DATA RECOVERY PLAN Priority 2:

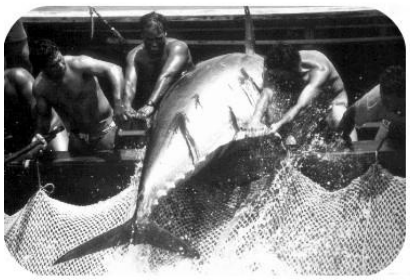
Historical genetic samples collected in old times in the Eastern Mediterranean Sea, in the Marmara Sea or in the Black Sea, including the genetic analyses of these samples"

Contract ICCAT GBYP 06-2013

Final Report

for

ICCAT



Task Leaders:

Fausto Tinti[1] and Gregory Neils Puncher[1]

Participants:

Gregory Neils Puncher[1], Francesco Massari[1], Elisabetta Cilli[2], Alessia Cariani[1], Fausto Tinti[1]

1. Laboratory of Genetics & Genomics of Marine Resources and Environment, Dept. Biological, Geological & Environmental Sciences, University of Bologna, Ravenna, Italy
2. Dept. of Cultural Heritage, University of Bologna, Ravenna, Italy

1. Overview

Mediterranean civilizations and the Atlantic bluefin tuna (*Thunnus thynnus*) have shared a dramatic history spanning millennia. Today, traditional fishing techniques are fading into the past and industrial fisheries threaten the survival of the world's largest tuna species. By analyzing the DNA contained in the bones of tuna recovered from archeological excavations throughout Europe, we hope to reconstruct the species' adaptive responses to fisheries pressure and climate change. Here we describe our efforts to reassemble part of the bluefin's history through ongoing research funded by ICCAT's GBYP Phase 4 Data Recovery Program. During the funded period we have successfully developed a new DNA extraction protocol for historical samples, conducted a comparison of DNA yields from different time periods and environments and designed a 96 SNP genotyping panel containing loci associated with selective traits and population structuring. Some two hundred historical and ancient samples will be dispatched for genotyping in the second half of January 2015 and the final deliverable products will be completed before March 2015, thereby fulfilling all goals proposed in the grant application.

2. Sample collection

2.1. Massimo Sella Archive

The Massimo Sella archive is a collection of remains (vertebrae, skulls and fins) belonging to some 5000 fish captured in the Mediterranean Sea during the early decades of the 20th century by Italian professor Massimo Sella. Originally housed at the Adriatic Marine Biology Institute in Rovigno, Croatia, the collection was moved to a University of Bologna facility in Fano, Italy during the Second World War. The vast majority of bones in the collection are Atlantic bluefin tuna (BFT) caught in tuna traps in Istria (Croatia), Messina (Italy) and Sliten (Libya). Several years ago geneticists from Italy used bones from this collection to investigate the temporal dynamics of population structuring using a host of microsatellites (Riccioni et al. 2010). In that study, they found weak signatures of population differentiation among modern ($F_{ST} = 0.014$, $P < 0.0001$) and two historical samples ($F_{ST} = 0.020$, $P < 0.0001$); although there remained no significant evidence of genetic

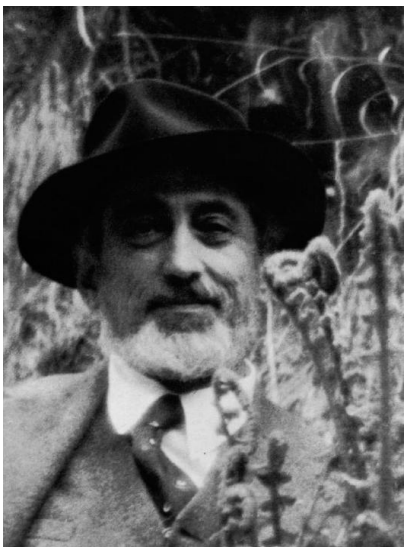


Fig. 1: Prof. Massimo Sella 1886-1959

structuring when Bayesian tools were employed. Taking these results into consideration, researchers were left with the following conclusions: 1) Data derived from the historical Libyan samples suggest that spatiotemporal shifts in BFT population structure and dynamics have occurred in the Mediterranean, 2) there is a weak signal of genetic structuring within the Mediterranean Sea, 3) the genetic markers used to date are likely to weak to demonstrate these genetic dynamics and, 4) a more robust sampling design using advanced techniques and higher performing genetic markers are required to address these questions (Cannas et al. 2012). For these reasons we have continued with this important work and included 150 samples from the Massimo Sella archive from the same locations used by Riccioni et al. (2010). DNA has been extracted from all of these samples and information concerning the capture site and date, weight and length of animal and sex has been recorded for each sample, whenever the data is available.

2.2. Istanbul, Turkey

Archeological excavations conducted between 2004 and 2013 in the Yenikapi neighbourhood of Istanbul have revealed the location of a Byzantine era harbour. Built by Emperor Theodosius in the 4th century CE, it was one of the largest and most important ports in the Roman Empire for nearly eight centuries. Many animal remains have been uncovered, including 150 vertebrae of Atlantic bluefin tuna, *Thunnus thynnus*, as well as a multitude of swordfish rostra, *Xiphias gladius*. All vertebrae are very well preserved and suitable for both morphological and molecular investigations. Butchery marks (Fig. 3) and historical accounts of the Ancient Greek and Roman tuna fishery (Oppian 177BCE; Aristotle 350BCE) in the region suggest that the vertebrae belonged to tunas caught in the immediate vicinity (Marmara or Black Sea). After the meat was removed from the bone, it was likely sold fresh to a local buyer or preserved in salt or processed into garum (fermented fish sauce) and traded abroad. Since a viable population of BFT has been absent from the Black Sea since the mid 1980s (Karakulak and Oray 2009), these vertebrae are of particular interest for the scientific community. The cause of their disappearance from the Black Sea remains unknown; however,

Fig. 2: Marmaray project archeological excavations at Yenikapi (Buket et al. 2009).

hypotheses abound, including eutrophication, overfishing, noise pollution, lack of prey and changing environmental conditions (Papaconstantinou and Farrugio 2000; Daskalov 2002; Sara et al. 2007; MacKenzie and Mariani 2012).

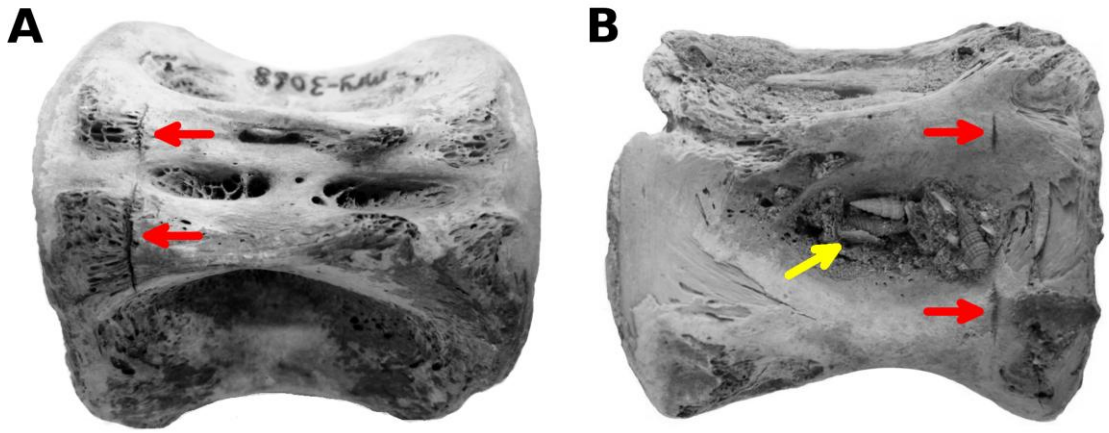


Fig. 3: *Thunnus thynnus* vertebrae recovered from excavations at Yenikapi, Istanbul, Turkey. Butchery marks (red arrows) and gastropod shells (yellow arrow) are indicated.

