UNIVERSITÀ POLITECNICA DELLE MARCHE
DIPARTIMENTO SCIENZE DELLA VITA E DELL'AMBIENTE

XXIX ciclo

Corso di Dottorato in

SCIENZE DELLA VITA E DELL'AMBIENTE

Curriculum Biologia ed Ecologia Marina

GENETIC DIVERSITY AND STOCK STRUCTURE IN TELEOSTS OF INTEREST TO FISHERIES

Dottoranda:

Dr. Tatiana Fioravanti

Tutor:

Chiar.mo Prof.
Vincenzo Caputo Barucchi

A.A. 2013/2016

CONTENTS

THESIS OBJECTIVES

The "*Food and Health Organization*" of the United Nations (FAO 2016) has showed that the global amount of fish capture in 2014 was 93 million tonnes, of which 81.5 million tonnes from marine fisheries and 11.9 million tonnes from inland fisheries. FAO (2016) has also reported that 31.4% of fish stocks are exploited at a biological unsustainable level and can be considered overfished. In addition, the collapse of marine species due to high fishing efforts can be also influenced by life history traits and climate variability (Pinsky & Byler 2015). The effects of climate changes severely affect freshwater biodiversity (Sala *et al.* 2000). In fact freshwater fishes are negatively impacted especially by high temperature, low levels of oxygen dissolved and modification in water flow (Ficke *et al.* 2007), which will lead to a drastic reduction in freshwater fish species in the next years (Xenopolous *et al.* 2005). The acquisition of information on the current genetic population structure and on how the genetic diversity has changed across time as results of overexploitation and/or environmental changes can be useful to develop new management and conservation strategies in order to prevent the size reduction of populations, their genetic impoverishment and losses in global biodiversity.

This PhD work consist of four parts, the first two provide useful information with regard to the genetic population structure of two pelagic species of interest to fisheries within the Mediterranean area (La Mesa *et al.* 2005; Morello & Arneri 2009), the Transparent goby (*Aphia minuta*) and the European anchovy (*Engraulis encrasicolus*). The genetic structure of these pelagic species was analysed using microsatellites as molecular markers and some hypotheses were provided to explain

which forces, among life history traits, historical events, hydrographic barriers and environmental conditions, shaped it. In addition, for *E. encrasicolus* a series of demographic fluctuations were observed in Adriatic Sea during the last 40 years, with a demographic collapse in 1987 (Santojanni *et al.* 2003). To evaluate if species with large census size, as European anchovy, can be negatively impacted at genetic level by overexploitation and/or environmental changes is extremely important for their management. With this aim, a temporal genetic analysis was performed using microsatellite loci on genomic DNA extracted from otoliths of *E. encrasicolus* provided by the ISMAR-CNR of Ancona (Italy), which collected them between 1978 and 2000 for aging fish and determining their growth patterns. For the last part of this PhD work, a freshwater fish (*Salmo trutta* complex), adapted to live in cold and well oxygenate waters (Jonsson & Jonsson 2009) and for which mtDNA genetic diversity was deeply examined within the Mediterranean area (Bernatchez 2001; Cortey *et al.* 2009), was used as target species. Specifically, *S. trutta* ancient bones coming from a stratigraphic succession that encompass a period of drastic climate changes (Pleistocene-Holocene transition) was analysed at genetic level using the mitochondrial DNA control region (D-loop) as molecular marker. The mtDNA haplotype diversity was analysed to understand if this temperature sensitive species was affected at genetic and ecological level by past climatic oscillations and to infer how will react to present and future environmental changes.

INTRODUCTION

Stock definition and identification

Since ancient times, humans have used fishes for their sustenance (O'Connor *et al.* 2011). Ancient fishing activities had a very low impact on marine and freshwater communities, but with the development of new and more efficient technologies and with the increase of human global population, the impact of fisheries on exploited species has become very high (Muscolino *et al.* 2012). The correct individuation of discrete units within species and the definition of methodologies to investigate genetic population structure seem necessary to manage current and future fishing activities. In this way, it will be possible to achieve the restoration of depleted populations and to avoid a future loss of fish biodiversity. One of the most used terms in fisheries management to describe this intraspecific unit is "*stock*". Several definitions of stock have been proposed by a lot of researchers (Waldman 2005). Unfortunately, there is not a universally accepted definition to date. Initially, the term stock was only attributed to the part of fish population harvested by anthropogenic activities (Dahl 1909). Because of the practical nature of this definition, it is better to refer to it as "*harvest stock*". However, the increasing knowledge in the field of genetics and the importance of this species sub-unit, not only for fisheries management but also for conservation biology, led to the introduction of other stock definitions mostly focused on biological, ecological and genetic aspects of fish populations. One of the most accepted definitions was provided by Ihssen *et al.* (1981). This author defines a stock as "an intraspecific group of randomly mating individuals with temporal or spatial integrity", while other definitions emphasized better the necessity to distinguish between a "*phenotypic stock*" (Booke 1981: "a group of fish characterized by phenotypic differences entirely induced by the

environment"), and a "*genetic stock*", defined as "a reproductively isolated unit, which is genetically different from other such units" (Carvalho & Hauser 1994).

As briefly reviewed above, the definition of stock is controversial. However, the identification of a stock is more important than its accurate definition for fisheries management. Despite different techniques can be used for this purpose (Begg & Waldman 1999), a holistic approach, based on several of these techniques, was proposed to define stock structure within fish species because it provides more reliable results (Begg & Waldman 1999). One of the first techniques applied in stock identification was the capture-mark-recapture method which allows, using physical or electronic tags, the tracking of fish movements, bringing also to information about homing and schooling beahviour and growth rates (Hall 2014). Hansen & Jacobsen (2003) analysed the origin of Atlantic salmon (*Salmo salar* L.) tagged in the Atlantic area north of the Faroe Islands and recaptured after their return to freshwaters. This study highlighted that the vast majority of individuals captured in the waters of Faroe Island belonged to the Norwegian stock and only a small fraction to the Scottish, Russian and Irish stocks. Among tagging methods, the use of parasites as natural markers is also very common. This technique is based on the fact that geographically separate populations of the same species can be infected by distinct parasites, or by the same parasite with different levels of infection (Catalano *et al.* 2014). MacKenzie *et al.* (2008) used larval nematodes (*Anisakis spp.* and *Hysterothylacium aduncum*) as a natural tag to clarify the stock structure of the Atlantic horse mackerel (*Trachurus trachurus* L.) within the European area. They observed a different pattern of infection between North Sea individuals and those from other areas confirming the management of the North Sea population as a separate stock. In order to infer the

population structure of species, also the analysis of phenotypic variation or the chemical analyses of calcified structures (otoliths and scales) may be useful. Specifically, "*phenotypic stocks*" can be revealed on the basis of meristic (countable) or morphological (descriptive) features (Cadrin *et al.* 2014), whereas chemical analyses of calcified structures indicate if individuals come from distinct stocks because differences in the elemental composition of otoliths and scales are caused by environmental conditions of different areas in which they live (Kerr & Campana 2014). For example, the management of the Icelandic cod (*Gadus morhua* L.) population as a single stock has been questioned by Jónsdóttir *et al.* (2006), which showed the existence of three different stocks on the basis of differences in otoliths chemistry. Finally, "*genetic stocks*" can be distinguished highlighting differences at genetic level through the use of genetic markers (Begg *et al.* 1999; Cuéllar-Pinzón *et al.* 2016).

Genetic markers

In order to reveal genetic variations within populations, researchers usually employ "*genetic markers*", molecular characteristics used to infer the genotype of individuals. Among them two different types can be recognized: protein and molecular or DNA markers (Abdul-Muneer 2014). The first genetic marker employed to study genetic variation were protein markers (Allendorf *et al.* 2013), called "*allozymes*" that represent alternative forms of an enzyme encoded by different alleles at the same locus. In this case, genetic variations determine differences in amino acid composition which can be detected by gel electrophoresis. In fact, amino acid modifications cause changes in structure and charge of different allozymes, leading to a different migration pattern during electrophoresis

(Kapuscinski & Miller 2007). Unfortunately, the genetic variability could be underestimated using allozymes because protein electrophoresis does not allow the detection of genetic changes that do not affect amino acid composition, such as silent substitutions within codons or changes in non coding regions (Allendorf *et al.* 2013).

From the 1990s, after the advent of the "*Polymerase Chain Reaction*" (PCR) (Saiki *et al.* 1985; Mullis & Faloona 1987), the number of molecular or DNA markers grew in importance. These latter can be distinguished between mitochondrial and nuclear markers on the basis of their collocation in nuclear or mitochondrial genome (Park & Moran 1994). Mitochondrial DNA (mtDNA, Fig. 1) is a circular DNA molecule (between 16,000 to 20,000 bp in length in fish) located inside each mitochondrion within the eukaryotic cell.

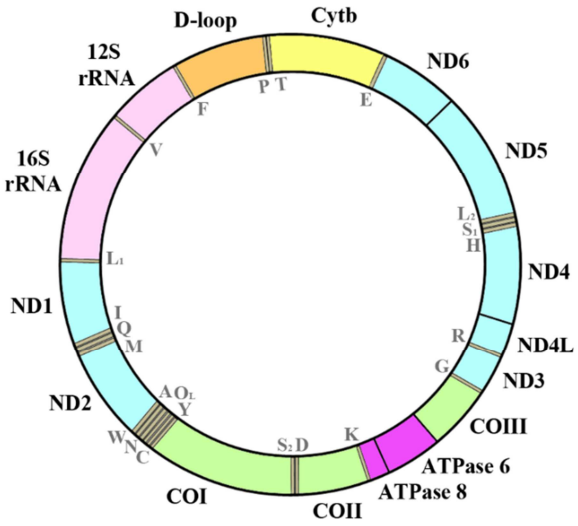


Fig. 1 Schematic representation of the vertebrate mitochondrial genome.

Several features make the mtDNA one of the most important markers used in phylogenetic, systematic and population genetic studies. Firstly, mtDNA is haploid, maternally inherited without recombination and exhibits a higher rate of mutation than nuclear genes (Brown 1985), which allows to create differences at species level

in a short interval of time. Indeed, it was observed that distinct regions of the mtDNA evolve at different rates compared to others (Meyer 1993). This feature makes distinct regions suitable for different studies. For example, the mitochondrial control region (D-loop) and the Cytochrome-b gene have a high mutation rate and seem useful for studies at species or population level (Brown *et al.* 1993). The same is for the mitochondrial Cytochrome-c oxidase I gene (CoI) that was selected as a standardize marker for identification of animal species in "the International Barcode of Life project" (iBOL) (Ratnasingham & Hebert 2007). In contrast, ribosomal genes (12S rRNA and 16S rRNA) evolve slowly and are useful for studies at higher taxonomic levels (Meyer et al. 1993). Secondly, analyses on mtDNA are relatively easy because this DNA molecule occurs with hundreds to thousands copies per cell, allowing its amplification also in degraded samples. In addition, variations in mtDNA sequences can be detected in a simple way by direct sequencing, or by the use of restriction enzymes which recognize and cut at specific nucleotide sites to generate "*Restriction Fragments Length Polymorphisms*" (RFLPs) which can be highlighted by gel electrophoresis. Unfortunately, mtDNA analyses are expensive and, due to the maternal inheritance of mtDNA, are not efficient to reveal male mediated processes (Kapuscinski & Miller 2007).

Another group of molecular markers is represented by nuclear DNA (nDNA) markers, which include "*Random Amplified Polymorphic DNA*" (RAPD), "*Amplified Fragment Length Polymorphisms*" (AFLPs), "*Variable Number of Tandem Repeats*" (VNTRs), in which minisatellite and microsatellite are included, and "*Single Nucleotide Polymorphisms*" (SNPs) (Abdul-Muneer 2014). Thanks to several features, the most common molecular markers used in population genetic studies are

microsatellites, also known as "*Simple Sequence Repeats*" (SSRs) (Tautz 1989) or "*Short Tandem Repeats*" (STRs) (Edwards *et al.* 1991). Microsatellites consist of short tandemly repeated units of 1-6 bp in length and are considered as "*neutral markers*" because they are principally found in non-coding regions of nDNA (Selkoe & Toonen 2006). Furthermore, they are Mendelian inherited and show a codominant expression, because allows the distinction between heterozygote and homozygote individuals. Microsatellites are also characterized by high mutation rates and high polymorphism because mutations lead to changes in the number of repeats and in sequence length, bringing to an increase in number of alleles at a specific locus (Selkoe & Toonen 2006). One of the most accepted mutation models for microsatellites is the Stepwise Mutation Model (SMM), in which each mutation results in the gain or loss of a single repeat unit (Slatkin 1995). However several others mutation models were proposed, but the best among them has not been yet accepted (Selkoe & Toonen 2006, Putman & Carbone, 2014). The analyses of microsatellites are very easy because they are flanked by conserved sequences of DNA that allow the design of specific primers for their PCR amplification (Selkoe & Toonen 2006). In addition, distinct alleles that differ in length can be simply revealed by gel electrophoresis or by automated DNA sizing, if during PCR amplification fluorescent-labeled primers are used. Although working with microsatellites seems very simple and convenient for genetic population studies, many disadvantages can be recognized. Firstly, for some species there are not available primers to amplify microsatellites and the isolation of new microsatellite loci and the design of specific primers could be expansive in term of resources and time (Zane *et al.* 2002). Secondly, a series of errors can occur during PCR amplification, such as the development of shadow or stutter bands, the non amplification of an allele due to

mutations in flanking regions that inhibit the primers annealing (null alleles), or the non amplification of longer alleles that are disadvantaged respect to shorter ones during amplification (large allele dropout). In particular, the presence of stutter bands induce to interpret a homozygote as heterozygote, instead the occurrence of null alleles or large allele dropout can lead to an incorrect genotyping of heterozygote individuals as homozygotes (Selkoe & Toonen 2006).

Genetic variations and stock management

The Hardy-Weinberg principle (Hardy 1908; Weinberg 1908) is based on the concept of "*ideal population*" that is defined as an infinite population in which random mating, no mutations, no immigration and no natural selection take place. Consequently, this ideal popoulation does not evolve and allele and genotype frequencies are constant between generations. However, natural populations deviate from HW equilibrium because are subjected to several forces, such as mutation, migration, genetic drift and natural selection, which determine changes in genetic diversity within and between populations. The major sources of genetic variation in populations are mutations (Allendorf *et al.* 2013) that occur naturally at chromosomal and molecular level (affect single nucleotides or more nucleotides). In several cases, mutations are lethal because lead to phenotypic changes that have a negative effect on the fitness of individuals. In other cases positive effects occur and new genetic variants are moved from generation to generation, thus increasing the genetic variation within populations. In fact, mutations that occur at molecular level modify DNA sequences, creating different alleles and increasing the level of polymorphism at a specific locus. A population can acquire new genetic variants also thanks to the immigration of individuals from other populations of the same species

(Allendorf *et al.* 2013). This migration process, that includes a transfer of new alleles, is known as "*gene flow*" and leads to an increase in the genetic variation within a population. A high immigration rate can cause a genetic homogenization of the involved populations bringing to the formation of a new larger single population. An additional evolutionary force is the natural selection (Allendorf *et al.* 2013) that occurs when some genotypes may bring to a greater fitness than others under certain environmental conditions. These advantageous variants tend to be more widespread in a population and differ between distinct populations subjected to different environmental pressures, increasing the level of genetic differentiation between populations.

Genetic variations between different populations can be detected using genetic markers. Its detection can led to the identification of "*genetic stocks*" (Begg *et al.* 1999; Cuéllar-Pinzón *et al.* 2016). Generally, freshwater species have been always considered as more genetically structured than marine ones, because the marine environment was considered without barriers which limit gene flow between populations (Waples 1998; Avise 2000). However, several studies have highlighted that other factors could determine the genetic structure of marine species, for example historical and paleoclimatic events, hydrographic barriers, life history traits and selective pressure (Avise 2000; Patarnello 2007). For this reason a lot of fish species of commercial interest are divided into genetically distinct populations that must be considered as separate stocks to management purpose with the aim to avoid their reduction or extinction (Reiss 2009). For example, the Atlantic cod was subdivided into different genetic stocks using microsatellite loci (Hutchinson *et al.* 2001), whereas using allozymes (Mork *et al.* 1985) and mtDNA (Smith *et al.* 1989)

no structure had been highlighted. Recently, a genetic structure was also evaluated for other marine fishes of commercial interest, such as Atlantic herring (*Clupea harengus* L.) in Baltic Sea (Jørgensen 2005), sardine (*Sardina pilchardus*) in Adriatic Sea (Ruggeri *et al.* 2013) and Yellowfin tuna (*Thunnus albacares*) in Pacific Ocean (Aguila *et al.* 2015).

A stock can be considered endangered when its size is reduced by natural or anthropogenic factors, because usually small populations may undergo a process known as "*genetic drift*", a change in allele frequencies caused by random sampling of individuals (Allendorf *et al.* 2013). Special cases of genetic drift can be determined by a drastic reduction in the number of individuals in a population, an event known as "*bottleneck*", or by a "*founder effect*" that occur when a new population arises from a very limited number of founders (Allendorf *et al.* 2013). In the long term, genetic drift leads to a loss of variation within populations and to inbreeding. Inbreeding occurs when related individuals randomly mate in a small population. This leads to an increase in the number of homozygotes and to a reduction in genetic diversity, sometimes with negative effects on the survival of individuals. In fact disadvantageous alleles, recessive in heterozygote individuals, become more frequently expressed in a population rich of homozygotes, bringing to a reduction in fitness and limiting survival probabilities of the affected population, a process known as "*inbreeding depression*" (Frankham 1999; Charlesworth & Willis 2009). Sometimes, when the gene flow between different populations of the same species is reestablished, individuals belonging to them can interbreed. Unfortunately, this process could lead to the "*outbreeding depression*", a reduction in fitness of the progeny produced by genetically distant individuals (Frankham 1999). The

inbreeding is best known for fishes that have suffered drastic reduction in size or employed in captive breeding programs, such as various salmonids (Wang *et al.* 2002). With the aim to evaluate if the endangered population of Chinook salmon (*Oncorhynchus tshawytscha*) of the Sacramento River can be negatively impacted by inbreeding, Arkush *et al.* (2002) analysed the genetic resistance to three pathogens in salmon individuals from broodstock. This study highlighted that homozygote individuals for MHC (Major Histocompatibility Complex) genes were more affected by pathogens than heterozigotes. This observation prove that a continue decrease in population size and an increase in homozygotes due to inbreeding, will bring the Sacramento river population of Chinook salmon to a major susceptibility to pathogens infection and ultimately to the extinction. Among salmonid species also the outbreeding process seems rather apparent. McGinnity *et al.* (2003) demonstrated that the deliberated or accidental introduction of the Atlantic salmon (*Salmo salar* L.) from farms into the wild could negatively affect the fitness and survival of endangered natural populations due to a reduced survival of farm and hybrids individuals in freshwater and marine environments.

