene-flow species. Running de novo pipeline in *Stacks*, a total of 6772 high-quality genome-wide SNPs were identified across Atlantic, Indian and Pacific population samples representing all major distribution areas. Preliminary analyses showed shallow but significant population structure among oceans ($F_{ST} = 0.0273$; P-value < 0.01). Discriminant Analysis of Principal Components endorsed the presence of genetically discrete yellowfin tuna populations among three oceanic pools. Although such evidence needs to be corroborated by increasing sample size, these results showed the efficiency of this genotyping technique in assessing genetic divergence in a marine fish with high dispersal potential.

1. Introduction

Yellowfin tuna (*Thunnus albacares*, YFT) has relevant biological and economic importance at the global scale, being an apex predator in oceanic ecosystem and representing the second largest tuna fishery worldwide (FIGIS, 2010–2015). Currently, YFT is managed in four distinct stocks under the jurisdiction of four independent Regional Fisheries Management Organizations (RFMOs). Although a proper fish stock management needs accurate knowledge on the stock structure and its genetic variation with respect to environmental and ecological conditions (Papetti et al., 2013), YFT genetic population structure has not been resolved yet. Different studies provided discordant patterns of YFT global-scale genetic differentiation (Ward et al., 1997; Ely et al., 2005; Appleyard et al., 2001), together with a genetic structuring detected at the regional level (Dammannagoda et al., 2008; Kunal et al., 2013; Li et al., 2015). This discordance was likely due to the YFT life history traits (e.g. high fecundity, large population sizes), which make detecting patterns of genetic differentiation among population samples very difficult (Ely et al., 2005; Juan-Jordá et al., 2013). Moreover,

* Corresponding author at: Dept. Biological, Geological and Environmental Sciences (BIGEA), University of Bologna, Via Selmi 3, 40126 Bologna, Italy.
E-mail addresses: carlo.pecoraro2@unibo.it (C. Pecoraro), massimiliano.babbucci@unipd.it (M. Babbucci), adriana.villamor@unibo.it (A. Villamor), rafaella.franch@unipd.it (R. Franch), chiara.papetti@awi.de (C. Papetti), brunol@spc.int (B. Leroy), sortega@ipn.mx (S. Ortega-Garcia), jmuir@hawaii.edu (J. Muir), rookerj@tamug.edu (J. Rooker), farocha@udo.edu.ve (F. Arocha), hmurua@azti.es (H. Murua), emmanuel.chassot@ird.fr (E. Chassot), nathalie.bodin@ird.fr (N. Bodin), fausto.tinti@unibo.it (F. Tinti), luca.bargelloni@unipd.it (L. Bargelloni), alessia.cariani@unibo.it (A. Cariani).
[1] These two authors contributed equally to this work.

http://dx.doi.org/10.1016/j.margen.2015.12.002
1874-7787/

population genetic studies reporting significant differences relied upon a relatively small number of molecular markers, hence, covering only a very limited portion of the genome (Appleyard et al., 2001; Díaz-Jaimes and Uribe-Alcocer, 2006). Failing to detect population structure, due to limited genetic resolution of classical markers, can potentially be misleading for management purposes, driving to local overfishing and severe stock decline (Ying et al., 2011).

According to the uncertainty about both population structure and size of YFT stocks, there is an evident need for developing alternative approaches based on genomics, that allow screening a larger number of markers across the entire genome, including neutral and non-neutral loci. This might enable detecting YFT population structure, quantifying the extent of spatial demographic changes and discover imprints of local adaptation, which represent priority focus for implementing any effective management plan.

The rapid advent of next-generation sequencing (NGS)-based genotyping methods has significantly improved our ability to analyse thousands of Single Nucleotide Polymorphism (SNP) markers across the entire genome, increasing the precision in detecting small genetic differentiation among geographical populations (Waples et al., 2008; Allendorf et al., 2010; Davey et al., 2011; Narum et al., 2013; Andrews and Luikart, 2014). Although SNPs are characterized by a low diversity due to the only four possible allelic states, this limitation is largely outweighed by their abundance, being as frequent as one SNP every few hundred base pairs (Morin et al., 2004, 2009). Moreover, SNPs are becoming the marker of choice for many applications in population ecology, evolution and conservation genetics, having a high potential for genotyping efficiency, data quality and low-scoring error rates, genome-wide coverage and analytical simplicity (Milano et al., 2014).