There is a strong possibility that the BFT of the Black Sea were spawning in isolation, since their migratory movements were likely determined by the whereabouts of their prey which migrated into the warming waters of the Black Sea in the spring, returning to the warmer waters of the Mediterranean Sea in autumn. Aristotle recorded this migratory behaviour in 350 BCE and his observations have been verified by 20th century marine scientists (Akyuz and Artuz 1957; Mather et al. 1995, Karakulak and Oray 2009). This would place the local population in the Black Sea during the spawning season, far from all other spawning groups. If spawning was taking place in the Black Sea, then adaptations to the local hydrographic conditions would have been required. During the spawning season, salinity, density and temperature are lower in the Black Sea than all other BFT spawning areas (Mackenzie and Mariani 2012). Unless physical adaptations provided increased egg buoyancy (increased size, decreased dry weight and thinner chorions with fewer lamellae), the eggs of BFT would quickly sink into the Black Sea's hypoxic waters below 100m (Mackenzie and Mariani 2012). Additional adaptations would have been required of the sperm, unfertilized eggs and developing embryo and larvae. The genetic programming for these adaptations could be recorded in the bones uncovered at Yenikapi and might still be traced to the remaining descendants of this isolated tribe, if any have survived. Information concerning the distribution of the remaining Black Sea BFT can provide information about population structuring, migratory

behaviour, genetic diversity and adaptive potential. By comparing the genetic code of modern and ancient BFT we can shed light on the evolution of the species genome in response to nearly two millennia of fisheries pressure, a changing climate and pollution of the sea. The rapid inundation of the harbour with fine silts has preserved all vertebrae in excellent condition (Fig. 3). Growth rings can be easily counted and growth rates estimated with ease. To date, this magnificent collection of BFT bones at the University of Istanbul haven't been utilized for any research purposes (Fig. 4). Caretaker of this overwhelming collection of bones is the Director of animal remains from the Marmaray excavation project, Professor Vedat Onar, of the Osteoarchaeology Laboratory, Department of Anatomy, Faculty of Veterinary Medicine at Istanbul University. We were fortunate enough to acquire 5 samples from this collection for preliminary analyses of DNA quality; however, additional access to these bones will require a contract with the Turkish Ministry of Culture and Istanbul University.

Figure 4: Bluefin tuna vertebrae unearthed at Yenikapi in the archive (left) or on display at the University of Istanbul.

The following SCRS document (submitted to ICCAT in September 2014) describes the historical context of this collection and discusses its scientific significance in further detail:

GN Puncher, V Onar, NY Toker, F Tinti (2014) A multitude of Byzantine era bluefin tuna and swordfish bones uncovered in Istanbul, Turkey. **ICCAT SCRS/2014/167.**

2.3. Iberian Peninsula

In the archives of the Faculty of Archeozoology at the Autonomous University of Madrid, Professor Arturo Morales watches over a vast collection of animal bones collected from archeological sites around the world (Fig. 5). From this collection, we were able to acquire 30 DNA genetic samples

from “giant” tuna (~400-500kg) vertebrae captured during the 1985 *almadraba* (annual tuna slaughter) at Barbate (Spain, n=10), vertebrae from medium sized adults (150-200kg) captured by Late Iron Age and republican roman colonists (2^{nd}-1^{st} century B.C.) in Tavira (Portugal, n=10) and Baelo Claudia (Spain, n=10). The samples collected from Tavira had been used by home builders as building materials and were incorporated into the walls of the buildings within the small Iberian roman colony. This was done in an effort to provide an organic matrix that strengthens instead of weakens architectural structures. Conversely, the bones from Baelo Claudia seemed to have been buried in a refuse pile or midden.

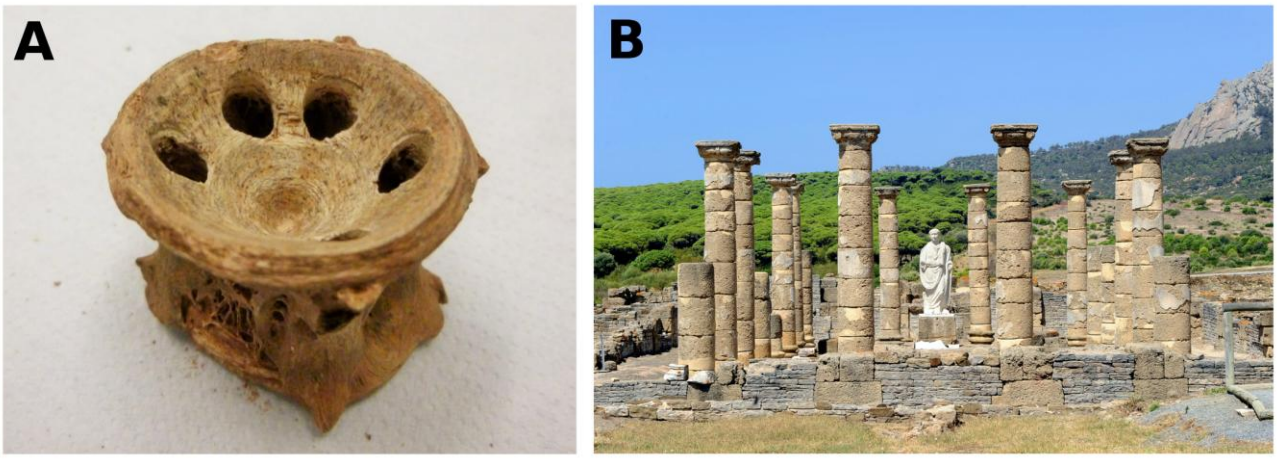


Fig. 5: A) BFT vertebrae from late iron age settlement at Castro Marim with drill holes after powder removal. **B)** Archeological site of ancient Roman fishing town of Baelo Claudia.

Preliminary plans discussed in the project proposal suggested that samples from a fish factory destroyed during the 1755 earthquake and tsunami would be analyzed; however, we were only able to acquire 2 of these samples. We hope that by including a second sample site from the 2^{nd} century B.C., we will more than make up for this absent stratum.

3. DNA extractions

*S*ample preparation (bone surface decontamination, drilling and powder collection) was conducted at the location of each collection, using a portable sampling set and UV lamp. Enzymatic and chemical extraction of DNA from all samples occurred in Ravenna (Italy) at the Ancient DNA Laboratories at the University of Bologna's Environmental Science campus and Department of Cultural Heritage.

3.1 Sample preparation

All bones were sprayed with 1.5-2% sodium hypochlorite (diluted commercial bleach) and left to soak for up to ten minutes, following published protocols for sterilization of bone surfaces (Kemp and Smith 2005). Some of the very old samples were much more porous and absorbent and were therefore exposed

to less bleach for a shorter duration of time. Following this treatment, bones were rinsed with distilled water, wiped clean with paper towel, rinsed and air-dried. Once dry, a thin surface layer was removed using sand paper and the surface was once again rinsed with bleach and distilled water. All bones were left to dry for a minimum of 15 minutes under UV light (254nm wave length; Fig 6) and holes were drilled into each bone to simultaneously gain access to the internal matrix and produce bone powder. For bones that were too small for drilling, surfaces were diligently cleaned and a hand saw was used to bisect the bone, at which time the internal matrix of the bone was scraped free and pulverized.

Fig. 6: A cleaned vertebra prepared for drilling and bone powder pulverized with a mortar pestle.