The effects of genetic drift are more pronounced in small populations. However, in natural populations not all individuals have the same reproductive success and the census size (N_c) is not representative of the real number of individuals that breed. In order to evaluate the effects of genetic drift is better refer to the effective population size (N_e) and not to the census size (N_c) of populations (Frankham 1999). The effective population size (N_e) is defined as "the size of an ideal population with the same rates of genetic change as observed one, and thus provides information on evolutionary change as well as ecological differences between observed and ideal

population" (Hauser & Carvalho 2008). A decrease in N_e within a population could bring about a reduction in the reproductive and evolutionary potential, increasing the probability of deleterious effect of inbreeding and genetic drift (Hauser & Carvalho 2008). The N_e and N_e/N_c ratio estimation are necessary, together with the evaluation of genetic diversity, to understand if fish populations are threatened by over-exploitation or natural changes, and to organize correct management and conservation strategies. Hoarau *et al.* (2005) have estimated N_e parameter in an overexploited flatfish, the European plaice (*Pleuronectes platessa* L.), using microsatellite loci on historical samples from North Sea. Very low levels of *Ne* with reference to the large census size of this species were estimated and inbreeding depression was highlighted as a probable effect of overexploitation occurred from the 1960s.

Usefulness of ancient and historical DNA studies

The ancient DNA (aDNA) field of research was born during the mid 1980s when, using bacterial cloning, fragments of DNA extracted from the museum-preserved skin of the extinct quagga (*Equus quagga quagga*, Boddaert 1785) (Higuchi *et al.* 1984) and from an Egyptian human mummy (Pääbo *et al.* 1985) were amplified. Unfortunately molecular cloning of aDNA turns to be a very inefficient technique, and it was only with the development of the Polymerase Chain Reaction (PCR, Saiki *et al.* 1985; Mullis & Faloona 1987), a procedure that allows the exponential amplification of specific DNA sequences even if present in few copies, that there was an outbreak in aDNA studies.

"*Ancient DNA*" should be defined as DNA extracted from biological materials coming from paleontological and archaeological sites or natural history collections. Specifically, the main sources of aDNA are: skins, hair, feathers, bones, teeth, scales, otoliths, pinned insects, herbarium specimens and formalin-fixed tissues (Wandeler *et al.* 2007). Usually, based on the age of samples, two different terms could be used to identify DNA extracted from old samples: "*historical* or *archived DNA*" which is DNA extracted from samples taken decades or even centuries and truly "*ancient DNA*" which is DNA from samples hundreds to several thousand years old (Nielsen & Hansen 2008). However, the definition "ancient" is not strictly based on the age of samples but especially on biochemical conditions of DNA molecules. Indeed, after the death of an organism, DNA is rapidly attacked by endogenous nucleases and microorganisms (Eglinton 1991) which, together with hydrolytic and oxidative processes, degrade DNA molecules in fragments ranging from 100 to 500 bp in length. In addition, other reactions could cause strand breaks, intra- and inter-molecular crosslinks, blocking and miscoding lesions, which could bring to the formation of sequencing artefacts and to the interruption of PCR amplification with an increasing risk of contamination by exogenous DNA (Lindhal 1993; Willerslev & Cooper 2005). In order to avoid or detect contamination, the nine "gold criteria" of authenticity proposed by Cooper & Poinar (2000) must be always followed when working on aDNA. They mainly include the physical isolation of the working area, preparation of negative controls for both extraction and amplification of the samples, cloning of PCR amplicons and reproducibility of results.

The majority of studies on aDNA were focused on relationships between extinct and extant species (Höss *et al.* 1996; Cooper *et al.* 2001), generally through the use of

mtDNA which is present with hundreds of copies per cell, making aDNA studies more feasible. However, this methodology was also applied in many other fields of research, for example: i) to understand diet and behaviour of extinct and extant species by extracting aDNA from coprolites (Poinar *et al.* 1998, 2001); ii) to analyse sedimentary DNA from permafrost and caves with the aim to detect the presence of different organisms for which no macrofossil remains are present (Willerslev *et al.* 2003, 2004); iii) to reveal details of domestication process of plants and animals (Troy *et al.* 2001; Jaenicke-Despres 2003); up to iv) the identification of pathogens that have caused diseases in humans for several centuries (Salo *et al.* 1994; Reid *et al.* 1999). In addition, genetic analyses of historical and ancient samples can provide a lot of useful information for the protection of threatened and endangered species (Leonard 2008), adopted also in fish species management and conservation (Nielsen & Hansen 2008). In particular, if one pays attention to this last area of interest, habitat modification and recreational fisheries have often led to the introduction of non-native species in areas outside their original distribution range. In these cases, the comparison between historical and modern samples could be useful to evaluate the level of introgression between native and non-native individuals (Hansen 2002; Larsen *et al.* 2005; Sønstebø *et al.* 2008) or to highlight if the indigenous population was completely replaced by non-native species (Nielsen *et al.* 1997). These studies are also important to identify pure native individuals that can be used in supportive-breeding and restoration programs (Hansen *et al.* 2006). Moreover, the comparison of historical and present genetic diversity in fish populations subjected to an intense fishing effort could be useful to identify a loss in genetic variation within populations, or also a modification in population structure in time. For example, Heath *et al.* (2002) highlighted a change in time in population structure in the

Rainbow trout (*Oncorhynchus mykiss*) from northern British Columbia, and a reduction in genetic diversity was observed in over-exploited species, such as the New Zealand snapper (*Pagrus auratus*) by Hauser *et al.* (2002). The estimation of genetic diversity within population is usually related with the evaluation of the effective population size (N_e), a parameter that allows the estimation of the number of individuals in a population that contribute genetically to the next generation. A decrease in N_e within a population could bring about a reduction in the reproductive and evolutionary potential, increasing the probability of extinction. The N_e estimation is necessary, together with the evaluation of genetic diversity, to understand if fish populations are threatened by over-exploitation, and to organize correct management strategies. Generally, high N_e values were always hypothesized for marine species, because they usually have a large census size (N_c). However several studies on historical samples have highlighted that some marine fish have values of N_e and N_e/N_c ratio lower than previously expected (Hoarau *et al.* 2005; Ruggeri *et al.* 2012). Unfortunately, not only anthropogenic activities, but also natural environmental changes could affect fish population stability. One of the most important problems in species conservation is to understand how species react to climate changes. Studies on ancient samples could be useful also in this case as suggested by Pauls *et al.* (2013), because they may show how species have reacted to past climate changes. This will allow us to predict how these species will adapt to the changed conditions. However, this field of research has mainly been focused on terrestrial and marine mammals rather than on fish species.

Target species of this PhD work

Aphia minuta

The transparent goby *(Aphia minuta*, Risso 1810) (Fig. 2) is a species belonging to the monotypic genus *Aphia* (family Gobiidae, order Perciformes). In several countries this species has a great commercial interest and a series of local common names are used: "*transparent goby*" in United Kingdom, "*chanquete*" in Spain and "*rossetto*" in Italy (La Mesa *et al.* 2005). In contrast to the other gobies, *A. minuta* is a pelagic and neritic species, widely distributed along the eastern part of the Atlantic Ocean, from the Baltic Sea to Morocco, within the Mediterranean Sea, Black Sea and Azov Sea (Tortonese 1975; Relini *et al.* 1999). It is characterized by progenesis (*sensu* Gould 1977), a mechanisms that allow the retention of larval traits into adulthood as a consequence of the early achievement of sexual maturity.

Fig. 2 *Aphia minuta* (by Stefano Guerrieri from http://colapesce.xoom.it/PescItalia/pisces/Perciformi/Gobidae/Rossetto.htm).

For this reason, this species shows reduced dimensions (< 60 mm in length), lateral-compressed and transparent body with few melanophores only on the median fin and on the head, lateral eyes, reduced fins, persistence of the swim bladder and a short and straight alimentary canal (Miller 1973). Progenesis seems to be an adaptation to pelagic and planktotrophic habits. In fact, the reduction in size of *A. minuta* allows a

net energy increase by feeding on zooplankton, but implies a higher risk of predation for both adults and juveniles. To avoid the negative effects related with predation, this species has adopted a reproductive strategy characterized by an early sexual maturity and a high reproductive effort (Miller 1979). The reproductive strategy of *A. minuta* was clarified by Caputo *et al.* (2001) through the histological analysis of gonads, which highlighted the abbreviate iteroparity of this species. Indeed, individuals spawn two times during one year and after the second reproduction die. The death of all breeders is caused by an irreversible apoptotic phenomenon that affects the intestinal epithelium after the second spawning event, probably induced by an endogenous timer (Caputo *et al.* 2002).

The spawning period of this species is extremely various across different geographic regions, also within the Mediterranean Sea. For example, in the western Mediterranean Sea spawning occurs two times during a year, as result of the presence of two different annual spawning cohorts, from December to April and from June to November (Iglesias & Morales-Nin 2001). A spawning period between March and September was observed in the central Tyrrhenian Sea (Auteri *et al.* 2000), whereas the breeding season in Adriatic Sea is from April to September (La Mesa 1999; Caputo *et al.* 2001). During the one year life cycle, *A. minuta* goes through three different phases (Auteri *et al.* 2000; Iglesias & Morales-Nin 2001). After hatching from demersal eggs, larvae constitute large schools in shallow waters near the coast. This "*pelagic phase*" is followed by the "*aggregated phase*", that occur in winter and during which juveniles recruit to the bottom to form very large schools. In spring, mature adults migrate offshore and disperse on the bottom. This is the "*demersal phase*". In winter, during the "*aggregated phase*", schools are easily

detected by echo-sounders and captured using surrounding nets (Auteri *et al.* 2000; Iglesias & Morales-Nin 2001).

Within the Mediterranean basin, the fishing effort for this species is very high, because it has an important commercial value especially for Italy and Spain (Fig. 3), where *A. minuta* reaches costs from 20 to 40 € per Kg (La Mesa *et al.* 2005).

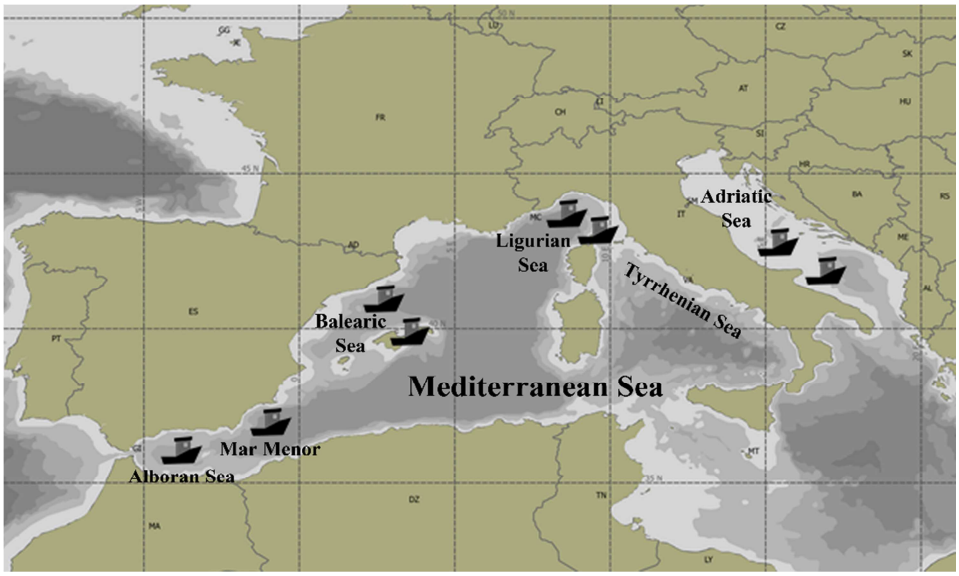


Fig. 3 Main fishing areas for *Aphia minuta* as reported by La Mesa *et al.* (2005).

The transparent goby fishery is considered "*special fishery*" because it is regulated by the directives of the Commission of the European Communities that authorize the exploitation of this resource only in specific areas in which it is abundant, and by a restricted number of vessels (La Mesa *et al.* 2005). In Spain about 70 vessels are employed for *A. minuta* fishery and total catches amount to about 100 tons per year. In Italy, total catches are about 160 tons per year and the fishing fleets is constituted by 400 vessels (La Mesa *et al.* 2005). A collapse in biomass of *A. minuta* was reported in Malaga in the 1970s and determined a crisis in local fishery. This event made clear that over-exploitation and/or environmental changes can have negative

effects on *A. minuta* survival, highlighting the necessity to plan correct management and conservation strategies that have to take into account the biological characteristics of this species, real biomasses estimated before each fishing season and the possibility of by-catch on other species, as larvae of clupeids and other small gobies (La Mesa *et al.* 2005).

The first study aimed to evaluate the population structure of the transparent goby within the Mediterranean basin was carried out as molecular marker the mtDNA (Giovannotti *et al.* 2009). Giovannotti *et al.* (2009) sampled *A. minuta* individuals from Adriatic Sea and Western Mediterranean basin and screened them for a 2,400 bp segment of the mtDNA, including the NADH dehydrogenase subunits 3, 4 and 4L, using the RFLP techniques in which all amplicons were digested with seven restriction enzymes (AluI, ApaI, AvaII, HinfI, HpaII, RsaI, TaqI). Results showed a weak genetic structure of *A. minuta* into the Mediterranean basin probably due to its dispersal capability and peculiar life cycle. Indeed, pelagic habits of this species are responsible of a high gene flow between different populations that results genetically homogeneous. In addition, the peculiar life cycle of this species made it extremely vulnerable to Pleistocene climatic changes, which brought to a severe demographic collapse with only a recent recolonization of the entire Mediterranean basin by founder effect (Giovannotti *et al.* 2009).

Engraulis encrasicolus

The European anchovy (*Engraulis encrasicolus*, Linnaeus 1758) (Fig. 4) belongs to the family Engraulide (order Perciformes). This species is widely distributed along the eastern part of the Atlantic Ocean, from Norway to South Africa, within the

Mediterranean Sea, Black Sea, Azov Sea and also in the Gulf of Suez and Suez canal (Whitehead *et al.* 1988). This species is characterized by a tapered body, a maximum length of 20 cm, with a back coloured in blue-green, and a silver belly. The large head brings big eyes and a mouth with a prominent maxilla.

Fig. 4 *Engraulis encrasicolus*.

Other peculiar features of this species are the tendency to form large monospecific groups, known as "*schooling behavior*" and the capacity to accomplish diel vertical migrations, during which individuals living at the surface during the night and return to the bottom during the day (Morello & Arneri 2009).

As for the reproduction, *E. encrasicolus* is an iteroparous species, usually spawning between April and October within the Adriatic Sea, with a peak in May-June (Vučetić 1971) and another in August-September (Regner 1972). The sexual maturity is reached at one year of age and for each reproductive season there are numerous spawning events. For this reason it is considered as a "*multiple spawner*" (Sinovčić & Zorica 2006). Together with other small pelagic species, *E. encrasicolus* has an important ecological role within the marine food web because it transfers energy from the lower trophic levels to the higher ones, feeding principally on zooplankton. In addition, through a wasp-waist control, it rules the abundance of both prey and predators (Cury *et al.* 2000; Coll *et al.* 2007). The importance of the ecological role of *E. encrasicolus* is amplified in ecosystems dominated by the pelagic compartment, as the Adriatic Sea (Coll *et al.* 2007).

The European anchovy has an important role in the Adriatic Sea, also because it is a species of great interest to fisheries together with sardine (*Sardina pilchardus*). In Italy, the fishing of small pelagics (anchovies and sardines) takes place mainly by "*lampara*" and "*volante*" vessels. "*Volante*" are midwater pelagic pair trawlers that fish during the day on schools of anchovy located near the bottom, while "*lampare*" are purse seiners that fish during the night using lights to group anchovies that are near the surface (Cingolani 1996; Falco 2007). To better manage fishing activities in this basin, the Adriatic Sea was subdivided into two "*Geographical Sub-Areas*" (GSA) (GFCM 2001), the GSA 17 includes the central and northern Adriatic, while the southern Adriatic is represented by the GSA 18 (Fig. 5).

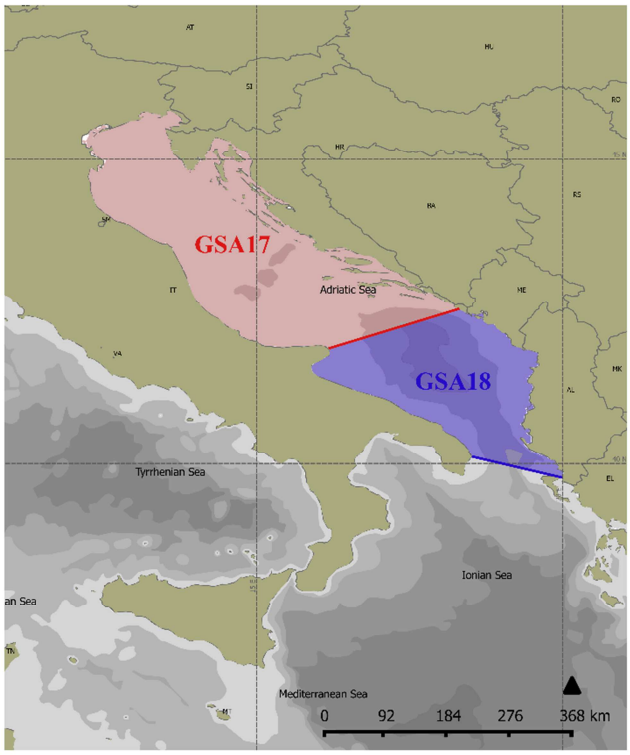


Fig. 5 Geographical Sub Areas (GSA) of the Adriatic Sea as indicated in GFCM (2001).

It seems that GSA 17 has recently been affected by over-fishing (Carpi *et al.* 2015), but it is not the first time within the Adriatic basin. In effect, biomass estimation

using different methodologies has shown a peak in biomass around 1978, with a rapid decline culminating with a demographic collapse in 1987, followed only by a partial recovery (Santojanni *et al.* 2003). These demographic fluctuations in anchovy biomasses seem affected by fishing activities. In fact, between 1978-1979 an absolute maximum in fisheries production within the Adriatic Sea can be observed (ca. 64,492 t), with a collapse in captures in the 1987 (7,055 t) (Morello & Arneri 2009). Besides fishing, other factors could adversely affect anchovy biomasses in Adriatic Sea as environmental changes, which limit the recruitment of larvae and juveniles, or an increase in predators (Regner 1996). Due to the high fishing effort on the European anchovy, the investigation of a possible stock composition seems necessary.

