Population structure and genetic management unit delineation in the bluefin tuna using a genotyping-by-sequencing approach.

Gregory N. Puncher (1,2), Alessia Cariani (1), Gregory E. Maes (3), Jeroen Van Houdt (4,5), Koen Herten (4,5), Aitor Albaina (6), Rita Cannas (7), Naiara Rodríguez-Ezpeleta (8), Haritz Arrizabalaga (8), Piero Addis (7), Angelo Cau (7), Nicloas Goñi (8), Igaratza Fraile (8), Urtzi Laconcha Santamaria (8), Fausto Tinti (1)

(1) Dept. of Biological, Geological and Environmental Sciences / Laboratory of Genetics and Genomics of Marine Resources and Environment (GenoDREAM), University of Bologna, Ravenna, Italy; (2) Biology Dept., Research Group Marine Biology, Ghent University, Ghent, Belgium; (3) Centre for Sustainable Tropical Fisheries and Aquaculture, Comparative Genomics Centre, James Cook University, Townsville, Australia; (4) Center for Human Genetics, KU Leuven, Leuven, Belgium; (5) Genomics Core, UZ Leuven, Leuven, Belgium; (6) Laboratory of Genetics Faculty of Science & Technology, Dept. of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Leioa, Spain; (7) Dept. of Life and Environmental Sciences, University of Cagliari, Cagliari, Italy; (8) AZTI Tecnalia, Marine Research Division, Pasaia, Gipuzkoa, Spain. Presenter contact details: greg@respectnature.com, Phone +39 3470424350

Summary
A genotyping-by-sequencing approach was used to detect single nucleotide polymorphisms (SNPs) throughout the Atlantic bluefin tuna (BFT; *Thunnus thynnus*) genome while simultaneously analyzing the species’ population structuring using 723 individuals captured between 2007-2013 from fourteen locations throughout the species’ range. Due to the lack of an existing reference genome and sub-optimal performance of the restriction enzyme used (*Ape*KI), a novel data analysis approach was developed. By pooling the data from all individuals within geographic sampling locations, allele read count frequencies were calculated and compared. From a pool of 184,895 candidate SNPs, a high performance 96 SNP genotyping panel was developed *in silico* by selecting loci with pronounced differences in read count frequencies among young fish (larvae and young-of-the-year). Genotyping results revealed significant differentiation of samples from the Gulf of Mexico and Mediterranean (*F*<sub>ST</sub> = 0.013). The composition of five mixed feeding aggregations in the Atlantic Ocean was determined via population assignment analysis of 230 adult fish. Our results provide evidence of persistent population structuring across broad geographic areas and can be used to estimate mixing rates, and with further development, may lead to traceability tools.

Introduction
SNPs have been used in several fisheries in recent years for determination of population structuring, traceability, hybridization rates and migratory dynamics (Nielsen et al. 2012; Albaina et al. 2013; Houston et al. 2014; Larson et al. 2014). The motivation to search for discriminatory SNPs within the BFT genome is a product of unresolved and often conflicting results from research that has used other molecular markers (ICCAT 2013). The ease by which data can be shared between research groups and the fact that SNPs can be sampled from throughout the entire genome, including genes influenced by selective pressures are additional benefits to this approach. Herein we describe efforts to discover SNPs in the *T. thynnus* genome, develop a SNP genotyping panel and investigate the species’ genetic population structure using this new tool.

Materials and Methods
Bluefin tuna (BFT) larvae (n=105) and young-of-the-year (YOY; n=450) were captured in the Mediterranean Sea, Gulf of Mexico and Cape Hatteras from 2007 to 2012 during the late spring and early summer months. Genomic DNA was extracted, purified, digested with the *Ape*KI restriction enzyme and GBS libraries were prepared following Elshire et al. (2011). Libraries were then sequenced on the Illumina HiSeq2500. The sequencing reads
were trimmed and aligned with a new genomic reference assembled using data generated by genome sequencing of a single individual captured offshore from the Balearic Islands. SNP calling was performed and after quality control filtering of the data, a total of 184,895 SNPs were found to be sufficiently represented among all samples. From this large number of candidate SNPs, 384 were selected for a genotyping panel based on pairwise comparisons between differences in allele frequencies among all samples. The results were used to develop a reduced 95 SNP panel which was then used to genotype 556 individuals from various age classes (99 larvae, 227 YOY and 230 medium/large adults).

Pairwise comparisons of young tuna samples showed consistent differentiation between Gulf of Mexico and Mediterranean Sea samples. No clear pattern of structuring was detected among Mediterranean samples using Fst values. Young and adult tunas were assigned to spawning areas of origin, considering as reference for assignment, individuals divided in two broad classes: Western Atlantic and Mediterranean. A minimum threshold of 70% for the scores of assignment was implemented, resulting in 80.3% of young tuna and 79.1% of adults assigned to one of the two regions of origin. The proportion of individuals assigned to the Western Atlantic (16.0%, average score = 87%) and Mediterranean Sea (64.3%, average score = 91.6%) was determined.

Discussion
Our SNP panel provides novel insights into the population structuring and spatial dynamics of Atlantic Bluefin Tuna. The proportion of individuals assigned with high assignment scores to probable spawning areas of origin was high, serving as an indication of the high performance of the selected SNPs. In order to improve the analysis of assignment to Western Atlantic and Mediterranean spawning areas, additional young tuna samples are required from the Gulf of Mexico and adult samples are required from the eastern Mediterranean. With further development and increased sampling, our capacity to trace individuals back to a Western or Eastern origin is expected to improve. The results from the analysis of young tunas provide further evidence of persistent population structuring across broad geographic areas, can be used to estimate mixing rates and with further develop, may lead to traceability tools.

References