3.2 Chemical/Enzymatic DNA extraction

Various protocols for the chemical extraction of DNA from bones were developed and compared in order to maximize DNA yields. Through a collaboration with the Department of Cultural heritage at the University of Bologna, which hosts specialists in the extraction of DNA from ancient humans, we have developed new protocols that yield higher concentrations of DNA from archeological remains than previously used methods (Riccioni et al. 2010). Based on a previously published protocol by Dabney et al. (2013), the new protocol uses the physicochemical properties of silica filter columns and guanidine to isolate DNA from digested bone lysates (Fig. 7).

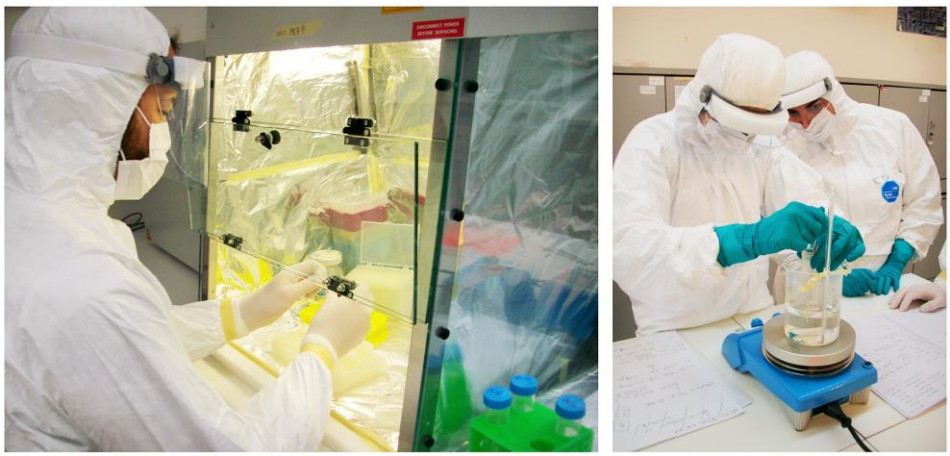


Fig. 7: Researchers wearing protective gear during DNA extractions of historical tuna remains.

In order to verify the species of each individual, all samples have had mini-barcodes of the following genes amplified via polymerase chain reaction: cytochrome oxidase subunit 1 (CO1), mitochondrial control region and internal transcribed spacer (ITS1). Sequencing of amplicons was out-sourced to Macrogen Europe and all alignments were conducted at UNIBO.

3.3 Quantification of historical DNA

All DNA has undergone rigorous quality and quantity analysis using multiple platforms (NanoDrop, Agilent BioAnalyzer, QuBit, qPCR). Most vertebrae from the Massimo Sella archive contained sufficient DNA for SNP genotyping purposes, as determined by the QuBit spectrophotometer (Table 1). However, samples older than 500 years possess concentrations of DNA lower than the threshold that can be quantified in this manner.

Table 1: Concentration of DNA extracted from samples from the Massimo Sella archive collected between 1911 -1926. DNA has been extracted from 50 vertebrae from each location.

	Average [DNA] (ng/µl)	**Maximum [DNA] value**	**Minimum [DNA] value**
Sliten trap samples	5.2 ± 3.24	15.7	0.37
Istrian trap samples	1.0 ± 1.61	10.3	0.05
Messina trap samples	5.4 ± 5.78	20.8	0.42

Spectrophotometry platforms (i.e. QuBit, NanoDrop) measure a sample's total DNA concentration, not the separate proportions of DNA belonging to the specimen (endogenous) and that which can be attributed to sources of contamination (exogenous). Preliminary analysis of historical DNA extractions, using the far more expensive Agilent BioAnalyzer platform, revealed that a large proportion of the total DNA in a handful of extractions consisted of long chains of nucleic acids (Fig. 8, Table 2). Since DNA is expected to fragment and degrade with time (exposure to catalytic enzymes, chemical decompostion and digestion by fungi and bacteria), longer strands of DNA are likely from exogenous sources.

Table 2: Proportion of total DNA belonging to historical sources as measured by the Agilent BioAnalyzer in four historical samples (CCF=La Chanca de Conil de la Frontera, 1755; PJ= Palacio de Justicia, Cadìz, 300-100 BCE).

Sample	**Time period**	**% of total DNA**	**[aDNA] (pg/µl)**	**Average size (bp)**
PJ48	300-100BC	39%	517.87	103
PJ114	300-100BC	48%	119.42	74
CCF86	1755	26%	7338.65	89
CCF131	1755	29%	1706.42	124

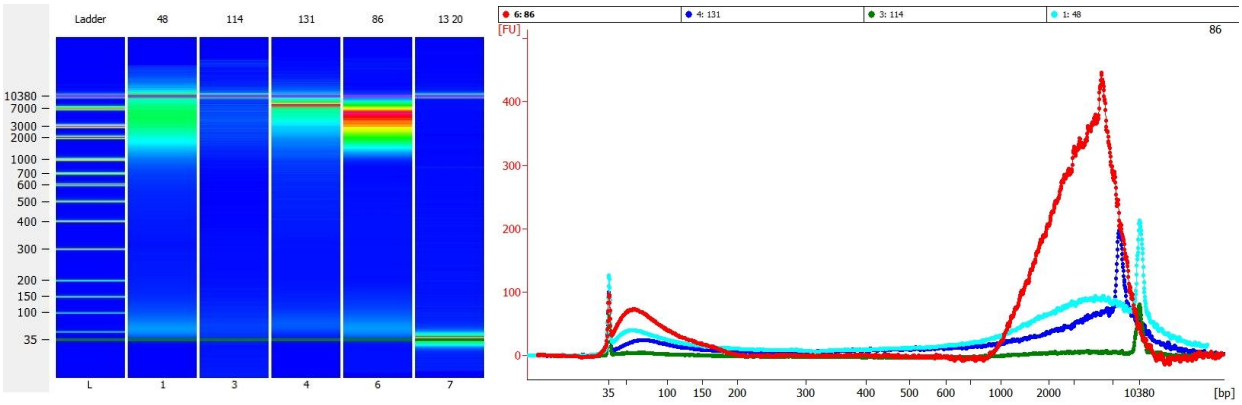


Fig. 8: Results from the analysis of four ancient samples from Spain (2 samples from La Chanca de Conil de la Frontera, 1755 and two samples from Palacio de Justicia, Cadìz, 300-100 BCE) using the Agilent BioAnalyzer. The picture on the left shows the length of DNA fragments on the y-axis with concentrations of DNA in columns, expressed as a gradient of colours. The picture on the right shows a peak of short fragment DNA (historical DNA; 40-300bp) and much larger peaks of long fragments of DNA (modern exogenous DNA; 400-10000bp).