Before the advent of molecular techniques, one of the first studies to evaluate population structure in this species was performed on the basis of phenotypic features and the estimation of growth rates by otoliths analysis. Levi *et al.* (1994) observed that European anchovies sampled in Northern Adriatic basin were characterized by a silver colouration and a slow growth rate, while anchovies sampled in Central and Southern Adriatic Sea showed a dark colouration and a fast growth rate. These observation led to hypothesize the presence of two separate stocks within the Adriatic Sea. The first molecular markers applied to identify a stock structure in *E. encrasicolus* were allozymes. Bembo *et al.* (1996) highlighted that anchovies sampled from different geographic areas showed differences in the expression of some allozymes, and these differences were extremely clear between individuals from shallow waters (Northern and Eastern Adriatic Sea) and those of deeper waters (Central and Southern Adriatic Sea), confirming the hypothesis of two

stocks as proposed by Levi *et al.* (1994). More recent studies (Borsa 2002; Borsa *et al.* 2004), analysing morphological and genetic markers, hypothesized the presence of two different species of anchovy within the Adriatic basin. The first, *E. encrasicolus* or "*blue anchovy*" was the most common in oceanic waters and the second, *E. albidus* or "*silver anchovy*" was a species of brackish waters common in the Northern Adriatic area and in the Gulf of Lione. Magoulas *et al.* (1996, 2006), on the basis of analyses on mtDNA, subdivided *E. encrasicolus* inhabiting the Mediterranean Sea in two different haplogroups. These different mitochondrial clades were evolved geographically isolated and, at a later time entered in contact in the Mediterranean Sea. Specifically, clade A originated in the Atlantic area while clade B originated within the Adriatic Sea. In addition, Magoulas *et al.* (2006) have not observed haplotype differences between individuals from northern Adriatic area and those from other areas. These authors hypothesized that *E. albidus* is only a population belonging to the clade B and adapted to live in brackish waters.

Salmo trutta complex

The *Salmo trutta* species complex should be considered as a complex of species belonging to the genus *Salmo* (family Salmonidae, order Salmoniformes) (Kottelat & Freyhof 2007). Individuals belonging to this species complex are characterized by different phenotypes depending on their habitat, a silver colouration is typical of lacustrine and anadromous forms, while bright colourations with a different pattern of dark spots (parr marks) are typical of riverine forms (Fig. 6). In addition, a fusiform body, an adipose fin and a toothed vomer (an adaptation to feed on invertebrates and small fishes) are always observed (Kottelat & Freyhof 2007).

Fig. 6 Mediterranean trout (*Salmo trutta* complex).

Several species belonging to the *S. trutta* complex could be recognized (Kottelat & Freyhof 2007), but the taxonomy of the genus *Salmo* is extremely doubtful. Bianco (2014) clarified this issue for the Italian peninsula defining five species belonging to the *S. trutta* complex: *Salmo cettii* (Rafinesque 1810), *Salmo marmoratus* (Cuvier 1829), *Salmo farioides* (Karaman 1938), *Salmo Fibreni* (Zerunian & Gandolfi 1990) and *Salmo carpio* (Linnaeus 1758). Specifically, *S. fibreni* and *S. carpio* are lacustrine species which are endemic to Lake Posta Fibreno and Garda Lake, respectively. *S. marmoratus* is distributed in rivers and lakes of northern Italy, and could be distinguished by its marble pigmentation. *Salmo cettii* (Rafinesque 1810) is distributed in Tyrrhenian rivers of Italy, Sardinia, Corsica and Sicily, and in some cases it is erroneously confused with *Salmo macrostigma* (Dumèril 1858), a species of the genus *Salmo* distributed in North-western Africa. *Salmo farioides*, a name that initially described trouts of the Eastern Adriatic area, is currently used to define also those of all Italian Adriatic rivers (Bianco 2014). Finally, also the brown trout is widespread within the Italian peninsula (*Salmo trutta*, Linnaeus 1758). This native of the Atlantic area and was introduced by human activities all over the world, including the Mediterranean area (Kottelat & Freyhof 2007).

Whereas the taxonomy of the genus *Salmo* is hazy, the phylogeography of the brown trout is more clear thanks to several studies on mtDNA control region (D-loop) diversity. These studies highlighted the presence of six different evolutionary lineages (Bernatchez 2001; Suárez *et al.* 2001; Cortey *et al.* 2009): Atlantic (AT), Danubian (DA), Mediterranean (ME), Adriatic (AD), *marmoratus* (MA) and Duero (DU), an endemism of the Iberian Peninsula. In addition, the Atlantic lineage could be differentiated in two distinct groups (Suárez *et al.* 2001; Cortey *et al.* 2009): a North Atlantic group and a South Atlantic group, which include three distinct clades (AT3-1; AT3-2; AT3-3, Cortey *et al.* 2009) (Fig. 7). Currently, the South Atlantic group is widespread in Iberian Peninsula, Morocco (Suárez *et al.* 2001; Cortey *et al.* 2009; Snoj *et al.* 2011) and Sicily (Schöffmann *et al.* 2007; Fruciano *et al.* 2014).

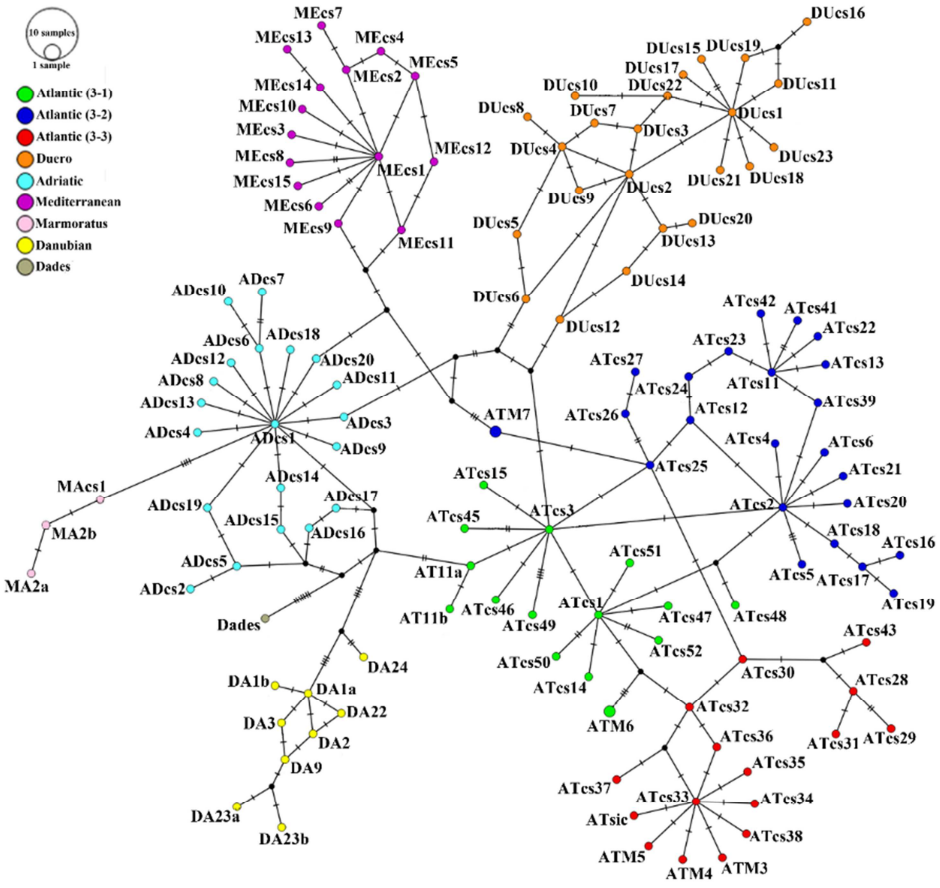


Fig. 7 Haplotype network showing the relationship between *Salmo trutta* lineages.

This genetic structure was probably designed by the major climatic and environmental changes occurred during the Pleistocene that have allowed the geographic isolation and the allopatric evolution of these lineages. Probably ca. 700,000 yrs ago, the most ancient separation led to the evolution of AT, DA and ME lineages, respectively distributed within Atlantic, Ponto-Caspian and Mediterranean areas, and almost simultaneously to the divergence of ME, AD and MA lineages within the Mediterranean basin (Bernatchez 2001). Later, the succession of glacial and interglacial cycles during the Pleistocene, led to the differentiation of Atlantic lineages in different clades (Cortey *et al.* 2009). In addition, the higher diversity observed in the southern AT group, in contrast to the lower diversity of the northern one, suggested that the northern group evolved only at the end of the Pleistocene, because during glacial periods northern AT populations decreased and migrants took refuge in southern areas. During inter-glacial phases, they move back northward, bringing only a part of their original genetic diversity (Cortey *et al.* 2009).

As mentioned above, the phenotypic plasticity is one of the most important features of *S. trutta* and within this species complex three different ecotypes should be recognized (Kottelat & Freyhof 2007). The "*anadromous trout or sea trout*" born in streams, smoltifies and migrates to the sea or to estuaries to forage, reach the sexual maturity and return to their native stream to spawn. The "*lake trout*", born in streams, smoltify and migrate to the lake to forage, then return to their native stream to spawn. Sometimes, some individuals do not perform the migration but remain in river for all their lives, and are called "*resident trout*" (Klemetsen *et al.* 2003; Kottelat & Freyhof 2007). Within the genus *Salmo*, one or more of these ecotypes are known and the developing of one or another of this life history strategies are determined by both

environmental and genetic factors (Kottelat & Freyhof 2007). Usually, the reproduction of *S. trutta* occurs in autumn-winter. In this period, females spawn their eggs on gravel bottoms of running rivers and they are rapidly fertilized by males (Klemetsen 2003). In spring larvae, called alevins, hatch and remain on the bottom, feeding on yolk residues. When the yolk is fully absorbed, individuals (fries) migrate to the open stream bed and start to predate (Klemetsen 2003). Juveniles (parrs) can mature and continue all the life cycle in fresh waters as "*resident trout*" or "*lake trout*" or, after one or several months, can go through the parr-smolt transformation (smoltification), a process that occur prior the seaward migration. The smoltification is regulated by hormones that induce morphological and physiological changes to allow the survival in salt waters, like a silver colouration and the proliferation of ion pumps in gills for secreting excess salts (McCormick 2013). After this process, "*smolts*" can migrate to the sea, where they actively feed also for several years, and at the maturation return to their native stream to spawn (Klemetsen 2003). Anadromous populations of *S. trutta* are present in all the north Atlantic area but are unknown in the entire Mediterranean basin (Elliot 1994), in which only some Atlantic individuals used in restocking activities seems to have this life strategy (Snoj *et al.* 2002). Probably, native migratory individuals are unknown in the Mediterranean Sea due to the high salinity (38‰) and temperature (ca. 25°C in summer) of this basin (Tortonese 1970).

Salmo trutta, like all salmonids species, lives in cold and well oxygenate waters, so the survival of *Salmo* species is mainly defined by temperature, because it influences the oxygen dissolved in water (Jonsson & Jonsson 2009). The upper limit of temperature for *S. trutta* survival is at 24.7°C and the lower limit is at 0°C, but the

growth ceases around 4°C. Moreover, high temperature could have effects on several phases of the *S. trutta* life cycle, such as a delayed spawning, accelerated embryo development and hatching of alevins and, for anadromous individuals, an early migration that has negative effects on post-smolts survival in sea water (Jonsson & Jonsson 2009; Elliott & Elliott 2010). An increase in temperature, due to present and future climate changes, seems to be the most alarming threat for the survival of temperature sensitive species, like salmonids. Probably, for *S. trutta* the most impacted by climate changes will be the southern populations, for which a local decline or extinction, and/or a northward shift to reach more suitable thermal habitats, are hypothesized (Jonsson & Jonsson 2009; Almodóvar *et al.* 2012). In addition, for anadromous individuals a reduction in this behaviour will be observed with an increasing temperature, because more suitable habitats for feeding and growing in rivers will make the seaward migration less favored (Jonsson & Jonsson 2009).

References

Abdul-Muneer P.M. (2014) Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. *Genetics research international*, **2014**.

Aguila R.D., Perez S.K.L., Catacutan B.J.N., Lopez G.V., Barut N.C., Santos M.D. (2015) Distinct yellowfin tuna (*Thunnus albacares*) stocks detected in Western and Central Pacific Ocean (WCPO) using DNA microsatellites. *PloS one*, **10**, e0138292.

Allendorf F.W., Luikart G., Atiken S.N. (2013) *Conservation and the Genetics of Populations*. Blackwell Publishing.

Almodóvar A., Nicola G.G., Ayllón D., Elvira B. (2012) Global warming threatens the persistence of Mediterranean brown trout. *Global Change Biology*, **18**, 1549-1560.

Arkush K.D., Giese A.R., Mendonca H.L., McBride A.M., Marty G.D., Hedrick P.W. (2002) Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 966-975.

Auteri R., Baino R., Abella A. (2000) Biology and population dynamic of the transparent goby: a locally important resource of artisanal fisheries. *Biologia Marina Mediterranea*, **7**, 144-157.

Avise J.C. (2000*) Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA: 3-108.

Begg G.A., Friedland K.D., Pearce J.B. (1999) Stock identification and its role in stock assessment and fisheries management: an overview. *Fisheries Research*, **43**, 1-8.

Begg G.A., Waldman J.R. (1999) An holistic approach to fish stock identification. *Fisheries research*, **43**, 35-44.

Bembo D.G., Carvalho G.R., Cingolani N., Arneri E., Giannetti G., Pitcher T.J. (1996) Allozymic and morphometric evidence for two stocks of the European anchovy *Engraulis encrasicolus* in Adriatic waters. *Marine Biology*, **126**, 529–538.

Bernatchez L. (2001) The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution*, **55**, 351–379.

Bianco P.G. (2014) An update on the status of native and exotic freshwater fishes of Italy. *Journal of Applied Ichthyology*, **30**, 62-77.

Booke H.E. (1981) The conundrum of the stock concept-are nature and nurture definable in fishery science? *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1479-1480.

Borsa P. (2002) Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). *Biological Journal of the Linnean Society*, **75**, 261–269.

Borsa P., Collet A., Durand J.D. (2004) Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean, *Comptes Rendus Biologies*, **327**, 1113–1123.

Brown J.R., Bechenbach A.T., Smith M.J. (1993) Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). *Molecular Biology Evolution*, **10**, 326-341.

Brown W.M. (1985) The mitochondrial genome of animals. In: MacIntyre R.J. (Ed), *Molecular evolutionary genetics*. Plenum, New York: 95-130.

Cadrin S.X., Kerr L.A., Mariani S. (2014) *Stock Identification Methods, Second ed.* Academic Press, San Diego.

Caputo V., Candi G., Arneri E., La Mesa M., Cinti C., Provinciali M., Nisi Cerioni P., Gregorini A. (2002) Short lifespan and apoptosis in *Aphia minuta*. *Journal of Fish Biology*, **60**, 775–779.

Caputo V., Candi G., La Mesa M., Arneri E. (2001) Pattern of gonad maturation and the question of semelparity in the paedomorphic goby *Aphia minuta*. *Journal of Fish Biology*, **58**, 656–669.

Carpi P., Santojanni A., Donato F., Colella S., Čikeš Keč V., Zorica B. *et al.* (2015) A joint stock assessment for the anchovy stock of the northern and central Adriatic Sea: comparison of two catch-at-age models. *Scientia Marina*, **79**, 57–70.

Carvalho G.R., Hauser L. (1994) Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries*, **4**, 326–350

Catalano S.R., Whittington I.D., Donnelan S.C., Gillanders B.M. (2014) Parasites as biological tags to assess host population structure: Guidelines, recent genetic advances and comments on a holistic approach. *International Journal for Parasitology: Parasites and Wildlife*, **3**, 220-226.

Charlesworth D., Willis J.H. (2009) The genetics of inbreeding depression. *Nature Reviews Genetics*, **10**, 783-796.

Cingolani N., Giannetti G., Arneri E. (1996) Anchovy fisheries in the Adriatic Sea. *Scientia Marina*, **60**, 269-277.

Coll M., Santojanni A., Palomera I., Tudela S., Arneri E. (2007) An ecological model of the Northern and Central Adriatic Sea: analysis of ecosystem structure and fishing impacts. *Journal of Marine Systems*, **67**, 119-154.

Cooper A., Lalueza-Fox C., Anderson S., Rambaut A., Austin J., Ward R. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, **409**, 704-707.

Cooper A., Poinar H.N. (2000) Ancient DNA: do it right or not at all. *Science*, **289**, 1139.

Cortey M., Vera M., Pla C., García-Marín J.L. (2009) Northern and Southern expansions of Atlantic brown trout (*Salmo trutta*) populations during the Pleistocene. *Biological Journal of the Linnean Society*, **97**, 904–917.

Cuéllar-Pinzón J., Presa P., Hawkins S.J., Pita A. (2016) Genetic markers in marine fisheries: Types, tasks and trends. *Fisheries Research*, **173**, 194-205.

Cury P., Bakun A., Crawford R.J., Jarre A., Quinones R.A., Shannon L.J., Verheye H.M. (2000) Small pelagics in upwelling systems: patterns of interaction and structural changes in "wasp-waist" ecosystems. *ICES Journal of Marine Science*, **57**, 603-618.

Dahl K. (1909) The problem of sea fish hatching. Special Part B, No. 5. In: Hjort J. (Ed), *Rappurt sur les Travaux de Commision A, Dans La Période 1902-1907, Conseil Permanent International Pour L'Exploration de La Mer. Rapports et Procès-Verbaux, Vol. X.*

Edwards A.L., Civitello A., Hammond H.A., Caskey C.T. (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American journal of Human Genetics*, **49**, 746-756.

Eglinton G., Logan G.A., Ambler R.P., Boon J.J., Perizonius W.R.K. (1991) Molecular preservation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **333**, 315-328.

Elliot J.M. (1994) *Quantitative ecology and the brown trout*. Oxford University Press, Oxford.

Elliott J.M., Elliott J.A. (2010) Temperature requirements of Atlantic salmon *Salmo salar*, brown trout *Salmo trutta* and Arctic charr *Salvelinus alpinus*: predicting the effects of climate change. *Journal of fish biology*, **77**, 1793-1817.

Falco P., Belardinelli A., Santojanni A., Cingolani N., Russo A., Arneri E. (2007) An observing system for the collection of fishery and oceanographic data. *Ocean Science*, **3**, 189-203.