Here, for the first time, we applied the 2b-RAD Genotyping-By-Sequencing (GBS) technique (Wang et al., 2012) for testing its potential for investigating population genetic structure in a non-model, large pelagic and highly migratory fish species. This novel genomic tool is based on sequencing reduced representation libraries produced by type IIB restriction endonucleases, which cleave genomic DNA upstream and downstream of their target site, generating tags of uniform length that are ideally suited for sequencing on existing NGS platforms (Wang et al., 2012). This method permits parallel and multiplexed sample sequencing of tag libraries for the rapid discovery of thousands of SNPs across the entire individual's genome, with a very cost-effective procedure resulting in high genome coverage. The 2b-RAD method allows screening in parallel almost every restriction site in the genome, whereas other GBS methods can only target a subset of total restriction sites to counterbalance loss of PCR amplification and sequencing efficiency due to large size of restriction fragments. This technique also allows fine-tuning the marker density by means of selective adapters in order to sequence fewer loci with higher coverage, for applications such as population genetics (Puritz et al., 2014; Andrews and Luikart, 2014). Given these attributes, the 2b-RAD method has the potential to discriminate the existence of genetic differentiation with a high statistical power, generating genome-wide data for genetic structure analysis at different spatial scales for YFT populations.

In this study, we: i) first examine the utility of Technical Replicates (TRs) for optimizing genotyping procedure, comparing the results obtained running the *denovo_map.pl* and the *ref_map.pl* programs in *Stacks* (Catchen et al., 2011, 2013); and ii) finally assess the applicability of 2b-RAD for future investigations in this highly migratory species.

2. Results and discussion

A similar number of reads was obtained among TRs, before and after quality filtering (Table 1), which underlines the reliability of this technique in genotyping individuals.

Among the different *Stacks* settings considered, the *-m* value was the parameter that most affected the genotyping results, in particular the number of detected SNPs. Sensitivity tests performed on the TRs showed a decrease in the number of SNPs, from 5753 to 4490, when increasing *-m* from 5 to 15 (Fig. 1 and Supplementary Material 1 for values with associated Standard Error). The percentage of error rate varied approximately from 1% to 5%, with a decreasing trend when increasing the *-m* values (Fig. 1 and Supplementary Material 1). The percentage of heterozygous SNPs remained constant with increasing *-m* values (Fig. 1 and Supplementary Material 1).

An increase in true heterozygous SNPs calls was observed using the *bounded SNP calling model* compared to the *default SNP model* and reducing the upper bound values, in agreement with the results obtained by Mastretta-Yanes et al. (2015). In fact, reducing the upper bound on the maximum-likelihood of ε decreases the possibility of calling a homozygote instead of a true heterozygous genotype (Catchen et al., 2013). The proper genotype calling was further checked for a subsample of the total reads obtained, in the *Stacks* web interface, verifying the sequences alignment and monitoring the genotyping inference when the results were exported. This procedure was repeated each time when changing the different model's *upper bound* values.

By relaxing the number of mismatches within each locus (*-n*) and among loci (*-M*), an increase in the number of SNPs and error rate was observed (Supplementary Material 2).

Mapping 2b-RAD reads against the genome of *Thunnus orientalis* allowed a high percentage of successfully mapped sequences (86.59%). The outputs obtained on the mapped data from TRs with the *ref_map.pl* program, confirmed the trends observed with the *denovo_map.pl* program (Fig. 1). However, the absolute number of SNPs was lower than that obtained with the *denovo_map.pl* program, likely due to the incompleteness of the reference genome used (the only *Thunnus* sp. genome available to date, Nakamura et al., 2013) and the phylogenetic distance between YFT and *T. orientalis*.

Aligning reads to the reference genome, before calling a locus, can filter out erroneous stacks generated by contaminants (e.g. bacteria) possibly present in very small amount in the starting gDNA sample. Moreover the error rate also showed a less evident decreasing pattern when increasing *-m*, confirming however a low error rate in the genotyping call (<5%). On the contrary, the percentage of heterozygous SNPs identified using *T. orientalis* genome as reference, showed a slight increase from 35.6% to 40.7%, when higher values of *m* were used (Fig. 1).

Table 1
Details on the technical replicates: acronym (Sample ID), Oceanic origin, genomic DNA concentration in ng/μL, library concentration in nm/μL, number of raw reads obtained, retained reads after quality filtering, and their corresponding percentage.

Sample ID	Oceanic origin	gDNA ng/μL	Library nm/μL	No. raw reads	No. filtered reads	% retained reads
34_2_Y_2R1	Atlantic Ocean	333.80	185.38	2,276,239	1,772,927	78%
34_2_Y_2R2	Atlantic Ocean	333.80	207.83	2,672,917	1,914,181	72%
34_2_Y_2R3	Atlantic Ocean	333.80	197.03	2,309,039	1,805,181	78%
77_2_Y_15R1	Pacific Ocean	218.02	238.81	2,212,559	1,788,988	81%
77_2_Y_15R2	Pacific Ocean	218.02	240.53	2,292,834	1,850,871	81%
77_2_Y_15R3	Pacific Ocean	218.02	208.91	2,085,658	1,802,820	86%
51_1_Y_7R1	Indian Ocean	177.54	166.05	2,330,292	1,767,345	76%
51_1_Y_7R2	Indian Ocean	177.54	170.81	2,300,342	1,824,635	79%