3.4 Quantitative Polymerase Chain Reaction

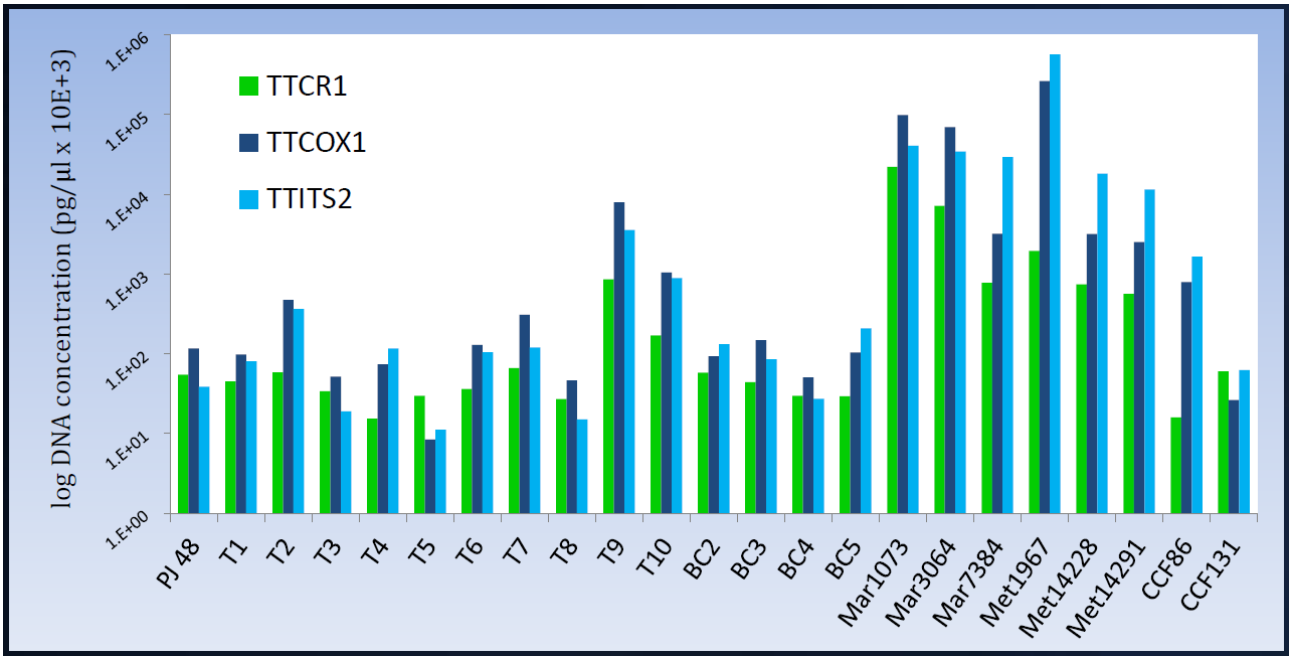
At the University of Gent (Belgium) a subset of all samples have been analyzed using quantitative polymerase chain reactions (qPCR). Several species-specific primer pairs were designed and used to execute all qPCRs in order to quantify the concentration of mitochondrial and nuclear DNA successfully extracted from each sample. Reactions were carried out using two mitochondrial genes (Control Region [TTC1R] and Cytochrome Oxidase subunit 1 [TTCOX1]) and one nuclear gene (Internal Transcribed Spacer Region 1 [TTITS1]). Comparisons have been made between the amount of DNA contained in each sample and their age and the environmental conditions which the bones had been exposed to. A manuscript detailing the novel DNA extraction techniques used and the conclusions concerning preservation of DNA in excavated remains is currently in development and will hopefully be submitted to a peer reviewed journal in early 2015.

Table 3: Number of samples analyzed with qPCR from each of ten different excavation sites and archives.

N. Samples	Location	Time period
5	Ionian Sea, Italy	1925-1926
5	Tyrrhenian Sea, Italy	1911
2	La Chanca de Conil de la Frontera, Spain (CCF)	1755
2	Marmaray, Yenikapi, Istanbul, Turkey (Mar)	300-1400 CE
3	Metro station, Yenikapi, Istanbul, Turkey (Met)	300-1400 CE
10	Baelo Claudia, Tarifa, Spain (BC)	100-200 BCE
10	Roman Balsa, Tavira, Portugal (T)	200 BCE
2	Palacio de Justicia, Cadìz, Spain (PJ)	300-100 BCE
1	Castro Marim, Portugal	200 BCE
2	Castro Marim, Portugal	400-300 BCE

Both nuclear and mitochondrial gene markers performed consistently for all reactions. Primer pairs used to amplify short fragments of CO1 and ITS1 performed better than the CR primers. Bones from the twentieth century collection yielded such high concentrations of DNA (mean yield = 1.8ng/µl) that they were excluded from further comparisons among the more ancient samples. Samples from Istanbul contained the highest concentration of DNA, while samples from Tavira yielded the lowest amount of DNA among the bones of ancient origin (Fig. 9).

Fig. 9: qPCR results for 23 samples of ancient origin using three different molecular markers (CO1, mtCR and ITS1). Samples are arranged chronologically. Name codes for sampling locations are featured in Table 3.



Focusing strictly on the Iberian samples, extractions with the highest concentration of DNA are perhaps the oldest amongst the collection from Castro Marim. Sample ROM542 is estimated to date back to the early roman republic (3rd-2nd century BC) and samples IA504 and IA535 were butchered by Iron Age fishermen. Samples with the lowest quantity of DNA are the vertebrae from the Roman settlement of Balsa in the Algarve, close to Tavira. Dried tuna vertebrae were incorporated into the building materials of the ancient Romans in the Algarve, presumably to add structural integrity as a strong organic brick of sorts. The low concentration of DNA in these bones may be a result of the way in which the bones were treated (boiling and drying in the hot summer sun) before being cemented into walls and floors. Alternatively, once inside of the walls they would have been exposed to extreme temperature variations between very high temperatures during the summer as the clay walls heated up during the day and cooled down at night, all of which is damaging for DNA. The age of bones clearly appears to have an influence

on DNA yield; however, the DNA extracted from ancient sources using our new technique is of sufficient quality and quantity to achieve our next challenge: high throughput SNP genotyping.

The results of this work were presented as a poster at the Mares Conference Marine Ecosystems Health and Conservation, Olhao, Portugal November 17-21, 2014:

GN Puncher, A Morales, V Onar, F Massari, A Cariani, E Cilli, H Arrizabalaga, F Tinti (2014) Unlocking the evolutionary history of the mighty bluefin tuna using novel paleogenetic techniques and ancient tuna remains (Poster).

4. SNP genotyping panel for 480 samples

A genotyping panel has been developed which contains SNPs that are indicative of genetic population structuring in modern populations as well as SNPs associated with genes that may hold selective or evolutionary significance. Within the ICCAT/GBYP Phase 4 contract for the *Data Recovery* programme, it plainly states that "The genetic data of these historical samples must be fully comparable with the existing genetic data of contemporary BFT populations already included in the ICCAT/GBYP database and obtained with the programs GBYP-Phase 3 and Phase 4 for maximizing data exploitability and cost reduction." Unfortunately, validation of the SNP panel developed during Phase 4 of the *Biological and Genetic Sampling and Analysis* (BGSA) Programme by the Genetics Consortium was only recently completed in December 2014. As a result, development of the SNP panel to be used for genotyping of historical samples has been delayed. Even so, in the first weeks of January 2015, the SNP panel was completed and an order will be placed with Fluidigm, before February for the development of all necessary chips for the genotyping of 184 historical samples and additional modern samples to ensure comprehensive comparisons across all time periods (Table 4). The SNPs contained in the panel were developed using 3 separate approaches: 1) Targeted Gene Approach, 2) Genotyping-by-Sequencing SNP development from BGSA Phase 4, and 3) Transcriptomic SNP development from BGSA Phase 2.

4.1. Targeted Gene Approach

Species specific primers have been developed to amplify regions of DNA that are likely to be associated with genes subject to selective pressures, such as those that influence metabolism, growth and immune response (myosin, heat shock protein, interleukin factor, hemoglobin, methylmercury binding protein). DNA fragments from these genes have been amplified (Fig. 10) from a group of 60 modern samples from 5 different locations in the species' geographic distribution (Gulf of Mexico, Balearic

Table 4: Samples to be genotyped using a high performance 96 SNP panel containing loci associated with population structuring and selective traits.