FAO (2016) *The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all*. Rome: 200.

Ficke A.D., Myrick C.A., Hansen L.J. (2007) Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries*, **17**, 581–613.

Frankham R. (1999) Quantitative genetics in conservation biology. *Genetics Research*, **74**, 237-244.

Fruciano C., Pappalardo A.M., Tigano C., Ferrito V. (2014) Phylogeographical relationships of Sicilian brown trout and the effects of genetic introgression on morphospace occupation. *Biological Journal of the Linnean Society*, **112**, 387–398.

GFCM (2001) Report of the twenty-six session. Lacco Ameno, Ischia, Italy, 10-13 September 2001. GFCM Report. No 26. FAO, Rome: 27.

Giovannotti M., La Mesa M., Caputo V. (2009) Life style and genetic variation in teleosts: the case of pelagic (*Aphia minuta*) and benthic (*Gobius niger*) gobies (Perciformes: Gobiidae). *Marine Biology*, **156**, 239–252.

Gould S.J. (1977) *Ontogeny and Phylogeny*. Harvard University Press, Harvard, MA: 303-350.

Hall D.A. (2014) Conventional and radiofrequency identification (RFID) Tags. In: Cadrin S.X., Kerr L.A., Mariani S. (Eds), *Stock Identification Methods, Second ed.* Academic Press, San Diego: 365-395.

Hansen L.P., Jacobsen J.A. (2003) Origin and migration of wild and escaped farmed Atlantic salmon, *Salmo salar* L., in oceanic areas north of the Faroe Islands. *ICES Journal of Marine Science*, **60**, 110-119.

Hansen M.M. (2002) Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: An approach using microsatellite DNA analysis of historical and contemporary samples. *Molecular Ecology*, **11**, 1003-1015.

Hansen M.M., Bekkevold D., Jensen L.F., Mensberg K.L.D., Nielsen E.E. (2006) Genetic restoration of a stocked brown trout *Salmo trutta* population using microsatellite DNA analysis of historical and contemporary samples. *Journal of Applied Ecology*, **43**, 669-679.

Hardy G.H. (1908) Mendelian proportions in a mixed population. *Molecular and General Genetics MGG*, **1**, 395-395.

Hauser L., Adcock G.J., Smith P.J., Ramírez J.H.B., Carvalho G.R. (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences*, **99**, 11742-11747.

Hauser L., Carvalho G.R. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333-362.

Heath D.D., Busch C., Kelly J., Atagi D.Y. (2002) Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). *Molecular Ecology*, **11**, 197-214.

Higuchi R., Bowman B., Freiberger M., Ryder O.A., Wilson A.C. (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature*, **312**, 281-284.

Hoarau G., Boon E., Jongma D.N., Ferber S., Palsson J., Van der Veer H.W. *et al.* (2005). Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (*Pleuronectes platessa* L.). *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 497-503.

Höss M., Dilling A., Currant A., Pääbo S. (1996) Molecular phylogeny of the extinct ground sloth *Mylodon darwinii. Proceedings of the National Academy of Sciences*, **93**, 181-185.

Hutchinson W.F., Carvalho G.R., Rogers S.I. (2001) Marked genetic structuring in localised spawning populations of cod *Gadus morhua* in the North Sea and adjoining waters, as revealed by microsatellites. *Marine Ecology Progress Series*, **223**, 251-260.

Iglesias M., Morales-Nin B. (2001) Life cycle of the pelagic goby *Aphia minuta* (Pisces: Gobiidae). *Scientia Marina*, **65**, 183–192.

Ihssen P.E., Booke H.E., Casselmann J.M., McGlade J.M., Paine N.R., Utter F.M. (1981) Stock Identification: Materials and Methods. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1838-1855.

Jaenicke-Despres V., Buckler E.S., Smith B.D., Gilbert M.T.P., Cooper A., Doebley J., Pääbo S. (2003) Early allelic selection in maize as revealed by ancient DNA. *Science*, **302**, 1206-1208.

Jónsdóttir I.G., Campana S.E., Marteinsdottir G. (2006) Stock structure of Icelanding cod *Gadus morhua* L. based on otoliths chemistry. *Journal of Fish Biology*, **69**, 136-150.

Jonsson B., Jonsson N. (2009) A review of the likely effects of climate change on anadromous Atlantic salmon *Salmo salar* and Brown trout *Salmo trutta*, with particular reference to water temperature and flow. *Journal of Fish Biology*, **75**, 2381-2447.

Jørgensen H.B., Hansen M.M., Bekkevold D., Ruzzante D.E., Loeschcke V. (2005) Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. *Molecular Ecology*, **14**, 3219-3234.

Kapuscinski A.R., Miller L.M. (2007) *Genetic guidelines for fisheries management, Second ed.* Duluth, Minnesota: Minnesota Sea Grant, University of Minnesota.

Kerr L.A., Campana S.E. (2014) Chemical composition of fish hard parts as a natural marker of fish stocks. In: Cadrin S.X., Kerr L.A., Mariani S. (Eds), *Stock Identification Methods, Second ed.* Academic Press, San Diego: 205-234.

Klemetsen A., Amundsen P.A., Dempson J.B., Jonsson B., Jonsson N., O'connell M.F., Mortensen E. (2003) Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. *Ecology of freshwater fish*, **12**, 1-59.

Kottelat M., Freyhof J. (2007) *Handbook of European freshwater fishes*. Publications Kottelat.

La Mesa M. (1999) Age and growth of *Aphia minuta* (Pisces, Gobiidae) from the central Adriatic Sea. *Scientia Marina*, **63**, 147-155.

La Mesa M., Arneri E., Caputo V., Iglesias M. (2005) The transparent goby, *Aphia minuta*: review of biology and fisheries of a paedomorphic European fish. *Review in Fish Biology and Fisheries*, **15**, 89–109.

Larsen P.F., Hansen M.M., Nielsen E.E., Jensen L.F., Loeschcke V. (2005) Stocking impact and temporal stability of genetic composition in a brackish northern pike population (*Esox lucius* L.), assessed using microsatellite DNA analysis of historical and contemporary samples. *Heredity*, **95**, 136-143.

Leonard J.A. (2008) Ancient DNA applications for wildlife conservation. *Molecular Ecology*, **17**, 4186-4196.

Levi D., Andreoli M.G., Arneri E., Giannetti G., Rizzo P. (1994) Otolith reading as a tool for stock identification. *Fisheries Research*, **20**, 97-107.

Lindahl T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709-715.

MacKenzie K., Campbell N., Mattiucci S., Ramos P., Pinto A.L., Abaunza P. (2008) Parasites as biological tags for stock identification of Atlantic horse mackerel *Trachurus trachurus* L. *Fisheries Research*, **89**, 136-145.

Magoulas A., Castilho R., Caetano S., Marcato S., Patarnello T. (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Molecular phylogenetics and evolution*, **39**, 734-746.

Magoulas A., Tsimenides N., Zouros E. (1996) Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Molecular Biology and Evolution*, **13**, 178-190.

McCormick S.D. (2013) Smolt physiology and endocrinology. *Fish physiology*, **32**, 199-251.

McGinnity P., Prodöhl P., Ferguson A., Hynes R., Ó Maoiléidigh N., Baker N. *et al.* (2003) Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 2443-2450.

Meyer A. (1993) Evolution of the mitochondrial DNA in fishes. In: Mochachka P.W., Mommsen T.P. (Eds), *Biochemistry and molecular biology of fishes*. Elsevier, Press Amsterdam, New York: 1-38.

Miller P.J. (1973) The species of Pseudaphya (Teleostei: Gobiidae) and the evolution of aphyiine gobies. *Journal of Fish Biology*, **5**, 353–365.

Miller P.J. (1979) Adaptiveness and implications of small size in teleosts. *Symposium of the Zoological Society of London*, **44**, 263–306.

Morello E.B., Arneri E. (2009) Anchovy and sardine in the Adriatic Sea-an ecological review. *Oceanography and Marine Biology: An Annual Review*, **47**, 209–256.

Mork J., Ryman N., Ståhl G., Utter F., Sundnes G. (1985) Genetic variation in Atlantic cod (*Gadus morhua*) throughout its range. *Canadian Journal of Fisheries and Aquatic Sciences*, **42**, 1580-1587.

Mullis K.B., Faloona F.A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in enzymology*, **155**, 335-350.

Muscolino M.S. (2012) Fishing and Whaling. In: McNeill J.R., Mauldin E.S. (Eds), *A Companion to Global Environmental History*. Chichester: 279-296.

Nielsen E.E., Hansen M.M. (2008) Waking the dead: the value of population genetic analyses of historical samples. *Fish and Fisheries*, **9**, 450-461.

Nielsen E.E., Hansen M.M., Loeschcke V. (1997) Analysis of microsatellite DNA from old scale samples of Atlantic salmon *Salmo salar*: a comparison of genetic composition over 60 years. *Molecular Ecology*, **6**, 487-492.

O'Connor S., Ono R., Clarkson C. (2011) Pelagic fishing at 42,000 years before the present and the maritime skills of modern humans. *Science*, **334.6059**, 1117-1121.

Pääbo S. (1985) Molecular cloning of Ancient Egyptian mummy DNA. *Nature*, **314**, 644-645.

Park L.K., Moran P. (1994) Developments in molecular genetic techniques in fisheries. *Reviews in Fish Biology and Fisheries*, **4**, 272-299.

Patarnello T., Volckaert F.A.M., Castilho R. (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, **16**, 4426–4444.

Pauls S.U., Nowak C., Bálint M., Pfenninger M. (2013) The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, **22**, 925-946.

Pinsky M.L., Byler D. (2015) Fishing, fast growth and climate variability increase the risk of collapse. *Proceedings of the Royal Society of London B: Biological Sciences*, **282**, 20151053.

Poinar H.N., Hofreiter M., Spaulding W.G., Martin P.S., Stankiewicz B.A., Bland H. *et al.* (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science*, **281**, 402-406.

Poinar H.N., Kuch M., Sobolik K.D., Barnes I., Stankiewicz A.B., Kuder T. *et al.* (2001) A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences*, **98**, 4317-4322.

Putman A.I., Carbone I. (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, **4**, 4399-4428.

Ratnasingham S., Hebert P.D. (2007) BOLD: The Barcode of Life Data System (http://www. barcodinglife.org). *Molecular ecology notes*, **7**, 355-364.

Regner S. (1972) Contribution to the study of the ecology of the planktonic phase in the life history of the anchovy in the central Adriatic. *Acta Adriatica*, **14**, 3-43.

Regner S. (1996) Effects of environmental changes on early stages and reproduction of anchovy in the Adriatic Sea. *Scientia Marina*, **60**, 167-177.

Reid A.H., Fanning T.G., Hultin J.V., Taubenberger J.K. (1999) Origin and evolution of the 1918 “Spanish” influenza virus hemagglutinin gene. *Proceedings of the National Academy of Sciences*, **96**, 1651-1656.

Reiss H., Hoarau G., Dickey-Collas M., Wolff W.J. (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, **10**, 361-395.

Relini G., Bertrand J., Zamboni A. (1999) Sintesi delle conoscenze sulle risorse da pesca dei fondi del Mediterraneo centrale (Italia e Corsica). *Biologia Marina Mediterranea*, **6**, 1– 868.

Ruggeri P., Splendiani A., Bonanomi S., Arneri E., Cingolani N., Santojanni A. *et al.* (2013) Searching for a stock structure in *Sardina pilchardus* from the Adriatic and Ionian seas using a microsatellite DNA-based approach. *Scientia Marina*, **77**, 565-574.

Ruggeri P., Splendiani A., Bonanomi S., Arneri E., Cingolani N., Santojanni A. *et al.* (2012) Temporal genetic variation as revealed by a microsatellite analysis of European sardine (*Sardina pilchardus*) archived samples. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**, 1698-1709.

Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A., Arnheim N. (1985) Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**, 1350-1354.

Sala O.E., Chapin F., Stuart A., Juan J., Berlow E., Bloomfield J. *et al.* (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

Salo W.L., Aufderheide A.C., Buikstra J., Holcomb T.A. (1994) Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *Proceedings of the National Academy of Sciences*, **91**, 2091-2094.

Santojanni A., Arneri E., Barry C., Belardinelli A., Cingolani N., Giannetti G. *et al.* (2003) Trends of anchovy (*Engraulis encrasicolus*, L.) biomass in the northern and central Adriatic Sea. *Scientia Marina*, **67**, 327–340.

Schöffmann J., Sušnik S., Snoj A. (2007) Phylogenetic origin of *Salmo trutta* L 1758 from Sicily, based on mitochondrial and nuclear DNA analyses. *Hydrobiologia*, **575**, 51-55.

Selkoe K.A., Toonen R.J. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters*, **9**, 615-629.

Sinovčić G., Zorica B. (2006) Reproductive cycle and minimal length at sexual maturity of *Engraulis encrasicolus* (L.) in the Zramaja River estuary (Adriatic Sea, Croatia). *Estuarine Coastal and Shelf Science*, **69**, 439-448.

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457-462.

Smith P.J., Birley A.J., Jamieson A., Bishop C.A. (1989) Mitochondrial DNA in the Atlantic cod, *Gadus morhua*: lack of genetic divergence between eastern and western populations. *Journal of Fish Biology*, **34**, 369-373.

Snoj A., Marčeta B., Sušnik S., Melkič E., Meglič V., Dovč P. (2002) The taxonomic status of the "sea trout" from the north Adriatic Sea, as revealed by mitochondrial and nuclear DNA analysis. *Journal of Biogeography*, **29**, 1179–1185.

Snoj A., Marić S., Bajec Sušnik S., Berrebi P., Janjani S., Schöffmann J. (2011) Phylogeographic structure and demographic patterns of brown trout in North-West Africa. *Molecular Phylogenetics and Evolution*, **61**, 203–211.

Sønstebø J.H., Borgstrøm R., Heun M. (2008) High genetic introgression in alpine brown trout (*Salmo trutta* L.) populations from Hardangervidda, Norway. *Ecology of Freshwater Fish*, **17**, 174-183.

Suárez J., Bautista J.M., Almodóvar A., Machordom A. (2001) Evolution of the mitochondrial control region in Palaearctic brown trout (*Salmo trutta*) populations: the biogeographical role of the Iberian Peninsula. *Heredity*, **87**, 198–206.

Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids research*, **17**, 6463-6471.

Tortonese B. (1975) *Osteichthyes (Pesci ossei). Parte seconda: Fauna D'italia XI.* Edizioni Calderini, Bologna: 636.

Tortonese E. (1970) *Osteichthyes (Pesci ossei). Parte prima: Fauna D'italia, Vol. 10.* Edizioni Calderini, Bologna: 1-565.

Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P. *et al.* (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature*, **410**, 1088-1091.

Vučetić T. (1971) Fluctuations a long term du meroplancton dans l'Adriatique centrale; oeufs de *Sardina pilcardus*, Walb., d'*Engraulis encrasicolus*, L. et larves de different poissons. *Archivi di Oveanologia e Limnologia*, **17**, 141-156.

Waldman J.R. (2005) Definition of stocks: an evolving concept. In: Cadrin S.X., Friedland K.D., Waldman J.R. (Eds), *Stock Identification Methods–Applications in Fishery Science*. Elsevier Academic Press, Burlington, MA: 7-16.

Wandeler P., Hoeck P.E.A., Keller L.F. (2007) Back to the Future: museum specimens in population genetics. *Trends in Ecology & Evolution*, **22**, 634-642.

Wang S., Hard J.J., Utter F. (2002) Salmonid inbreeding: a review. *Reviews in Fish Biology and Fisheries*, **11**, 301-319.

Waples R.S. (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438–450.

Weinberg W. (1908) Über den Nachweis der Vererbung beim Menschen. Jahresh.Verein f. vaterl. Naturk. In: Wurtemmberg, **64**, 368-382.

Whitehead P.J.P., Nelson G.J., Wongratana T. (1988) FAO species catalogue. Vol 7. Clupeoid fishes of the world (Suborder Clupeoidei). An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shades, anchovies and wolf-herrings. Part 2-Engraulidae. *FAO fisheries synopsis*, **125**, 305-579.

Willerslev E., Cooper A. (2005) Review paper. Ancient DNA. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 3-16.

Willerslev E., Hansen A.J., Binladen J., Brand T.B., Gilbert M.T.P., Shapiro B. *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791-795.

Willerslev E., Hansen A.J., Poinar H.N. (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution*, **19**, 141-147.

Xenopoulos M.A., Lodge D.M., Alcamo J., Märker M., Schulze K., Van Vuuren D.P. (2005) Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biol*ogy, **11**, 1557–1564.

Zane L., Bargelloni L., Patarnello T. (2002) Strategies for microsatellite isolation: a review. *Molecular ecology*, **11**, 1-16.

CHAPTER I

Evidences for the role of demography, selective pressure and hydrographic boundaries in shaping the genetic structure of the transparent goby, *Aphia minuta*, Risso 1820.

Abstract

The transparent goby (*Aphia minuta*) is a small progenetic goby that displays a particular life cycle and represents a valuable fishery resource harvested by artisanal fleets in most part of its distributional range. This study was aimed to define genetic variability in *A. minuta* on 5 sampling sites within the Mediterranean Sea and one in the Atlantic Ocean through the analysis of 11 nuclear microsatellite loci. Results revealed lower levels at several genetic diversity estimators (H_E, H_O, N_A, R_S) in the Atlantic Ocean rather than in the Mediterranean Sea suggesting the role of past or current demographic events to shape this pattern. The genetic structure was investigated using both classical genetic differentiation descriptors (θ_{ST}) and Bayesian approaches (STRUCTURE and GENELAND), and defining the current (GENECLASS) and the past migration rates (MIGRATE). The results obtained revealed high genetic structure within the Mediterranean Sea and suggested very low current migration rate. The pattern of historical migration suggested the possible role of hydrographic boundaries to shape genetic structure identified in this species. In addition, the detection of loci under selection, within those analyzed, suggested the possible implication of selective pressures connected with the peculiar life cycle of this gobiid fish. The high genetic subdivision identified with nuclear markers underlined possible implications in fishery management for the transparent goby in the Mediterranean area.

Introduction

Several evolutionary approaches have demonstrated that marine organisms show populations that are genetically less structured than the terrestrial or freshwater ones (see DeWoody & Avise 2000). This fact has for longstanding stoked the conception that marine environment was a free-barrier container where the mixing between individuals of a species was easily allowed (Ward *et al.* 1994, Avise 2000, DeWoody & Avise 2000). Advances in population biology and ecology have proved that life cycle, environmental features and historical events represent the main factors that shape population connectivity and genetic structure in marine species. Life history traits, like pelagic/demersal spawning behavior, vagility, and duration of larval phase, have been proved to be correlated to the genetic structure of several marine fish species (Riginos *et al.* 2011, Faurby & Barber 2012). Similarly, both selective pressure, caused by environmental characteristics, and hydrographical boundaries can act in transitional areas between biogeographic breaks and could be the main reason for the genetic patchiness observed in many fish populations (Avise 2000, Patarnello *et al.* 2007). Finally, historical events are responsible for local extinction and recolonization. Palaeoclimatic events played an important role to drive population dispersal and the effects produced by these phenomena are reflected by the current species distribution (Avise 2000, Wilson & Veraguth 2010). The Pleistocene climatic fluctuations determined, in the European area, the occurrence of repeated separation and secondary contact in several fish species (Magoulas *et al.* 1996, Viñas *et al.* 2004).

The gobiid fish represents the family with more marine species than any other fish family (Nelson 2006). Their high level of diversification could be connected with the evolutionary radiation occurred presumably after the transitional stage that led these

organisms from freshwater to marine habitat (Thacker 2009). This group of vertebrate has mostly a benthic life style, mainly inhabits the marine near-shore area (Nelson 2006) and is represented by both specialized (i.e. shrimp-goby mutualism, coral-dwelling gobies) (Thacker *et al.* 2011, Duchene *et al.* 2013) and highly adaptive forms (i.e. invasive species) (Stepien & Tumeo 2006). A gobiid species that shows a life style far different from benthic behavior is the transparent goby, *Aphia minuta* (Risso 1810). This is a neritic and nektonic goby occurring from the north-eastern Atlantic to the Mediterranean and Black Sea regions (Miller 1986). As several other species (i.e. genera *Crystallogobius* and *Pseudaphya*) adapted to a semipelagic life style, *A. minuta* is a paedomorphic fish that shows the persistence into adulthood of larval behavior and anatomic features (Miller 1973). It was proposed that the paedomorphic condition is probably of adaptive significance for a planktotrophic life in water column (Miller 1979). In addition to the retention of larval features, this species is also characterized by a short life span (usually less than one year) with a rapid achievement of sexual maturity. This kind of life cycle can be considered as an example of progenesis (sensu Gould 1977) and is associated to abbreviate iteroparity, with one or two spawning events for each generation (Caputo *et al.* 2001, Caputo *et al.* 2003), followed by the death of adults, due to apoptosis of intestinal epithelium (see Caputo *et al.* 2002).

Despite gobiid fish seems to show a high tendency toward quite structured populations in relation with their life history features (such as a relatively limited dispersal ability), it is not always possible to found a clear pattern in these species. In fact, studies concerning the genetic variability in gobies showed both high genetically structured species and cases with low levels of genetic differentiation in spite of dispersal ability (Pampoulie *et al.* 2004, Berrebi *et al.* 2005, Earl *et al.* 2010,

Kokita *et al.* 2013). To the best of our knowledge, the genetic variability of *A. minuta* in the Mediterranean basin was previously explored only by Giovannotti *et al.* (2009), who demonstrated a weak genetic structure within the Mediterranean area as a result of i) potential dispersal ability, suspected to increase the gene flow between populations, and ii) the effects of palaeoclimatic events occurred during the Pleistocene on this almost semelparous and short-living goby. In addition, differently from other gobies, that rarely have a significant commercial value, *A. minuta* is a target species for fishery and is highly appreciated in the Mediterranean region. The fishery of the transparent goby is considered by the Commission of the European Communities as a ‘‘special fishery’’ (fishing activities managed locally under derogation of European rules) and involves mainly artisanal fishing fleets from Spain and Italy (see La Mesa *et al.* 2005 for a review). The intense fishing effort applied to this goby, in association with its particular life cycle (pelagic life style, paedomorphosis and a short life span), pose the need for a better knowledge of the genetic composition of its populations throughout the Mediterranean area.

In order to provide information useful to improve the knowledge about the factors driving the genetic structure of marine fishes in relation to their life history traits and specifically the management of this fishery resource, a nuclear molecular marker (microsatellite DNA) was used to investigate the genetic variability of *A. minuta* from the Mediterranean Sea and the Atlantic Ocean.

Materials and Methods

Sample collection, DNA extraction, PCR amplification 82 and visualization on gel

A total of 310 adults and sub-adults individuals of transparent goby were collected from the Adriatic (AN, CR and OR), Tyrrhenian (LI) and Western Mediterranean Seas (PM) and from the Atlantic Ocean (GU) (Table 1). Whole animals were preserved in absolute ethanol at -20°C until DNA extraction. The DNA was extracted using standard phenol-chloroform procedures described in Taggart *et al.* (1992). All samples were screened at eleven microsatellite loci described in Ruggeri *et al.* (2013) (Table 2). PCR conditions for microsatellite loci typed were optimized as in Ruggeri *et al.* (2013) and samples were run as described in Ruggeri *et al.* (2012).

Data analysis and statistical treatment of data

Standard genetic diversity estimates were obtained from genotype and allele frequencies of the microsatellite loci typed. The program MICROCHECKER 2.2.1 (Van Oosterhout *et al.* 2004) was used to assess the presence of null alleles and other genotyping errors (allele dropout and stutter peaks). In addition, the null allele frequencies were estimated per locus and per sampling locality using the algorithm of Dempster *et al.* (1977) available in FreeNa (Chapuis & Estoup 2007). Number of alleles observed for each locus (N_A), allelic richness (R_S), observed (H_O) and expected (H_E) heterozygosities were calculated for each population using FSTAT 2.9.3 (Goudet 2001).

The inbreeding coefficient, F_{IS} (Weir & Cockerham 1984) was used to evaluate the Hardy-Weinberg genotypic equilibrium (HW) in each sampling location. HW was assessed and tested with the exact test implemented in Genepop v.4.0.10 (Rousset

2008) using a Markov chain method of 100 batches of 10000 iterations each, with the first 1000 iterations discarded before sampling (Guo & Thompson 1992). Genepop was also applied to test for linkage disequilibrium between loci. The tests were conducted by means of 1000 batches of 2000 iterations each of Markov chain method (MCMC). P-values from multiple comparisons, like those derived from HW equilibrium and linkage disequilibrium tests, were corrected using a sequential Bonferroni correction (Rice 1989) with the aim to obtain more reliable results.

The presence of loci potentially affected by selection was investigated with LOSITAN (Antao *et al.* 2008). The simulations were carried out using 50000 iterations and taking into account both Infinite Allele Mutation model (IAM) and Stepwise Mutation Model (SMM), since the mutational model involved in the microsatellite set of loci here used was not known. Furthermore, analyses were carried out both including and excluding the Atlantic Ocean sample (GU) to verify the consistency or the role of loci identified as outliers. In addition, the sequences of loci under selective pressures were aligned on nucleotide BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The BLAST search was aimed to find the best match between sequence of loci under selection and *Maylandia zebra* and *Danio rerio* reference genomes described in GenBank.

Levels of genetic divergence between pairs of samples were estimated based on θ_{ST} (Weir & Cockerham 1984), an analogue of Wright's F_{ST}, using FSTAT, and compared with those obtained applying the Exclusion of Null Allele Method (ENA) as implemented in FreeNa.

The genetic structure was examined using two distinct Bayesian approaches with the aim to detect the number of clusters (K) of *A. minuta* at the HW equilibrium within the geographical area investigated. The first approach was carried out with the

Bayesian software GENELAND (Guillot *et al.* 2005a). This software incorporates geographical information (represented by the geographic coordinates of sampling sites) and produces accurate inference of population structure in marine species and can help to define the spatial distribution between the inferred clusters. The test was aimed to identify the most probable K from one to seven clusters and was performed by means of 10 independent runs with 200000 MCMC iterations, discarding 10000 iterations. An additional run of 200000 MCMC iterations, with the Poisson processes equal to the number of sampled individuals, was carried out with K fixed at the most probable number of cluster identified, as suggested in Guillot *et al.* (2005b) in order to estimate the other parameters.

The results obtained with GENELAND were compared with a more conservative approach using the program STRUCTURE v.2.3.2.1 (Pritchard *et al.* 2000, Falush *et al.* 2003). STRUCTURE has proved to be less performant in estimating the number of clusters when the F_{ST} levels are low (Latch *et al.* 2006). This is a typical condition for many marine fishes. Therefore, the results obtained with STRUCTURE were used as a baseline for the number of possible clusters exists in the geographic area investigated. The program was set assuming K from one to six and selecting an admixture model with correlated allele frequencies. Each K was performed with ten independent runs of a Monte Carlo Markov Chain of 106 iterations after a 105 131 burn-in replications period. Prior information about the basin to which each sample come from were provided and used in STRUCTURE simulations applying the logPrior mode. The Evanno *et al.* (2005) method, implemented in the program STRUCTURE HARVESTER (Earl & vonHoldt 2012), was applied to determine the most likely K value. On first instance the simulations with both GENELAND and

STRUCTURE were conducted using all the loci analyzed and the results were then compared with additional simulations carried out on the basis of only neutral loci.

The historical and contemporary migration rate among sampling locations was investigated using two different approaches on neutral loci data to avoid the bias of selection on estimates. The contemporary gene flow was evaluated applying a self-assignment test implemented in GENECLASS 2.0 (Piry *et al.* 2004). The detection of the first generation migrants, in each locality analyzed, was explored applying the criterion of Rannala & Mountain (1997) using 10000 simulated individuals on the basis of Paetkau *et al.* (2004) algorithm. The historical level of gene flow was estimated using a Coalescent-based approach implemented in MIGRATE-n v.3.5.1 program (http://popgen.sc.fsu.edu/Migrate/Migrate-n.html). As suggested in Beerli (2009) a Bayesian method was applied (Beerli 2006) and F_{ST} estimates (θ_{ST}) among localities were used as initial parameters for calculation of Θ (= $4N_e\mu$) and M (M= m/μ, where m is the immigration rate per generation and μ is the mutation rate). A Brownian motion model was used and mutation was considered constant over all loci. MCMC consisted of 1 long chain with 500000 recorded genealogies for each locus, with a burn-in of 10000 genealogies discarded.

The geographic areas associated with genetic discontinuities, that could promote a limitation in gene flow between localities analyzed, were identified with BARRIER v.2.2 (Manni *et al.* 2004). The program uses Monmonier's algorithm to identify the boundaries associated with the highest genetic heterogeneity and graphically plot them on a map where the samples are represented on the basis of their geographical coordinates. The reliability and the strength of each boundary identified were evaluated using a bootstrapping procedure over each of the 11 nuclear loci separately.

A Mantel test (using ISOLDE in Genepop) was applied to test the relationship between genetic differentiation and geographic distance (isolation by distance). The linear shortest sea-distance expressed in kilometers was regressed against the genetic differentiation estimates obtained with both ρRST and θST estimators and the significance of correlation between geographic and genetic matrices was obtained with a permutation test of 10000 iterations. An extended version of the POWSIM software (Ryman & Palm 2006) was used to compare the statistical power of the dataset for detection of population differentiation when all loci analyzed or only the neutral loci were taken into account. Simulations were carried out between the six samples using default parameters (1000 dememorizations, 100 batch and 1000 iterations per batch) and several combinations of population divergence time (t) and effective population size (*Ne*) [*Ne*/t: 1000/10, 1000/20; 5000/5, 5000/10, 5000/20; 10000/2, 10000/5, a simulation scenario with no divergence among samples (1000/0) was also tested]. Each Ne/t tested were simulated with 1000 replicates.

The program BOTTLENECK 1.2.02 (Piry *et al.* 1999) was applied to detect genetic signatures of changes in population size. The analysis was carried out with 10000 simulations under the stepwise mutation model (SMM) and the two-phase model (TPM) with 95% single step mutations and 5% multi-step mutations as recommended by Piry *et al.* (1999). P-values from the Wilcoxon's test, assessed for significance at the 0.05 level, and the distribution of allele frequencies (Mode shift Test) were used as evidence for past change in population size.

Results

PCR success and genetic variability within samples

The PCR success was high for the eleven microsatellite loci genotyped, and the missing data accounted for 2.81% of the global dataset. MICROCHECKER test showed a lack of PCR artifacts (allele dropout and stuttering) in all the microsatellite loci analyzed. However, signals for null alleles were identified at 14 out of 66 global tests and involved mainly loci Am_AGAT12, Am_AAT3. Null allele frequencies ranged from 0.051 (Am_AC1) to 0.113 (Am_AC34). When the Brookfield algorithm (Brookfield 1996) was applied to correct the dataset for null allele frequencies, a qualitative improvement of the results was achieved, since only 4 out of 66 tests remained still significant. Therefore, in subsequent analyses the corrected dataset was used. The microsatellite polymorphism was moderate to high, varying from 4 alleles at Am_AC22 locus to 49 alleles at Am_AC1 locus (Table 2). The mean number of alleles (N_A) per sample (over all loci) varied from 8.55 alleles in GU to 12.55 alleles in LI (average of 11.23±1.43). The allelic richness (R_S) showed values quite similar to N_A varying from 8.07 (GU) to 11.66 (LI), with an average of 10.59±1.31 alleles (Table 2). The expected heterozygosity (H_E) showed an average value of 0.587±0.006 (Table 2), while average value of the observed heterozygosity (H_O) was 0.558±0.006 (Table 2). The lowest values in H_E and H_O was identified in the GU sample, whereas the highest ones in the AN sample (Table 2). No significant departure from Hardy-Weinberg equilibrium was found in any sample. Significant linkage disequilibrium (LD) was identified in 5 out 330 comparisons and involved loci Am_AC22-Am- AGAT33 in all populations except Livorno (LI). LD test performed on pair of loci across all populations identified highly significant linkage

between Am_AGAT36-Am_AC3, Am_AGAT1-Am_AAT3 and Am_AC22-Am_AGAT33.

The outlier identification analysis showed analogous results when IAM or SMM models were applied and suggested selection at 4 out of 11 loci. Directional selection was identified at loci Am_AC34 and Am_AGAT36, whereas balancing selection at loci Am_AAT3 and Am_AC1 (Fig.1). When the sample GU was excluded from the analysis, the results obtained with both IAM and SMM models, still suggested the identification of directional selection at locus Am_AC34 and balancing selection at locus Am_AAT3. The alignment of clone sequences containing loci under selection on nucleotide BLAST was successful for all loci with the exception of Am_AC1. The sequence of locus Am_AGAT36 (accession number KC199979.1) aligned correctly 414 out of 462 bp on *Maylandia zebra* MetZeb1.1 Scaffold 00026 (accession number NW_004531736.1) and within this sequence a perfect match for 29 out of 33 bp associated to a *calpain-5-like isoform x2-x1* was obtained. The sequence of locus Am_AC34 (accession number KC199981.1) aligned 235 out of 431 bp on *Maylandia zebra* MetZeb1.1 Scaffold 00006 (accession number NW_004531716.1) and within this sequence matched 76 out of 84 bp associated to a RUNT-related transcription factor 3-like isoform x1-x4. Finally, the sequence of locus Am_AAT3 aligned correctly 171 out of 514 bp on *Danio rerio strain Tuebingen chromosome 5, Zv9* (accession number NC_007116.5) and within this fragment matched 58 out of 63 bp coding for a peroxisomal membrane protein-2.

Spatial genetic structure and migration rate between samples

Pairwise values of both θ_{ST} and ENA-F_{ST} over all loci were similar and ranged from 0.0005 to 0.0999 (θ_{ST}) and 0.0007 to 0.1000 (ENA F_{ST}), respectively. Three out of

15 pairwise θ_{ST} were non-significant ($P < 0.0033$) and involved the CR vs. LI, AN vs. OR and OR vs. LI sample pairs (Table 3). The application of GENELAND, even when loci under selection were excluded by the analysis, allowed the identification of a genetic structure represented by five independent clusters (Fig.S1; $K_{\text{all-loci}} = 5$, Log-Likelihood $_{\text{all-loci}} = -2678.79$; $K_{\text{neutral-loci}} = 5$, Log-Likelihood $_{\text{neutral-loci}} = -5436.28$) corresponding to each locality analyzed with the exception of AN and OR that were clustered together (Fig.2). The analysis carried out with STRUCTURE showed a slightly different scenario. In fact, $K_{\text{all-loci}} = 3$ was the higher posterior probability [Pr(X/K)] when all loci were taken into account, allowing the identification of three clusters: i) Atlantic Ocean; ii) Western Mediterranean Sea; iii) Adriatic Sea. The analysis carried out on the seven neutral loci produced a $K_{\text{neutral-loci}} = 2$ as main result, allowing the detection of an Atlantic Ocean and a Mediterranean Sea cluster (Fig.3; Fig.S2).

The contemporary gene flow suggested that 5 out of 310 individuals analyzed were not allocated to the same location where they were sampled. Two individuals from LI were assigned to CR, one individual from CR was assigned to OR and finally two individuals from GU were assigned to PM and CR. The historical gene flow suggested M values between 15.67 (M_{5-6}) and 63.67 (M_{6-5} and M_{3-4}). The Θ values were quite similar among all sampling locations and ranged from 0.0972 in LI to 0.0986 in GU. Asymmetrical gene flow was detected between OR-LI and PM-GU, and lower levels of M estimations were detected between AN-CR and OR-PM.

Theta values (Θ) per location and M values with relative directionalities between geographically connected locations were reported in Fig.4A-B The software BARRIER v.2.2 identified up to five discontinuities in genetic frequencies suggesting a limited gene flow between samples. The bootstrapping procedures

revealed higher support for three out of five barriers identified, separating i) GU from PM, ii) PM from the rest of Mediterranean samples (LI, OR, AN and CR), and iii) LI from the Adriatic samples (OR, AN and CR) (Fig.4C). A limited role was highlighted for barriers separating CR from AN-OR and between AN and OR (Fig.4C). The Mantel test showed a lack of possible structure referable to isolation by distance (IBD). In fact, non-significant values ($P > 0.05$) were obtained when both the pairwise $\theta_{ST}/(1- \theta_{ST})$ and $\rho R_{ST}/(1- \rho R_{ST})$ were plotted against the geographical distances (Fig S3). The POWSIM analysis showed that the statistical power of the dataset allows to detect true population differentiation (F_{ST}) values as large as 0.0020 with a probability between 89.2% (all loci) and 83.9% (neutral loci) on the basis of results from Fisher's exact test, and between 93.4% and 90.2% on the basis of the Chi Squared test (Table S1). The significance of $F_{ST} = 0.0000$ (no drift and sampling from the base population) was assessed at 6.7% and 5.8% from Fisher's exact test and 3.2% and 2.9% of probability from Chi Squared test. All of these tests are quite close to 5% of error rate suggesting good reliability of simulations (Table S1). The BOTTLENECK test for detection of recent bottleneck affecting the localities analyzed, showed a lack of significances with both SMM and TPM models. However, the shift-mode test allowed the identification of a shifted distribution in GU sample (Table 4).