Geographic region	Modern samples	1911-1926 (Massimo Sella Archive)	1755 (Cadiz, Spain)	4^{th}-15^{th} century AD (Istanbul, Turkey)	1^{st} century AD (Cadiz, Spain)	2^{nd} century BC (Tavira, Portugal)	4^{th}-2^{nd} century BC (Portugal + Spain)	**TOTAL**
Adriatic Sea	**40** (2011)	**50** (Istrian traps)						**90**
Tyrrhenian Sea	**80** (2011)	**50** (Traps of Pizzo and Messina-Cialona)						**130**
NW Ionian Sea (Libya and Malta)	**40** (2011)	**50** (Trap of Sliten)						**90**
Gibraltar + Portugal	**60** (2011)		**2**		**10**	**10**	**5**	**107**
Levantine Sea	**60** (2007 + 2011)	**2** (Istanbul)		**5**				**47**
Gulf of Mexico	**16** (2009)							**16**
Total	**296**	**152**	**2**	**5**	**10**	**10**	**5**	**480**

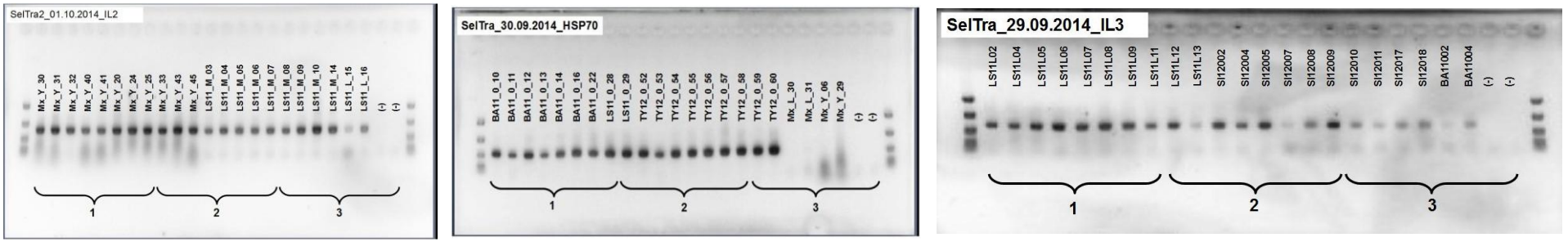


Fig. 10: Examples of electrophoresis gels showing successful amplification of targeted genes (interleukin factor and Heat shock protein 70) from ABFT samples collected throughout the species range.

Islands, Tyrrhenian Sea, Levantine Sea, Strait of Sicily). The PCR products of three of the most successful gene amplifications (Heat shock protein 70, Interleukin factor region 1 and 2) have been sequenced, aligned and analyzed for SNP loci to be used in the SNP genotyping panel.

4.2. Genotyping-by-Sequencing SNP discovery and comparison with annotated genes of related fish species.

As mentioned above, the 96 SNPs that were discovered and validated in Phase 4 of the BGSA Programme have been analyzed for their potential to be used for the purposes of detecting spatio-temporal shifts in genetic structuring of BFT populations as well as detect adaptation over time to environmental and fishery pressures. All SNP flanking sequences were compared with previously published sequences of annotated genes from other teleost species using the BLAST algorithm (Altschul et al. 1990) provided by the National Center for Biotechnology Information. A total of 35 SNP flanking sequences aligned with >100bp of previously annotated genes with an identity rate of >75%. The vast majority of these alignments occurred with well studied teleosts such as: *Dicentrarchus labrax, Takifugu rubripes, Poecilia formosa, Larimichthys crocea, Stegastes partitus, Oreochromis niloticus, Oryzias latipes, Maylandia zebra,* and *Haplochromis burtoni*. The genes with which these SNPs are likely to be associated are listed in Table 5.

Several of the SNPs that demonstrate higher capacity for population discrimination have been selected for additional investigation regarding the nature of their impact on gene expression. The flanking region sequences of these SNPs of additional interest have been aligned with several sequences from other teleost species and translated into the expected expression of amino acids (i.e. building blocks of proteins) in an effort to determine whether the different alleles of SNP modify the structure of the protein. For example, SNP_150 aligned with the *Myosin VIIB (myo7b)* gene which is an Actin-based motor molecule with ATPase activity that plays an important role in intracellular movements. In humans it is expressed in the eye and inner ear, the olfactory epithelium, brain, choroid plexus, intestine, liver, kidney, adrenal gland, and testis. During the development of the 384 SNP panel for Phase 4 BGSA, when SNPs were being selected due to extreme differences in pooled allele frequency differences across populations, this particular SNP was a strong candidate (Table 6). While guanine was the dominant nucleotide expressed in the Levantine Sea samples (60%), adenine was by far the dominant allele for samples captured in the Strait of Sicily and the Tyrrhenian Sea (92% and 93%, respectively).

Table 5: Possible genes associated with GBS derived SNPs.

Transmembrane protein 132 (TMEM132)
Leucine-rich repeats and immunoglobulin-like domains 1 (lrig1)
Pre-B-cell leukemia transcription factor-interacting protein 1
rho guanine nucleotide exchange factor 16
Unconventional myosin-VIIa
Zinc finger and BTB domain-containing protein 12
Regulator of G-protein signaling 20
Kappa-type opioid receptor
C4b-binding protein alpha chain
Protein disulfide-isomerase TMX3
Myosin heavy chain 1
Anaphase promoting complex subunit 2 (anapc2)
Bridging integrator 1 (bin1b)
Copper transporting ATPase 2 (ATP7B)
Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase
Excitatory amino acid transporter 1
Golgi brefeldin A resistant guanine nucleotide exchange factor 1 (gbf1)
GTP-binding protein Di-Ras1
HoxBb gene cluster
Immunoglobulin superfamily member 21
Kisspeptin 2 (Kiss2)
Laminin, alpha 5 (lama5)
S-adenosylmethionine mitochondrial carrier protein
Leucine-rich repeats and immunoglobulin
Lysyl oxidase homolog 1
Metabotropic glutamate receptor 2
Mothers against decapentaplegic homolog 4
Myc box-dependent-interacting protein 1
Myosin VIIB (myo7b)
Protein FAM73B
Protein kinase C delta
Retinol dehydrogenase 8
Sodium/myo-inositol cotransporter
Solute carrier family 5 (sodium/myo-inositol cotransporter), member 3 (slc5a3)
T-box brain protein 1 (TBR1)
Dermatan-sulfate epimerase-like protein
Anti-dorsalizing morphogenic protein

Table 6: Allele frequency of SNP associated with the Myosin VIIB gene from 555 samples pooled by geographic area from GBS SNP discovery by BGSA Genetics Consortium during GBYP Phase 4.

Location	**SNP read count**	**Allele A frequency**
Western Atlantic	205	0.77
Balearic Islands	816	0.55
Ligurian Sea	39	1.00
Tyrrhenian Sea	268	0.93
Strait of Sicily	188	0.92
Levantine Sea	498	0.40

After the validation of 96 high performance SNPs by the BGSA Genetics consortium, SNP_150 continued to show significant differences in allele frequency across sampling sites (Table 7).

Table 7: Allele frequency of SNPs associated with the Myosin VIIB gene from 167 samples pooled by sampling location from GBS SNP discovery by the BGSA Genetics Consortium during GBYP Phase 4 - 96 SNP panel validation.