Discussion

Genetic variability in Aphia minuta and differences in genetic diversity between Mediterranean and Atlantic samples

The present study is the first description of microsatellite genetic variation in the transparent goby (*Aphia minuta*). The variability of microsatellite loci in *A. minuta*

has proved to be moderate (N_{AM} = 10.23±1.48; H_E = 0.587±0.064) if compared to other marine species, typically showing mean levels of allelic variation of at least 20.6±11.8 and higher heterozygosity (H = 0.77±0.19) (for further considerations see DeWoody & Avise 2000). Within the Family Gobiidae, this species seems to show polymorphism and heterozygosis degrees more similar to those of goby species with limited dispersal abilities or inhabiting freshwater environments (Hirase *et al.* 2012, Hughes *et al.* 2012), rather than to marine species with a potential higher gene flow (Pampoulie *et al.* 2004).

A possible explanation for the relatively low levels of genetic diversity recorded in *A. minuta* could be searched in some life history traits of this species. The short life-cycle (less than one year) coupled with a rapid achievement of sexual maturity lead to a reduction in length of larval phase that could concur (together with a demersal spawning) to maintain a relative low gene flow and a high self-recruitment, as observed in many other littoral fish species (Carreras-Carbonell *et al.* 2007, Hirase *et al.* 2012). Interestingly, the sample that showed the lowest level of genetic variability at all four genetic diversity estimators was the one from the Atlantic area (GU). This observation could be the result of great contemporary or past demographic variation in population size. Historical events, like Pleistocene climatic oscillations, have often been invoked to support phylogeographic patterns in marine fish (Bargelloni *et al.* 2003, Wilson & Veraguth 2010; Knutsen *et al*, 2013) and could be the reason for a reduction in genetic size of the Atlantic population of the transparent goby. In fact, the regression in distributional range for many marine species in the Atlantic area during the Last Glacial Maximum (c. 23-18000 years before present) could have also affected *A. minuta* populations. Although the molecular marker employed in this study is not ideal to provide direct insights into the historical signatures that can be

recorded in the genome of an organism, it could be speculated that the peculiar life cycle of the species could have promoted, during glacial and interglacial phases, cyclical bottlenecks and founder events leading to a decrease in genetic variability. This scenario was proposed in another goby species (*Pomatoschistus microps*) as an explanation to the lower genetic variability found in populations from north-eastern Atlantic coasts (Gysels *et al.* 2004).

Alternatively, the lower genetic levels detected in the Atlantic sample could came from a recent and abrupt reduction in population size. During 1998 the area of Guadalquivir estuary was affected by a sudden heavy metal pollution originated by the Aznalcóllar mining spill (Gómez-Parra *et al.* 2000). This event have rapidly produced high levels of mortality in Guadalquivir estuary communities and therefore may have led in the transparent goby population to a recent bottleneck. Despite the search for a recent genetic collapse in our samples suggested no significances for such a demographic event, a signal of shift in allele frequencies was detected only in GU sample. This latter result put forward a probable connection between low levels in genetic diversity estimators in this sample and the Aznalcóllar mining spill disaster and would suggests that the GU sample could have undergone to severe bottleneck events and are still reaching or have recently reached a new population equilibrium.

Genetic structure in Aphia minuta and hypotheses on current and historical dispersal

Simulations carried out using both STRUCTURE and GENELAND suggested the existence of a substantial genetic structure in *A. minuta* at nuclear DNA level, leading to the identification from two to five genetic clusters. However, these methods produced quite dissimilar results. STRUCTURE identified at most three

independent populations that cover approximately three main basins (Atlantic Ocean, Western Mediterranean Sea and Adriatic Sea), whereas GENELAND detected five independent clusters representing each of the locality analyzed with the exception of AN and OR samples, that were clustered together. The low number of first generation migrants identified and the pairwise θ_{ST} estimations seem to give more credit to the GENELAND results suggesting a high degree of genetic differentiation between samples analyzed. In addition, Hubisz *et al.* (2009) discussed the limitation of STRUCTURE to identify clusters that do not show a strong signal of genetic differentiation, thus producing an underestimation in the detection of genetically distinct units rather than spatially explicit models (i.e. GENELAND), especially in marine fish in which F_{ST} values are usually lower than in freshwater or terrestrial organisms (Ward 2006).

The genetic structure that emerged from assignment tests was detected also in other marine species (Sala-Bozano *et al.* 2009, Durand *et al.* 2013). However, this high genetic subdivision highlighted by nuclear DNA seems to be in contrast with the very low genetic differentiation detected at mitochondrial DNA level in this species within the Mediterranean Sea (Giovannotti *et al.* 2009). Discordance between these two molecular markers could be related to the different inheritance modalities of mtDNA and nDNA. In fact, mtDNA is mainly informative to describe historical genetic structure, due to its lack of recombination and maternal inheritance, whereas nDNA (that is recombinant and has biparental inheritance) gives better hints in current population structure (Ballard & Whitlock, 2004). Therefore, a possible explanation could be found in a recently built up genetic structure among Mediterranean populations of *A. minuta*, that cannot be detected using mtDNA markers. The M values found between localities suggest higher historical migration

rate rather than the current exchange of individuals and add further support to the idea of an incipient establishment of genetic differentiation between Mediterranean samples.

The directionality of historical gene flow seems to be coherent with the existence of genetic structure and suggests a past convergent exchange from Atlantic Ocean toward Western Mediterranean Sea and from Southern Adriatic Sea toward Tyrrhenian Sea. This pattern seems also consistent with the identification of three main barriers to equal exchange of individuals between locations. These barriers could be related to the existence of hydrographic features (i.e. pattern of local circulation of water masses) or to topographical conformation of coastlines. In fact, the isolation of Atlantic Ocean and Adriatic Sea samples from the Central Mediterranean Sea could be connected to the role played by Gibraltar Strait or Almería-Oran oceanic front, on one hand, and by Otranto and/or Siculo-Tunisian Straits or the Ionian cyclonic circulation of water masses, to the other. These oceanographic features are often invoked to explain genetic discontinuities (phylogeographic breaks) among Atlantic Ocean vs Mediterranean Sea and Western vs Eastern Mediterranean regions, in several marine organisms, including gobiid species (Patarnello *et al.* 2007, Mejri *et al.* 2011) and could, therefore, have a role in shaping nuclear genetic structure in the transparent goby.

Selective pressures and their contribution to the genetic differentiation of Aphia minuta

An alternative hypothesis to a recent establishment of genetic structure in *A. minuta* within the Mediterranean Sea, could be searched in the role generated by selective pressures. This additional scenario is not excluded by the one proposed above, but

could be even complementary to it. Selective pressure was recently identified as the main factor that is shaping the genetic structure in a Japanese paedomorphic goby species (*Leucopsarion petersii*), affecting, in particular, the activity of *Neuropeptide Y (NPY)*, an important regulator of food intake and orexigenic agent (Kokita *et al.* 2013).

The detection of outlier loci, together with peculiar life history traits (i.e. short life cycle and post-spawning death) and heterogeneity in ecological and environmental features within Mediterranean Sea sub-basins, would suggest the involvement of natural selection determining the genetic structure in this species. Particularly, microsatellite loci identified under directional selection (Am_AC34 and Am_AGAT36) were found to contain homology in sequences with genes involved in the apoptotic cycle, like *calpain-5-like isoform* and *RUNT-related transcription factor-3-like isoform* (*RUNX3*). In fact, *calpain-5* is a calcium-activated protease, that was highly described as abundant in mammalian colon, small intestine and testis, and connected with epithelium degenerative programmed processes (Huang & Wang 2001). At the same time, *RUNX3* is a transcription factor that seems to be involved with gastric cells apoptosis in human (Vogiatzi *et al.* 2006). Both *calpain-5* and *RUNX3* genes expression suggest that selective processes are probably acting on intestine epithelium and could be associated with the almost semelparous cycle in *Aphia*, when an endogenous timer, based on apoptotic events, triggers the degeneration of intestinal epithelium leading individuals to death after their second spawning (Caputo *et al.* 2002). The reproduction in *A. minuta*, as a consequence of wide latitudinal range, varies extensively (for an exhaustive review e.g. La Mesa *et al.* 2005). The spawners mass mortality together with a different time in spawning peak between the individuals from the six localities investigated in this study, could

generate different expression of genes that regulate apoptosis and consequently lead to a pronounced genetic structure.

Consequences of genetic structure in Aphia minuta fishery management

Finally, the application of microsatellite DNA markers allowed the identification of a current limited gene flow between localities analyzed that could be related to the presence of hydrographical features and/or the occurrence of selective pressure on genes related to the life cycle of this species. These results should point the attention to conservation issues concerning the potential risk of local failure in recruitment connected with fishery and pollution. In fact, the transparent goby life cycle and its limited dispersal make the recruitment of next generation reliant essentially on the previous one. Therefore, each local population of transparent goby should be treated as single genetic stock (see Carvalho & Hauser, 1994) and carefully managed in order to avoid their collapse.

Even though low levels of genetic variability and a presumable past signature of genetic bottleneck in GU sample could be more easily attributable to Aznalcóllar mining spill, it should be taken into consideration that Guadalquivir estuary is also one of the most exploited fishing area along the Spain coasts (Sobrino *et al.* 2005). Fisheries have been several times proved to be implied in alteration of both fish population dynamics (Hutchings 2005) and fish genetic variability (Hauser *et al.* 2002, Ruggeri *et al.* 2012) and, therefore, it could be speculated that fishery can also affect local *A. minuta* populations. As a possible future application of microsatellite markers used in this work, this issue could be analyzed in further detail with the aim to define at what extent local exploitation contribute to reduce genetic variability in this valuable fishery resource.

Figures and Tables

Fig. 1 Graphical outputs from LOSITAN (Antao *et al.* 2008) software applied on the basis of IAM (Fig.1A) and SMM (Fig.1B). Blue dots represent interpolation between F_{ST} and H_E mean values in each marker (microsatellite locus) analyzed. Blue dots falling on light grey area represent neutral markers whereas the yellow and the light red areas represent marker under directional and balancing selection, respectively.

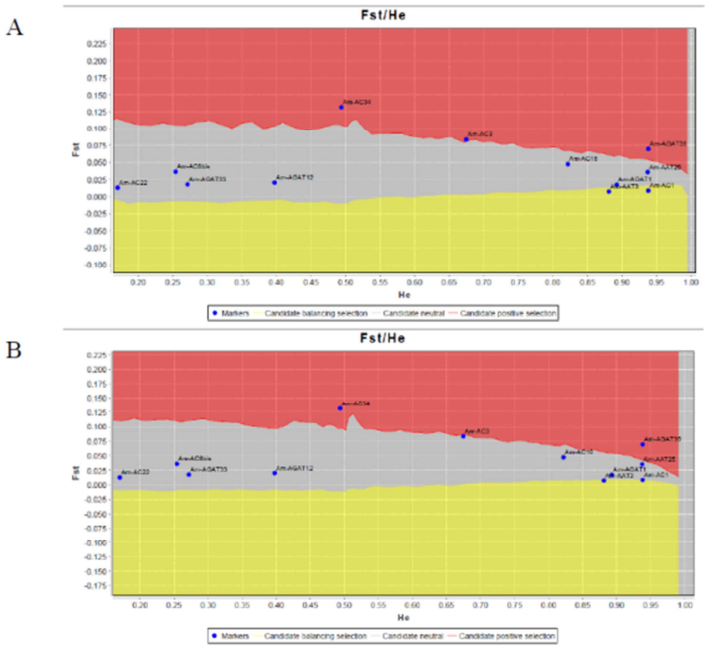


Fig. 2 Graphical outputs from GENELAND (Guillot *et al.* 2005a) simulations. The gradient maps (Fig.2A-E) explain (with lighter colours) the highest probable distribution of each cluster identified within the geographic area investigated. The Fig.2F summarize the geographical patch that identify each of the five clusters detected, using different colors

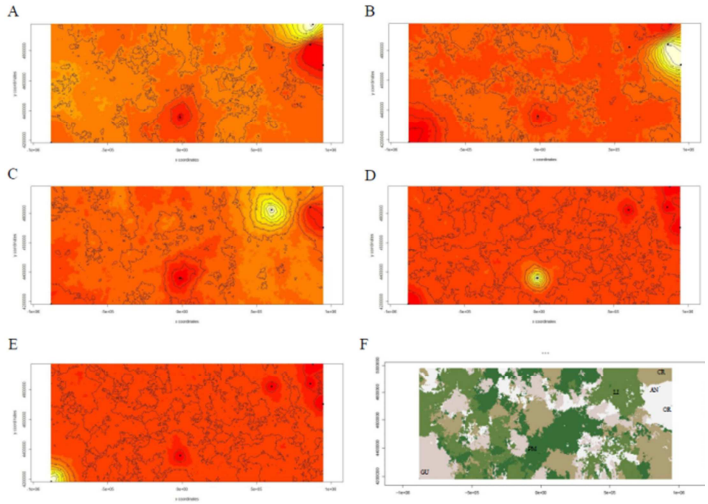


Fig. 3 Bar plot representation of STRUCTURE v.2.3.2.1 (Pritchard *et al.* 2000, Falush *et al.* 2003) results. In Fig.3A was showed a two cluster structure (red: CR, AN, OR, LI, PM; Green: GU) obtained using seven neutral loci, whereas in Fig.3B a three cluster structure (blue: CR, AN, OR; green: LI, PM; red: GU) obtained using all the eleven loci typed in this study.

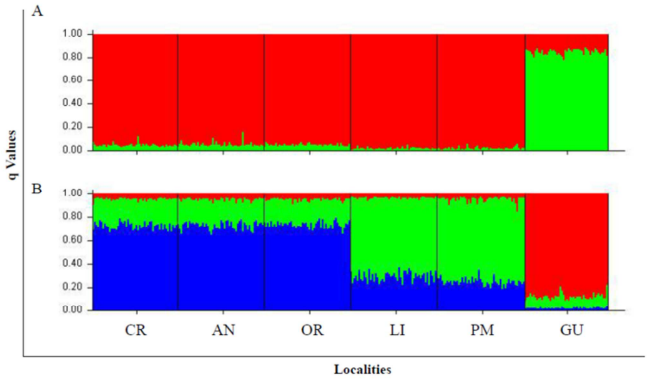


Fig. 4 The map in Fig.4A explains the historical gene flow (MIGRATE-n v. 3.2.1; Beerli *et al.* 2009) between locations taking into account for the main barriers identified by the software BARRIER v.2.2 (Manni *et al.* 2004; Fig.4C). M values are expressed between locations, whereas Θ values are reported for each location. M values are graphically described using arrows that explain directionality and strength (the arrow's length) of gene flow. The dashed square identify the area covered by Fig.4B. The map in Fig.4C explain the main barriers to gene flow detected with BARRIER v.2.2 software. The dashed lines represent the main connections belong to the Delaunay triangulation, the polygons with light grey solid lines represents areas identified by Voronoï tessellation method. The black solid lines identify the main barriers to equal gene flow and the thickness of each barrier roughly reflect its strength.

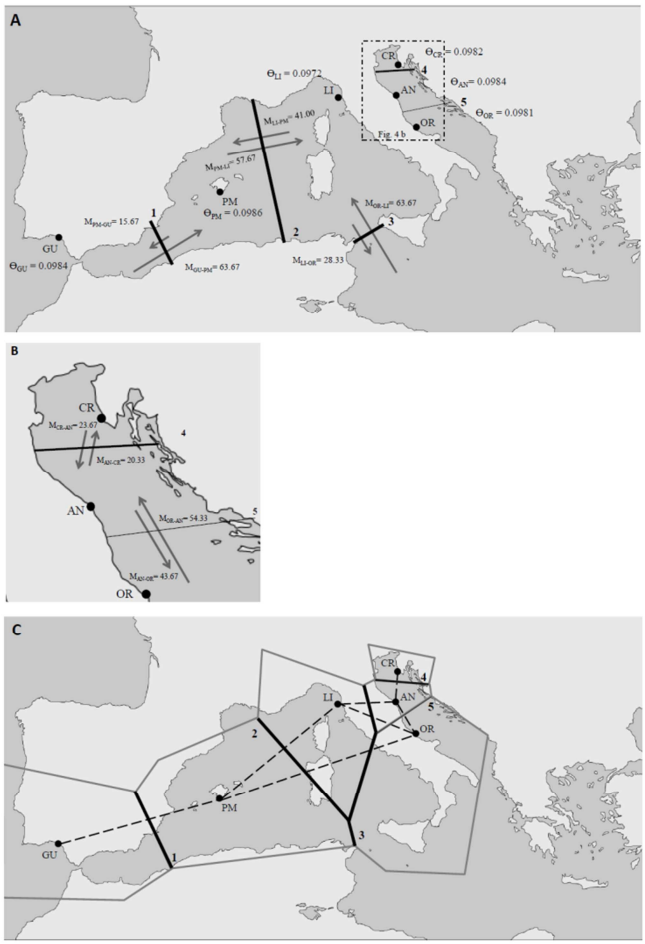


Table 1 Sampling information. The table resumes the name of the basin which each sample belongs to, the code for each sampling site, the geographical locality and coordinates of the sites where *Aphia minuta* individuals were sampled, the number of specimens analyzed for each locality and the date (month and year) when the samples were obtained.

Basin	Sample code	Locality	Coordinates	Number of individuals	Date of sampling
Adriatic Sea	CR	Pula (HR)	44°50'38"N; 13°47'58"E	51	March 2004
Adriatic Sea	AN	Ancona (I)	43°37'46"N; 13°29'30"E	52	January 2003
Adriatic Sea	OR	Ortona (I)	42°20'53"N; 14°25'58" E	52	June 2004
Tyhrrenian Sea	LI	Livorno (I)	43°32'04"N; 10°14'50"E	52	April 2003
Western Mediterranean Sea	PM	Palma de Mallorca (E)	39°13'41"N; 03°03'41"E	53	January 2003
Atlantic Ocean	GU	Guadalquivir Estuary (E)	36° 45'46"N; 06° 28' 05"O	50	February 2013

Table 2 Summary statistics of genetic variability at eleven microsatellite loci.

		CR	AN	OR	LI	PM	GU	$N_{A\text{-}tot}$
Am-AGAT1	N_A	20	16	14	22	21	19	34
	N	51	52	52	52	51	50	
	f_{null}	0,050	0,041	0,000	0,000	0,000	0,000	
	H_E	0,883	0,873	0,871	0,867	0,888	0,880	
	H_O	0,824	0,788	0,923	0,846	0,922	0,900	
	F_{IS}	0,068	0,098	-0,060	0,024	-0,038	-0,023	
	R_S	18,755	15,014	13,517	19,375	19,647	17,492	
Am-AC16	N_A	11	11	11	15	10	6	20
	N	51	52	52	52	53	49	
	f_{null}	0,039	0,000	**0,059**	0,000	0,002	0,038	
	H_E	0,850	0,820	0,822	0,850	0,835	0,519	
	H_O	0,745	0,846	0,731	0,885	0,849	0,469	
	F_{IS}	0,125	-0,033	0,112	-0,041	-0,017	0,097	
	R_S	10,41	10,102	10,278	13,973	9,711	5,673	
Am-AGAT36	N_A	23	20	21	23	24	13	36
	N	45	52	52	52	52	50	
	f_{null}	0,027	0,001	0,007	0,039	0,041	0,000	
	H_E	0,913	0,925	0,917	0,931	0,917	0,633	
	H_O	0,889	0,904	0,923	0,865	0,846	0,600	
	F_{IS}	0,027	0,023	-0,007	0,071	0,078	0,053	
	R_S	22,485	19,137	19,741	21,488	21,98	12,007	
Am-AGAT12	N_A	6	9	7	8	7	6	10
	N	44	52	52	47	49	50	
	f_{null}	**0,103***	0,000	0,031	**0,078**	**0,070**	0,000	
	H_E	0,214	0,443	0,466	0,451	0,330	0,433	
	H_O	0,136	0,462	0,423	0,362	0,265	0,480	
	F_{IS}	**0,365**	-0,042	0,093	0,200	0,197	-0,111	
	R_S	5,86	8,51	6,58	7,62	6,67	5,46	
Am-AC8	N_A	4	3	5	4	3	3	5

		CR	AN	OR	LI	PM	GU	$N_{A\text{-}tot}$
Am-AC34	N_A	5	4	2	5	6	3	8
	N	41	52	41	52	52	50	
	f_{null}	**0,113***	0,000	0,010	0,000	0,000	0,012	
	H_E	0,430	0,529	0,499	0,512	0,372	0,231	
	H_O	0,341	0,558	0,439	0,519	0,404	0,220	
	F_{IS}	0,207	-0,054	0,121	-0,014	-0,086	0,048	
	R_S	5,00	3,75	2,00	4,57	5,66	2,82	
Am-AC3	N_A	6	7	6	6	8	3	10
	N	46	46	51	52	52	50	
	f_{null}	0,028	**0,084**	0,016	0,042	**0,070**	0,000	
	H_E	0,784	0,720	0,680	0,783	0,594	0,150	
	H_O	0,717	0,630	0,647	0,692	0,519	0,160	
	F_{IS}	0,085	0,126	0,049	0,117	0,127	-0,067	
	R_S	5,89	6,88	6,00	6,00	7,37	2,82	
Am-AAT3	N_A	13	17	21	16	17	12	26
	N	48	46	48	48	51	46	
	f_{null}	0,000	**0,106***	**0,071**	**0,082***	0,047	**0,062**	
	H_E	0,839	0,894	0,885	0,859	0,911	0,860	
	H_O	0,833	0,761	0,792	0,750	0,824	0,783	
	F_{IS}	0,007	**0,150**	0,106	0,128	0,097	0,090	
	R_S	12,52	16,43	19,37	15,23	16,32	11,77	
Am-AC1	N_A	27	32	30	27	30	22	49
	N	46	52	52	51	49	48	
	f_{null}	0,010	0,008	0,034	**0,051**	**0,073**	0,039	
	H_E	0,940	0,938	0,924	0,919	0,929	0,929	
	H_O	0,913	0,885	0,846	0,824	0,813	0,833	
	F_{IS}	0,029	0,058	0,085	**0,105**	**0,126**	0,104	
	R_S	26,06	29,093	27,232	25,209	28,102	20,77	
Am-AGAT33	N_A	6	6	3	8	6	4	11

	N	46	52	52	52	53	48			N	51	52	52	52	53	50
	f_{null}	0,000	0,000	0,000	0,000	**0,059**	0,000			f_{null}	0,000	0,000	0,000	0,000	0,024	0,000
	H_E	0,239	0,146	0,469	0,332	0,126	0,158			H_E	0,375	0,383	0,161	0,250	0,242	0,188
	H_O	0,261	0,135	0,481	0,346	0,094	0,146			H_O	0,412	0,442	0,173	0,269	0,226	0,200
	F_{IS}	-0,094	0,080	-0,025	-0,044	0,251	0,076			F_{IS}	-0,100	-0,155	-0,075	-0,076	0,067	-0,064
	R_S	3,89	2,99	4,96	3,79	2,77	3,00			R_S	5,76	5,94	2,79	7,40	5,45	3,96
Am-AC22	N_A	3	3	2	4	3	3	4	**Average**	N_{AM}	11,27	11,64	11,09	12,55	12,27	8,55
	N	50	52	52	52	52	50			H_E	0,609	0,629	0,623	0,623	0,575	0,462
	f_{null}	0,000	0,000	0,010	0,000	0,036	0,000			H_O	0,574	0,607	0,596	0,587	0,538	0,445
	H_E	0,237	0,244	0,160	0,094	0,177	0,097			F_{IS}	0,059	0,035	0,045	0,058	0,065	0,037
	H_O	0,240	0,269	0,173	0,096	0,154	0,100			R_S	10,88	10,99	10,41	11,66	11,50	8,07
	F_{IS}	-0,015	-0,106	-0,085	-0,026	0,133	-0,029									
	R_S	3,00	3,00	2,00	3,57	2,79	2,96									

NA-tot, Number of alleles observed in a specific locus; NA, number of alleles observed per location; N, number of individuals correctly genotyped; NAM, mean number of alleles observed per location; fnull, null allele frequency calculated with Dempster method (Dempster *et al.* 1977); HO, observed heterozygosity; HE, expected heterozygosity; FIS, inbreeding coefficient estimates. RS, allelic richness estimates standardized at 41 individuals. Bold FIS values are significant (<0.05) after a sequential Bonferroni correction. Bold fnull values are those exceed the 5% null allele tolerance threshold. Asterisk (*) denotes estimates still showed null allele signals after Brookfield (1996) method.

Table 3 Pairwise genetic differentiation based on θ_{ST} (below) and ENA-F_{ST} (above).

	CR	AN	OR	LI	PM	GU
CR	-	0.0187	0.0244	0.0075	0.0232	0.1000
AN	**0.0196***	-	0.0007	0.0010	0.0273	0.0967
OR	**0.0267***	0.0005	-	0.00863	0.0341	0.0946
LI	0.0077	**0.0096***	0.0090	-	0.0173	0.0979
PM	**0.0231***	**0.0267***	**0.0357***	**0.0171***	-	0.0640
GU	**0.0999***	**0.0959***	**0.0960***	**0.0966***	**0.0631***	-

Asterisc (*) and bold θ_{ST} values are significant after Bonferroni correction (Rice 1989).

Table 4 Wilcoxon test for bottleneck detection based on SMM and TPM mutation models and test for allele frequencies distribution (Shape-Mode Test).

	Wilcoxon Test		**Shape Mode**
Samples	**SMM**	**TPM**	
CR	0.41	0.23	L-shaped
AN	0.21	0.18	L-shaped
OR	0.14	0.09	L-shaped
LI	0.45	0.45	L-shaped
PM	0.88	0.84	L-shaped
GU	0.14	0.12	Shifted

One-tail Wilcoxon Test P-values (significancies P<0.05) from SMM and TPM models. **SMM**: Stepwise Mutation Model; **TPM**: Two Phase Mutation Model. **L-shape** identify normal distribution in allele frequencies, whereas **Shifted** identify a bottleneck signature in allele frequencies distribution within the population.

References

Antao T., Lopes A., Lopes R.J., Beja-Pereira A., Luikart G. (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323-327.

Avise J.C. (2000*) Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA: 3-108.

Ballard J.W.O., Whitlock M.C. (2004) The incomplete natural history of mithocondria. *Molecular Ecology*, **13**, 729–744.

Bargelloni L., Alarcon J.A., Alvarez M.C., Penzo E., Magoulas A., Reis C., Patarnello T. (2003) Discord in the family Sparidae (Teleostei): divergent phylogeographical patterns across the Atlantic-Mediterranean divide. *Journal of Evolutionary Biology*, **16**, 1149–1158.

Beerli P. (2006) Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.

Beerli P. (2009) How to use migrate or why are markov chain monte carlo programs difficult to use? In: Bertorelle G., Bruford M.W., Hauffe H.C., Rizzoli A., Vernesi C. (Eds), *Population Genetics for Animal Conservation*). Cambridge University Press, Cambridge, UK: 42-79.

Berrebi P., Cattaneo-Berrebi G., Valade P., Ricou J.F., Hoareau T. (2005) Genetic homogeneity in eight freshwater populations of *Sicyopterus lagocephalus*, an amphidromous gobiid of La Réunion Island. *Marine Biology*, **148**, 179-188.

Brookfield J.F.Y. (1996) A simple method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Caputo V., Candi G., La Mesa M., Arneri E. (2001) Pattern of gonad maturation and the question of semelparity in the paedomorphic goby *Aphia minuta*. *Journal of Fish Biology*, **58**, 656–669.

Caputo V., Candi G., Arneri E., La Mesa M., Cinti C., Provinciali M., Nisi Cerioni P., Gregorini A. (2002) Short lifespan and apoptosis in *Aphia minuta*. *Journal of Fish Biology*, **60**, 775–779.

Caputo V., La Mesa M., Candi G., Nisi Cerioni P. (2003) The reproductive biology of the crystal goby with a comparison to that of the transparent goby. *Journal of Fish Biology*, **62**, 375–385.

Carreras-Carbonell J., Macpherson E., Pascual M. (2007) High self-recruitment levels in a Mediterranean littoral fish population revealed by microsatellite markers. *Marine Biology*, **151**, 719–727.

Carvalho G.R., Hauser L. (1994) Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries*, **4**, 326-350.

Chapuis M.P., Estoup A. (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Dempster A.P., Laird N.M., Rubin D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B (Methodological)*, **39**, 1–38.

DeWoody J.A., Avise J.C. (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology*, **56**, 461–473.

Duchene D., Klanten S.O., Munday P.L., Herler J., van Herwerden L. (2013) Phylogenetic evidence for recent diversification of obligate coral-dwelling gobies compared with their host corals. *Molecular Phylogenetics and Evolution*, **69**, 123–132.

Durand J.D., Blel H., Shen K.N., Koutrakis E.T., Guinand B. (2013) Population genetic structure of *Mugil cephalus* in the Mediterranean and Black Seas: a single mitochondrial clade and many nuclear barriers. *Marine Ecology Progress Series*, **474**, 243-261.

Earl D.A., von Holdt B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Earl D.A., Louie K.D., Bardeleben C., Swift C.C., Jacobs D.K. (2010) Rangewide microsatellite phylogeography of the endangered tidewater goby, *Eucyclogobius newberryi* (Teleostei: Gobiidae), a genetically subdivided coastal fish with limited marine dispersal. *Conservation Genetics*, **11**, 103–114.

Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Falush D., Stephens M., Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Faurby S., Barber P.H. (2012) Theoretical limits to the correlation between pelagic larval duration and population genetic structure. *Molecular Ecology*, **21**, 3419–3432.

Giovannotti M., La Mesa M., Caputo V. (2009) Life style and genetic variation in teleosts: the case of pelagic (*Aphia minuta*) and benthic (*Gobius niger*) gobies (Perciformes: Gobiidae). *Marine Biology*, **156**, 239–252.

Gómez-Parra A., Forja J.M., Delvalls T.A., Sáenz I., Riba I. (2000) Early contamination by heavy metals of the Guadalquivir Estuary after the Aznalcóllar mining spill (SW Spain). *Marine Pollution Bulletin*, **40**, 1115–1123.

Goudet J. (2001*)* FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from www.unil.ch/izea/softwares/fstat.html.

Gould S.J. (1977) *Ontogeny and Phylogeny*. Harvard University Press, Harvard, MA: 303-350.

Guillot G., Mortier F., Estoup A. (2005a) Geneland: a program for landscape genetics. *Molecular Ecology Notes*, **5**, 712-715.

Guillot G., Estoup A., Mortier F., Cosson J.F. (2005b) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261-1280.

Guo S., Thompson E. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Gysels E.S., Hellemans P., Pampoulie C., Volckaert F.A.M. (2004) Phylogeography of the common goby, *Pomatoschistus microps*, with particular emphasis on the colonization of the Mediterranean and the North Sea. *Molecular Ecology*, **13**, 403–417.

Hauser L., Adcock G.J., Smith P.J., Ramirez J.H.B., Carvalho G.R. (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences*, **99**, 11742-11747.

Hirase S., Kanno M., Ikeda M., Kijima A. (2012) Evidence of the restricted gene flow within a small spatial scale in the Japanese common intertidal goby *Chaenogobius annularis*. *Marine Ecology*, **33**, 481–489.

Huang Y., Wang K.K.W. (2001) The calpain family and human disease. *Trends in Molecular Medicine*, **7**, 355–362.

Hubisz M.J., Falush D., Stephens M., Pritchard J.K. (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322-1332.

Hughes J.M., Real K.M., Marshall J.C., Schmidt D.J. (2012) Extreme genetic structure in a small-bodied freshwater fish, the purple spotted gudgeon, *Mogurnda adspersa* (Eleotridae). *PLoS ONE*, **7**, e40546.

Hutchings J.A. (2005) Life history consequences of overexploitation to population recovery in Northwest Atlantic cod (*Gadus morhua*). *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 824-832.

Knutsen H., Jorde P.E., Gonzalez E.B., Robalo J., Albretsen J., Almada V. (2013) Climate change and genetic structure of leading edge and rear end populations in a northwards shifting marine fish species, the corkwing wrasse (*Symphodus melops*). *PLoS ONE*, **8**, e67492.

Kokita T., Takahashi S., Kumada H. (2013) Molecular signatures of lineage-specific adaptive evolution in a unique sea basin: the example of an anadromous goby *Leucopsarion petersii*. *Molecular Ecology*, **22**, 1341–1355.

La Mesa M., Arneri E., Caputo V., Iglesias M. (2005) The transparent goby, *Aphia minuta*: review of biology and fisheries of a paedomorphic European fish. *Review in Fish Biology and Fisheries*, **15**, 89–109.

Latch E., Dharmarajan G., Glaubitz J.C., Rhodes O.E.J. (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295-302.

Magoulas A., Tsimenides N., Zouros E. (1996) Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of European anchovy (*Engraulis encrasicolus*). *Moleculr Biology and Evolution*, **13**, 178-190.

Manni F., Guérard E., Heyer E. (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". *Human Biology*, **76**, 173–190.

Mejri R., Arculeo M., Hassine O.K.B., Lo Brutto S. (2011) Genetic architecture of the marbled goby *Pomatoschistus marmoratus* (Perciformes, Gobiidae) in the Mediterranean Sea. *Molecular Phylogenetics and Evolution*, **58**, 395–403.

Miller P.J. (1973) The species of Pseudaphya (Teleostei: Gobiidae) and the evolution of aphyiine gobies. *Journal of Fish Biology*, **5**, 353–365.

Miller P.J. (1979) Adaptiveness and implications of small size in teleosts. *Symposium of the Zoological Society of London*, **44**, 263–306.

Miller P.J. (1986) Gobiidae. In: Whitehead P.J.P., Bauchot M.L., Hureau J.C., Nielsen J., Tortonese E. (Eds), *Fishes of the North-Eastern Atlantic and Mediterranean*, Vol 3. UNESCO, Bungay: 1019-1085.

Nelson J.S. (2006) *Fishes of the world*, 4th edn. Wiley, New York, NY: 422-426.

Paetkau D., Slade R., Burden M., Estoup A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Pampoulie C., Gysels E.S., Maes G.E., Hellemans B., Leentjes V., Jones A.G., Volckaert F.A.M. (2004) Evidence for fine-scale genetic structure and estuarine colonisation in a potential high gene flow marine goby (*Pomatoschistus minutus*). *Heredity*, **92**, 434–445.

Patarnello T., Volckaert F.A.M., Castilho R. (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, **16**, 4426–4444.

Piry S., Luikart G., Cornuet J.M. (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Piry S., Alapetite A., Cornuet J.M., Paetkau D., Baudouin L., Estoup A. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.

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

Rannala B., Mountain J.L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences*, **94**, 9197–9201.

Rice W.R. (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

Riginos C., Douglas K.E., Jin Y., Shanahan D.F., Treml E.A. (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*, **34**, 566–575.

Rousset F. (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Ruggeri P., Splendiani A., Bonanomi S., Arneri E., Cingolani N., Santojanni A., Belardinelli A., Giovannotti M., Caputo V. (2012) Temporal genetic variation as revealed by a microsatellite analysis of European sardine (*Sardina pilchardus*) archived samples. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**, 1698–1709.

Ruggeri P., Splendiani A., Giovannotti M., Fioravanti T., Nisi Cerioni P., Caputo V. (2013) Isolation and characterisation of novel microsatellite loci in the transparent goby *Aphia minuta* (Perciformes, Gobiidae). *Conservation Genetics Resources*, **5**, 775–777.

Ryman N., Palm S. (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, **6**, 600-602.

Sala-Bozano M., Ketmaier V., Mariani S. (2009) Contrasting signals from multiple markers illuminate population connectivity in a marine fish. *Molecular Ecology*, **18**, 4811-4826.

Sobrino I., Baldó F., García-González D., Cuesta J.A., Silva-García A., Fernández-Delgado C., Drake P. (2005) The effect of estuarine fisheries on juvenile fish observed within the Guadalquivir Estuary (SW Spain). *Fisheries Research*, **76**, 229-242.

Stepien C.A., Tumeo M.A. (2006) Invasion genetics of Ponto-Caspian gobies in the Great Lakes: a 'cryptic' species, absence of founder effects, and comparative risk analysis. *Biological Invasions*, **8**, 61–78.

Taggart J.B., Hynes R.A., Prodöuhl P.A., Ferguson A. (1992) A simplified protocol for routine total DNA isolation from salmonid fishes. *Journal of Fish Biology*, **40**, 963–965.

Thacker C.E. (2009) Phylogeny of Gobiodei and placement within Acanthomorpha, with a new classification and investigation of diversification and character evolution. *Copeia*, **2009**, 93–104.