Sample	A	G
Cape Hatteras, YOY, 2008	0.588	0.412
Gulf of Mexico, Larvae, 2008	0.688	0.312
Gulf of Mexico, Larvae, 2009	0.500	0.500
Strait of Sicily, YOY, 2011	0.857	0.143
Strait of Sicily, YOY, 2012	0.542	0.458
Levantine Sea, YOY, 2011	0.654	0.346
Balearic Islands, YOY, 2011	0.604	0.396

In figure 11, the SNP_150 sequence is aligned with several sequences for the same gene from other teleost species. At the bottom of the figure, an "R" replaces the location of the SNP at which adenine (A) and guanine (G) are expressed by the species. Interestingly, all other sequences available for this particle gene possess adenine at this locus. When the alignment is translated into amino acids, a non-synonymous base substitution, or a change in the expressed amino acid, is revealed (Fig. 12-13). The analysis that has revealed this example of significant modifications to the genetic code will be conducted for all SNPs that are now part of the final 96 SNP panel to be used for historical sample genotyping.

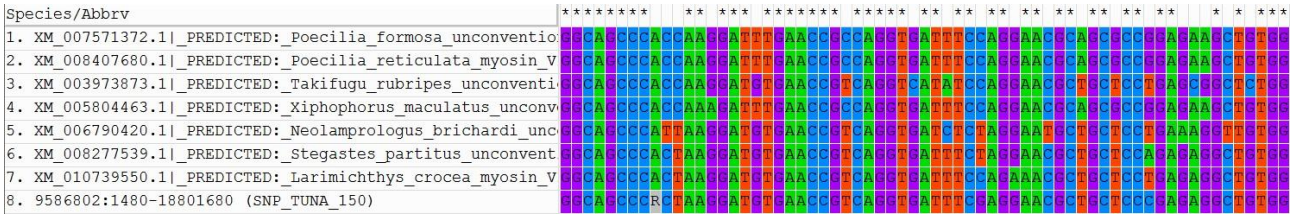


Fig. 11: DNA alignment of GBS derived SNP with predicted sequences (translated from sequenced proteins) from other teleost fishes of the Unconventional myosin VIIb gene. Note the 9th base in the tuna sequence, represented as an R, which indicates a SNP that is expressed as either guanine or adenine.

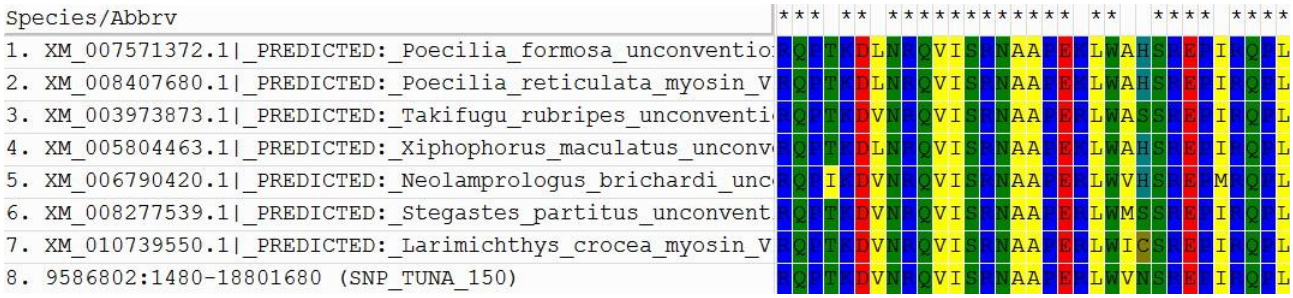


Fig. 12: Alignment of amino acids from various teleost fishes with a tuna sequence derived from translated DNA. Note the threonine amino acid present in position 4 when adenine is expressed in the genetic code.

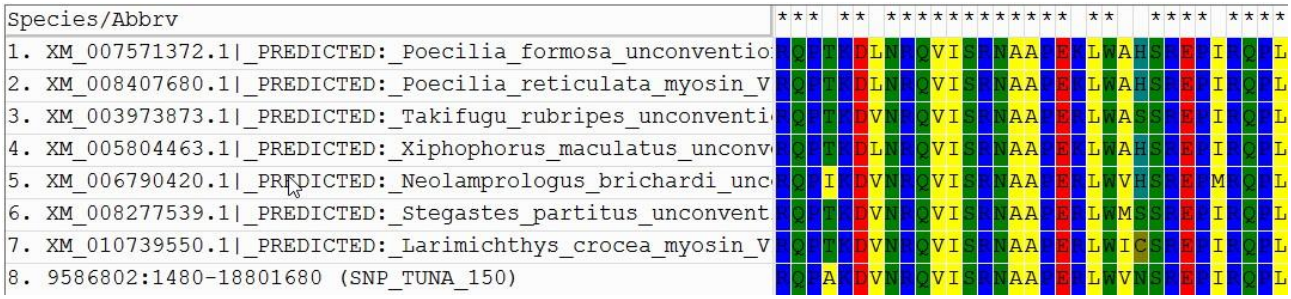


Fig. 13: Alignment of amino acids from various teleost fishes with a tuna sequence derived from translated DNA. Note the alanine amino acid present in position 4 when guanine is expressed in the genetic code.

4.3. Transcriptome SNP discovery and comparison with the annotated genome of Atlantic cod

All 384 SNPs discovered by the BGSA Genetic Consortium during Phase 2 were aligned with the publicly available annotated genome of the Atlantic cod, *Gadus morhua*. A total of 41 SNPs of the total number of SNPs derived from the analyzed transcriptome aligned with a high degree of similarity to annotated genes of cod (Table 8). These have all been added to the genotyping panel.

Table 8: Genes from the annotated cod genome with close flanking region sequence similarity with Transcriptomic derived SNPs from Phase 2.

Asparaginyl-tRNA synthetase
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c3 (subunit 9) genome duplicate b
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit
Betaine-homocysteine methyltransferase
Calponin 1, basic, smooth muscle, b
Collagen, type VI, alpha 1
Complement factor D (adipsin)
COP9 constitutive photomorphogenic homolog subunit 5
Elongation factor 1 alpha
Eukaryotic translation elongation factor 1 alpha 1a
Eukaryotic translation elongation factor 2a, tandem duplicate 2
Eukaryotic translation initiation factor 3, subunit G
Fructose-1,6-bisphosphatase 1b
Fructose-1,6-bisphosphatase 2
Hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit
Integral membrane protein 2Bb
Myomesin 1
Myosin, heavy chain 13, skeletal muscle
Myosin, light chain 12, genome duplicate 1
Myosin, light chain 9b, regulatory
Myozenin 1 and 1b
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9
Phosphoglucomutase 1
Ribosomal protein L18a, L27, L8, L28, L3, L31, L7, L9, S14, S18, S2, S23, SA
Thioredoxin interacting protein a
Ubiquinol-cytochrome c reductase hinge protein
Vacuolar protein sorting 28

To complete the 96 SNP panel, an additional 20 SNPs with high sampling location discrimination power (data derived from GBYP Phase 4 - Validation of 96 SNP panel by BGSA Genetics Consortium) were added to the 76 SNPs associated with annotated genes.

4.4 A quick note on the progress of the BGSA Genetics Consortium

The GBYP BGSA Genetics Consortium recently validated a 96 SNP panel using 167 young-of-the-year and larvae captured in the vicinity of known spawning areas. This validated panel is currently being used to genotype an additional 576 individuals, which will provide a more detailed estimate of the modern BFT population structure. That being said, the preliminary analysis conducted with a limited number of individuals during the validation process has yielded impressive results that give credit to the SNPs selected. Genetic admixture and population structuring was investigated using STRUCTURE (Pritchard et al. 2000), a software that employs a Bayesian model-based clustering method. Analysis of samples collected in the Mediterranean only, using a subset of 48 SNPs selected for high discriminatory power, suggested that the current Mediterranean stock descends from two ancestral populations. Furthermore, all individuals are admixed descendants of these two populations, each containing a unique proportion of genetic material derived from their ancestors (Fig. 14A). As such, there are unlikely to be "pure" individuals from any location, presenting a significant challenge to traceability and assignment efforts. When the number of ancestral populations are increased beyond the thresholds dictated by the STRUCTURE results, a more detailed illustration of the population structuring takes form (Fig. 14 B,C).

More impressive still are the analyses that incorporated samples from the Western Atlantic. Using a separate subset of 59 SNPs, the Bayesian analysis suggests descent from two ancestral populations (Fig.15A) and when additional clusters are used, the complexity of genetic structuring increases (Fig. 15 B,C). As the figures from both datasets demonstrate, individuals within each sampling group share similar proportions of ancestral genetic information. STRUCTURE also provided “alpha score” values (<1.0) for all scenarios, suggesting that the sampling locations reflect existing population structuring.

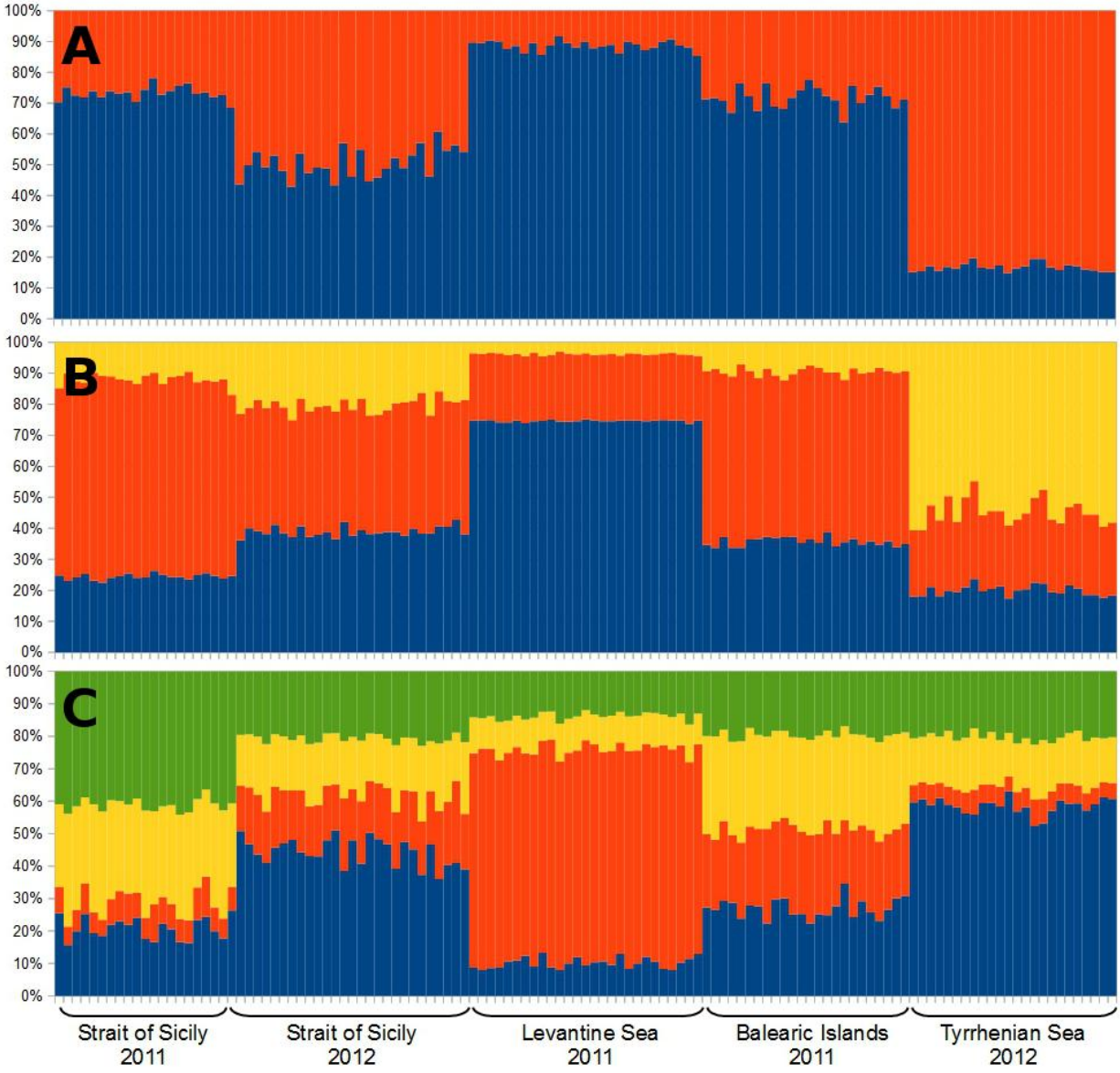


Fig. 14: STRUCTURE analysis of 123 BFT collected from the Mediterranean Sea using a subset of 48 SNPs from the validated 96 SNP panel developed by the BGSA Genetics Consortium. The model used assumes 2 (A), 3 (B) and 4 (C) ancestral populations.

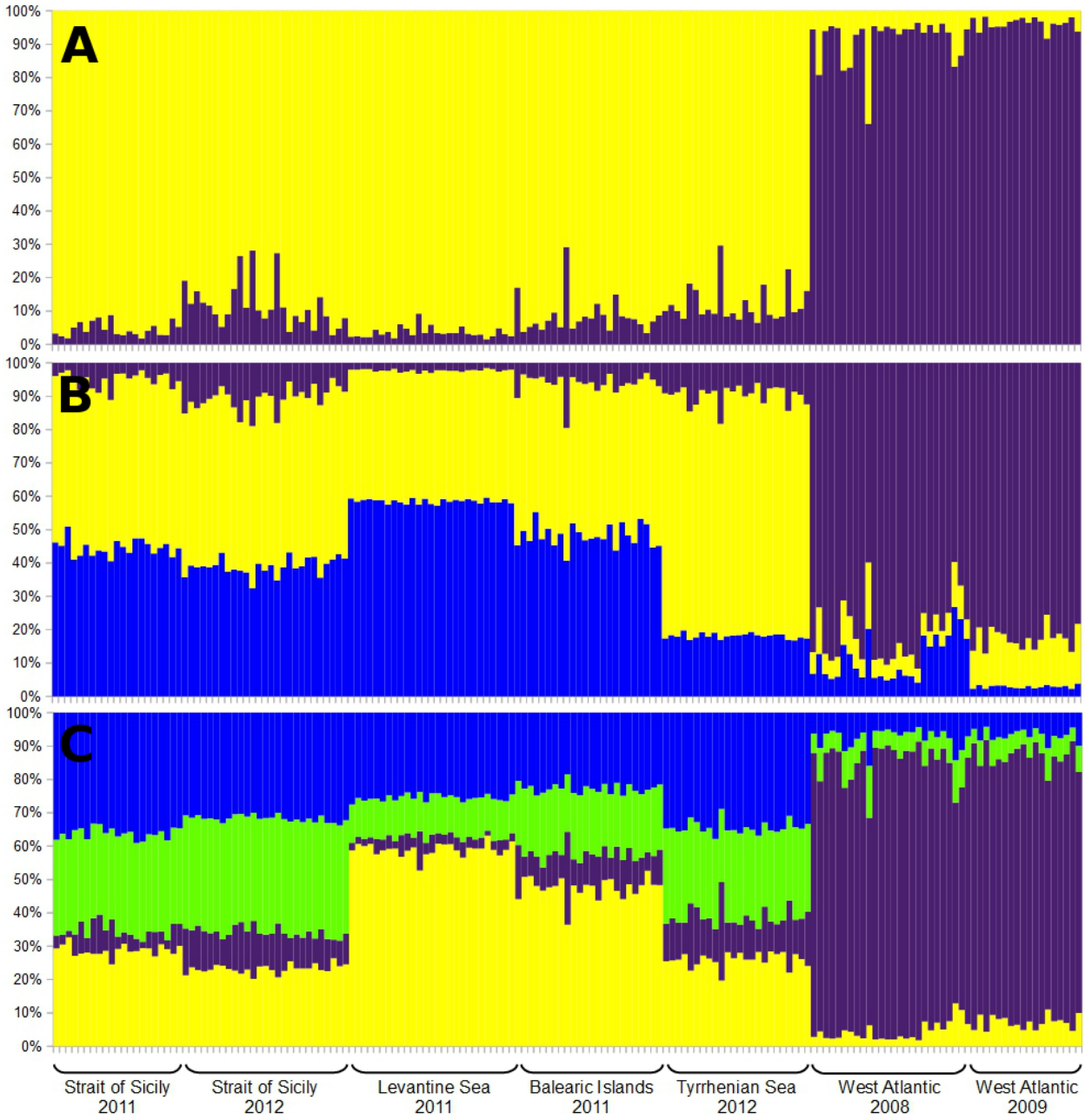


Fig. 15: STRUCTURE analysis of 167 BFT from the Mediterranean and Western Atlantic using a subset of 59 SNPs from the 96 SNP panel developed by the BGSA Genetics Consortium. The model used assumes 2 (A), 3 (B) and 4 (C) ancestral populations .

Genepop software (Raymond and Rousset 1995; Rousset 2008) was used to calculate the fixation index (Fst; a measure of population differentiation) for all pairs of populations within the Mediterranean Sea, using the aforementioned 48 SNPs. The population pair with the highest Fst value (0.0353) was the Tyrrhenian Sea and Levantine Sea. All samples were analyzed with a subset of 14 highly discriminatory SNPs and once again, the highest Fst values are reported for combinations of the Levantine Sea with the Balearic Islands, Strait of Sicily and the Tyrrhenian Sea (Table 9). Interestingly, the Tyrrhenian Sea and

Strait of Sicily pairwise comparison also has a high Fst value. Overall, these are by far the highest F_{st} values reported for paired BFT sampling locations in the Mediterranean Sea.

Table 9: F_{ST} matrix containing results from the analysis of five reference samples using 14 SNPs.

	Strait of Sicily 2011	Strait of Sicily 2012	Levantine Sea 2011	Balearic Islands 2011
Strait of Sicily 2012	0.0008			
Levantine Sea 2011	0.0471	0.0277		
Balearic Islands 2011	0.0332	0.0087	0.0563	
Tyrrhenian Sea 2012	0.0358	0.0316	0.0338	0.0252

These preliminary results from the BGSA Genetics Consortium clearly demonstrate the existence of genetic structuring within the Mediterranean Sea, within which BFT from the Levantine Sea form a distinct unit. Moreover, the evidence now suggests that the Tyrrhenian Sea spawning area is an additional source of genetically unique BFT. By genotyping historical samples using the same SNPs, it will be possible to determine whether this structuring has persisted over time. If hybridization of two ancestral populations occurred in the past, these tools will be able to provide an estimate of where and when this happened. This approach might also provide the whereabouts of the descendants of BFT that disappeared from the Black Sea in the mid-1980s.

5. Remaining duties

Manufacture of the final genotyping SNP panel will be completed by Fluidigm before the end of February and all samples will be genotyped shortly thereafter. After receiving the genotype data from Fludigm, a small amount of quality control will be required, after which data analysis can be completed within days. We expect to have the final deliverable for GBYP Phase 4 Data Recovery completed by the end of March 2015; however, we suspect that the results will have profound implications for both BFT management/conservation as well as the aquaculture/sea ranching industry (gene annotation). Beyond the preliminary results expected from this short contract, it will take some time to fully exploit the potential of all of the expected results from this ambitious project.

Contact information:

Gregory Neils Puncher	University of Bologna	gregoryneil.puncher2@unibo.it
Fausto Tinti	University of Bologna	fausto.tinti@unibo.it
Alessia Cariani	University of Bologna	alessia.cariani@unibo.it
Vedat Onar	University of Istanbul	onar@istanbul.edu.tr
Arturo Morales	Autonomous University of Madrid	arturo.morales@uam.es

References

Akyuz, E.F. and Artuz, G. (1957) Some observations on the biology of tuna (*Thunnus thynnus*) caught in Turkish waters. Proceedings and Technical Papers, General Fisheries Council for the Mediterranean, Rome 4(14): 93-99.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215: 403-410.

Aristotle (350 BCE) The history of animals. Book VIII. Chapter XIII. Translated by D'Arcy Wentworth Thompson. The internet classics archive [Online] http://classics.mit.edu/ Accessed 2014-09-12.

Cannas R., Ferrara G., Milano I., Landi M., Cariani A., Addis P., Cau A., Piccinetti C., Sella M. and Tinti F. (2012) Spatio-temporal genetic variation of Atlantic bluefin tunas from Sardinian and Mediterranean tuna traps. ICCAT-GBYP Symposium on Trap Fishery for Bluefin Tuna, Tangier (23-25 May 2011). Collect. Vol. Sci. Pap. ICCAT, 67(1): 351-358.

Dabney, J., Knapp, M., Glocke, I., et al. (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. PNAS 110(39): 15758-15763.

Daskalov, G.M. (2002) Overfishing drives a trophic cascade in the Black Sea. MEPS 225: 53-63.

Karakulak S.F. and Oray I.K. (2009) Remarks on the fluctuations of bluefin tuna catches in Turkish waters. Col. Vol. Sci. Pap. ICCAT, 63: 153-160.

Kemp, B.M., and Smith, D.G. (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. Forensic science international 154(1): 53-61.

MacKenzie B.R. and Mariani P. (2012) Spawning of bluefin tuna in the Black Sea: Historical evidence, environmental constraints and population plasticity. PloS ONE 7(7): e39998. doi:10.1371/journal.pone.0039998

Mather F. J.III, Mason J.M. and Jones A.C. (1995) Historical Document: Life History and Fisheries of Atlantic Bluefin Tuna. NOAA Technical Memorandum, NMFS-SEFSC, pp. 370.

Onar, V., Pazvant G., Armutak A. (2008) Radiocarbon dating results of the animal remains uncovered at Yenikapi Excavations. Istanbul Archaeological Museums, Proceedings of the 1st Symposium on Marmaray- Metro Salvage Excavations, May 5th-6th, Istanbul, 2008, pp. 249-256.

Onar, V., Alpak, H., Pazvant, G., Armutak, A., and Chroszcz, A. (2012) Byzantine horse skeletons of Theodosius harbour: 1. Paleopathology. Revue de Médecine Vétérinaire, 163: 139-146.

Oppian (177BCE) Halieuticks. *In* Diaper W. (1722) Oppian's Halieuticks of the nature of fishes and fishing of the ancients. Oxford. pp. 232

Papaconstantinou C., and Farrugio H. (2000) Fisheries in the Mediterranean. Mediterranean Marine Science 1(1): 5-18.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. Genetics 155(2): 945-959.

Raymond M. and Rousset F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity 86:248-249.

Riccioni, G., Landi, M., Ferrara, G., et al. (2010) Spatio-temporal population structuring and genetic diversity retention in depleted Atlantic bluefin tuna of the Mediterranean Sea. PNAS 107(5): 2102-2107.

Rousset, F. (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resources 8: 103-106.

Sara G., Dean, J. M., D'Amato, D., Buscaino, G., et al. (2007) Effect of boat noise on the behaviour of bluefin tuna *Thunnus thynnus* in the Mediterranean Sea. MEPS 331: 243-253.