Thacker C.E., Thompson A.R., Roje D.M. (2011) Phylogeny and evolution of Indo-Pacific shrimp-associated gobies (Gobiiformes: Gobiidae). *Molecular Phylogenetics and Evolution*, **59**, 168–176.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M., Shipley P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Viñas J., Alvarado Bremer J.R., Pla C. (2004) Phylogeography of Atlantic bonito (*Sarda sarda*) in the northern Mediterranean: the combined effects of historical vicariance, population expansion, secondary invasion and isolation by distance. *Molecular Phylogenetics and Evolution*, **33**, 32-42.

Vogiatzi P., De Falco G., Claudio P.P., Giordano A. (2006) How does the human RUNX3 gene induce apoptosis in gastric cancer? *Cancer Biology and Therapy*, **5**, 371-374.

Ward R.D. (2006) The importance of identifying spatial population structure in restocking and stock enhancement programmes. *Fisheries Research*, **80**, 9–18.

Ward R.D., Woodwark M., Skibinski D.O.F. (1994) A comparison of genetic diversity levels in marine, freshwater, and anadramous fishes. *Journal of Fish Biology*, **44**, 213–232.

Weir B.S., Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wilson A.B., Veraguth I.E. (2010) The impact of Pleistocene glaciation across the range of a widespread European coastal species. *Molecular Ecology*, **19**, 4535–4553.

CHAPTER II

Ruggeri P., Splendiani A., Occhipinti G., **Fioravanti T.**, Santojanni A., Leonori I., De Felice A., Arneri E., Procaccini G., Catanese G., Tičina V., Bonanno A., Cerioni Nisi P., Giovannotti M., Grant W.S., Caputo Barucchi V. (2016) Biocomplexity in Populations of European Anchovy in the Adriatic Sea. *PLoS ONE*, **11** (4), e0153061. doi:10.1371/journal.pone.0153061

Biocomplexity in Populations of European Anchovy in the Adriatic Sea

Abstract

The sustained exploitation of marine populations requires an understanding of a species' adaptive seascape so that populations can track environmental changes from short- and long-term climate cycles and from human development. The analysis of the distributions of genetic markers among populations, together with correlates of life-history and environmental variability, can provide insights into the extent of adaptive variation. Here, we examined genetic variability among populations of mature European anchovies (n = 531) in the Adriatic (13 samples) and Tyrrhenian seas (2 samples) with neutral and putative non-neutral microsatellite loci. These genetic markers failed to confirm the occurrence of two anchovy species in the Adriatic Sea, as previously postulated. However, we found fine-scale population structure in the Adriatic, especially in northern areas, that was associated with four of the 13 environmental variables tested. Geographic gradients in sea temperature, salinity and dissolved oxygen appear to drive adaptive differences in spawning time and early larval development among populations. Resolving adaptive seascapes in Adriatic anchovies provides a means to understand mechanisms underpinning local adaptation and a basis for optimizing exploitation strategies for sustainable harvests.

Introduction

A growing body of evidence indicates that marine organisms with potentially dispersive planktonic larvae can show genetically patchy distributions that may be

uncorrelated with spatial complexity (Selkoe *et al.* 2008; Hedgecock & Pudovkin 2011). This unexpected genetic population structure has been attributed to natural selection on life-history stages (Kelly & Palumbi 2010; Selkoe & Toonen 2011), on isolation by hydrographic barriers to gene flow (White *et al.* 2010) and on adaptation to local environments (Gaggiotti *et al.* 2009). The interplay between these forces is referred to as biocomplexity, which underpins population resilience and persistence (Hilborn *et al.* 2003; Ruzzante *et al.* 2006; Olsen *et al.* 2008). The use of ocean-wide environmental databases and genetic markers to understand the origins of biocomplexity is a relatively new area known as seascape genetics (Manel & Holderegger 2013; Riginos & Liggins 2013). Spatial patterns among marine populations can provide insights into interactions between environmental variables and life-history traits and can illuminate the roles of these variables in shaping genetic structure. Marine ecosystems are dynamic and complex, reflecting interactions between biotic and abiotic variables (Riginos & Liggins 2013) . Small-scale spatial and temporal heterogeneity can impose selective constraints on the development and survival of larvae and juveniles (Riginos & Liggins 2013; Selkoe *et al.* 2010) and hence on the distributions of adults (Lamichhaney *et al.* 2012; Riccioni *et al.* 2013; Milano *et al.* 2014).

Here, we focus on anchovy (*Engraulis encrasicolus* L.) populations in the Adriatic Sea. This species support large fisheries, as well as artisanal harvests along the coasts of Italy and the Balkan countries. The Adriatic is a semi-enclosed basin open to the Central-Eastern Mediterranean Sea and is divided into several areas with contrasting levels of productivity (Buljan & Zore-Armanda 1974). Four areas with different local hydrographic and physical properties can be recognized (Artegiani *et al.* 1997a,

1997b). The first area (Zone A, *sensu* Buljan & Zore-Armanda 1974) encompasses deep, oligotrophic waters in the southern central Adriatic and receives warm, high saline water from the Eastern Mediterranean Sea. A second area (Zone B) lies in shallow waters (mean depth < 40–50 m) of the northwestern Adriatic and receives large amounts of freshwater from the western side of the Adriatic. Zone B is characterized by high levels of productivity and strong seasonal swings in temperature, salinity and dissolved oxygen. A third area (Zone C) lies in the northeastern Adriatic Sea with depths greater than those in the western Adriatic, moderate levels of productivity, high salinities and stable environmental conditions throughout the year. A fourth area (Zone D) includes the lagoons and channels on both sides of the Adriatic. These areas have the highest levels of productivity in the Adriatic Sea, but represent only 1–2% of its surface (Buljan & Zore-Armanda 1974; Artegiani *et al.* 1997a, 1997b). High levels of productivity in the northern Adriatic basin support spawning and feeding grounds for many marine species of economic and ecological importance. One of these is the European anchovy which spawns several times from April to October in defined spawning areas (Morello & Arneri 2009). Potential spawning and juvenile nursery areas are located over the continental shelf in the northwestern Adriatic, where high inshore productivity enhances larval growth and promotes the recruitment of juveniles into adult populations (Santojanni *et al.* 2006).

Two anchovy stocks in the Adriatic were previously described based on color (Borsa *et al.* 2004) and allozyme-frequency variation (Bembo *et al.* 1996). A small silver anchovy occurs largely in the shallow and less saline waters of the northern Adriatic, and a larger bluish form inhabits the open waters of the central basin, for the most

part. These morphs are superficially similar to *E. albidus* (silver anchovy) and *E. encrasicolus* (blue anchovy) in the Gulf of Lyon, and led Borsa *et al.* (2004) to suggest that both species also inhabit the Adriatic Sea. Additionally, Bembo *et al.* (1996) found strong allozyme-frequency gradients among populations in the northwestern and southeastern areas of the Adriatic. However, the distributions of these gradients did not correspond entirely with the distributions of the two putative species. In the interim, no studies have been made with additional molecular markers to investigate these differences.

Here, we used microsatellite DNA markers to resolve the fine-scale genetic structure of anchovy populations in the Adriatic. Among the molecular markers used for population genetic studies, microsatellite DNA has proved to be a powerful marker in studies of population structure at regional and sub-basin scales due to its high level of polymorphism (reviewed in Chistiakov *et al.* 2006). We had four objectives: i) test whether multiple species of anchovies exist in the Adriatic basin as postulated by Borsa *et al.* (2004), ii) identify genetically discrete populations and local stocks, iii) estimate the directionality of gene flow and the level of connectivity among populations and iv) identify environmental variables influencing the adaptive seascape of Adriatic anchovies.

Materials and Methods

Ethical statement, Sample collection and DNA extraction

A total of 531 mature European anchovies were collected at 13 localities in the Adriatic between July and September 2012 and at two localities in the Tyrrhenian

Sea between May and June 2013 (Fig. 1A; Table 1A). Since samples of European anchovies were collected during research cruises, special ethical permits for scientific sampling were not required. Samples from the Adriatic, except the Boka Kotorska sample (KOT), were collected by midwater pelagic trawls during annual acoustic surveys between July and September 2012 in southern (July) and northern-central (September) areas of the Adriatic Sea. Tyrrhenian samples were collected with pelagic trawls during the EVATIR 2013 oceanographic research cruise that took place May–June 2013. No area under protection or special management was included in our sampling design. Permission to carry out surveys in Italian waters was granted by the MIPAAF "Ministero per le Politiche Alimentari, Agricole e Forestali", Stato Maggiore della Marina (MARISTAT), Istituto Maridrografico, Comandi Militari Marittimi di La Spezia and Taranto and the local Adriatic/Tyrrhenian Port Authorities. Sampling in the waters of Montenegro was permitted by the Local Ministry of Agriculture and Rural Development, and sampling off Albania was permitted by the Ministria Mjedisit, Pyjeve dhe Administrimit Ujrave Drejtorise Politikave te Peshkimit. Permission to conduct the acoustic survey and related biological samplings of small pelagic fish in the eastern part of the Adriatic Sea during September 2012 on board the r/v BIOS DVA was granted by the Croatian Ministry of Agriculture and Ministry of science, education and sports. Whole animals were preserved in absolute or 95% ethanol at -20°C until DNA extraction. Total DNA was extracted from gill or muscle tissue using a standard phenol-chloroform protocol (gill) (Taggart et al., 1992), or with the NucleoSpin Tissue Kit (muscle) (Macherey-Nagel).

Molecular markers and PCR amplification

All samples were screened at 14 previously described microsatellite loci (Chiu *et al.* 2002; Landi *et al.* 2005; Pakaki *et al.* 2009; Lin *et al.* 2011). Eleven loci were labeled with fluorescent dyes and multiplexed in three polymerase chain reactions (PCR) (Table S1; File S1). These loci were genotyped using an ABI-PRISM 3130xl Genetic Analyzer (Applied Biosystems). The remaining loci, Enja-148, Ee2-508 and Ee2-165b, were amplified individually due to moderate levels of polymorphism and were screened with a silver staining protocol (Bassam *et al.* 1991). Additional information about PCR protocols, amplification profiles, fragment genotyping, internal standards, silver staining adopted for the non-labeled PCR products and genotyping accuracy are provided in File S1.

Statistical treatment of microsatellite and mtDNA data

The incidences of null alleles, allelic dropout and stutter peaks were assessed with MICROCHECKER 2.2.1 (Van Oosterhout *et al.* 2004), and loci affected by null alleles were corrected with the Brookfield algorithm (Brookfield 1996) in MICROCHECKER. In addition, the Dempster algorithm (Dempster *et al.* 1977) in FreeNa (Chapuis & Estoup 2007) was used to estimate null allele frequencies per locus and per sampling locality. Number of alleles per locus (N_A), allelic richness (R_S), observed (H_O) and expected (H_E) heterozygosity, and the inbreeding coefficient (F_{IS}) were calculated for each sample using FSTAT 2.9.3 (Goudet 2001).

Conformation to Hardy-Weinberg proportions was tested with the exact test implemented in Genepop 4.0.10 (Rousset 2008) using a Markov Chain Monte Carlo (MCMC) method with 100 batches of 10,000 iterations each, with the first 1,000 iterations discarded before sampling. Genepop was also used to test for linkage

disequilibrium between loci by a MCMC chain executed with 1,000 batches of 2000 iterations each. A sequential Bonferroni adjustment of *P*-values was used to account for an increase in type-I error for multiple comparisons (Rice 1989).

Genetic differentiation between samples was estimated with Θ_{ST} (Weir & Cockerham 1984) and compared with divergences (F_{ST}) estimated with the Exclusion of Null Allele Method (ENA) in FreeNa (Chapuis & Estoup 2007). High heterozygosity and large number of alleles (20 or more) can inflate estimates of F_{ST} (Bird *et al.* 2011). To evaluate this effect, genetic differentiation between samples was also estimated with G_{ST} (Nei 1973), G'_{ST} (Hedrick 2005) and D_{EST} (Jost 2008). G_{ST}, G'_{ST} and D_{EST} were calculated with GenAlEx 6.5 (Peakall & Smouse 2012). Genetic population structure was also examined using a Bayesian approach implemented in STRUCTURE 2.3.2.1 (Pritchard *et al.* 2000; Falush *et al.* 2003) and by the multivariate ordination method in DAPC (Jombart *et al.* 2010), implemented in the adegenet package (Jombart 2008) in R 2.15.3 (R Development Core Team 2009; http://www.r-project.org).

Runs in STRUCTURE were made assuming $K = 1–10$ and imposing an admixture model with correlated allele frequencies. Each K value was replicated with ten independent runs of 10^6MCMC iterations, after a burn-in of 10^5 iterations. A set of STRUCTURE simulations were made using i) all loci, ii) only neutral loci and iii) only candidate outlier loci. These simulations were repeated with (LocPrior mode) and without Prior information. When Prior information was included, the following clustering criterion (based on geographical partition of data) was assumed: Cluster 1 (northwest Adriatic) [SLO+NAD+ANC+PEA+PEB]; Cluster 2 (southwest Adriatic) [BAA+BAB]; Cluster 3 (southeast Adriatic) [KOT+MNA+MNB]; Cluster 4

(northeast Adriatic) [RIJ+JAB+DUG] and Cluster 5 (Tyrrhenian Sea) [SPE+CDG]. The Evanno method (Evanno *et al.* 2005) in STRUCTURE HARVESTER (Earl & von Holdt 2012) was used to determine the most likely *K* value from all simulations and graphical representation of multiple runs per *K* values were produced by the CLUMPAK Server (http://clumpak.tau.ac.il) (Kopelman *et al.* 2015).

DAPC (Jombart *et al.* 2010) is a multivariate approach using principal components of genetic variation (PCA) that maximizes differences among groups while minimizing differences within groups. A set of 150 principal components was retained as predictors for discriminant analysis among samples.

The geographic boundaries among sampled populations were identified with BARRIER 2.2 (Manni *et al.* 2004), which uses Monmonier's algorithm to identify boundaries characterized by an abrupt break in gene flow (genetic barriers) and plots these boundaries on a map with sample coordinates. The strengths of putative boundaries were evaluated using a bootstrapping procedure separately for each locus.

Contemporary gene flow was evaluated with an individual-assignment test in GENECLASS 2.0 (Piry *et al.* 2004). For each locality, the criterion of Rannala & Mountain (1997) was applied using 10,000 individuals and the Paetkau *et al.* (2004) algorithm to detect first generation migrants.

The second method used to detect gene flow and connectivity was based on the coalescent calculation of historical migration rates between sampling locations using MIGRATE-n 3.5.1 (http://popgen.sc.fsu.edu/Migrate/Migrate-n.html). A Bayesian method was applied (Beerli 2006) and F_{ST} estimates among localities were used as baseline to calculate the mutation parameter Θ ($\Theta = 4N_e\mu$, where N_e is the effective

population size and μ is the locus mutation rate) and the historical migration rate, M ($M = m/\mu$, where m is the immigration rate per generation). A Brownian motion model was used and mutation rates were assumed to be constant over a set of 1000 neutral loci. The MCMC procedure consisted of one long chain of 500,000 genealogies for each locus, with the first 10,000 genealogies discarded as burn-in. The sum of M values from each location was used to determine directionality of gene flow.

The relationship between genetic and geographic distances between samples (isolation by distance) was tested using Mantel's difference matrix test implemented in Genepop (ISOLDE) (Rousset 2008). Genetic distances (Θ_{ST}) between samples were compared to linear shortest sea-distances between samples with a permutation test of 10,000 iterations.

Non-neutral loci were identified with four methods that are based on different statistical assumptions. First, the coalescent approach in LOSITAN (Antao *et al.* 2008) detects outlier loci from the joint distribution of F_{ST} and expected heterozygosity (H_E) under the island model of migration (Beaumont & Nichols 1996). Runs were made with both the Infinite Allele Mutation Model (IAM) and Stepwise Mutation Model (SMM) each with 100,000 simulations. A second, Bayesian approach implemented in BayeScan 2.0 (Foll & Gaggiotti 2008), uses a Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) algorithm to obtain posterior distributions and is useful for small sample sizes because it reduces bias. We used default parameters of 20 independent runs with 5000 iterations each, 50,000 iterations burn-in, sample sizes of 5000 and thinning of 10. BayeScan results were inferred following the Jeffrey's interpretation, as described in the BayeScan User

Manual. Third, a hierarchical finite island model (Excoffier *et al.* 2009) in Arlequin 3.5.1.3 (Excoffier & Lischer 2010) was used to identify potential candidate loci by partitioning the dataset into population clusters as inferred from the STRUCTURE results. This method is particularly efficient for species showing significant population structure, which reduces false positives in the hierarchical analysis of genetic differentiation. Replicate coalescent runs of 100 demes were performed with 20,000 iterations. Only loci identified as outliers in all three approaches were considered as true outliers. Forth, we used the LnHR method (Kauer *et al.* 2003), which is specifically designed to detect outlier microsatellites by evaluating changes in H_E between pair of samples under the following formula:

$$LnRH_{popA\text{-}popB} = Ln\,[(1/(1-He_{popA}))^2 - 1/(1/(1-He_{popB}))^2 - 1]$$

We used a threshold of ±2.58 and assumed that LnRH values under neutrality were normally distributed after data standardization (mean = 0, SD = 1). The values of 99% of neutral loci are expected to fall within this range.

A set of 13 environmental variables (listed in Table S2) was recovered from the Sea Data-Net Climatologies Pan-European Infrastructure for Ocean and Marine Data Management web site (http://gher-diva.phys.ulg.ac.be/web-vis/). In this database, chlorophyll and oxygen values were reported as averages over 1890 and 2008, whereas salinities and temperatures were reported as averages between 1900 and 2009. These values were used as independent variables in a redundancy discriminant analysis (RDA) implemented in CANOCO 4.5 (Ter Braak & Šmilauer 2002). These ordinations were compared to the allelic-frequencies distributions of outlier loci to test for allele-environment associations. Alleles with frequencies less than 5% were removed to reduce noise, and the environmental data series was tested for a neutral

distribution by a Shapiro-Wilk test. Non-neutral variables were Log_{10} or Ln transformed for seascape analyses. The forward selection of environmental variables at each locus was performed with a Monte Carlo permutation test of 999 iterations ($P < 0.01$). When a locus was associated with more than one variable, a permutation test as above was repeated to establish the conjunct significance of these variables on the distribution of allele frequencies. The associations between allele frequencies in samples and environmental variables were further investigated with a Canonical Correspondence Analysis (CCA) in PAST 2.17 (Hammer *et al.* 2001) with 10,000 permutations to determine significance.

Detection of multiple anchovy species in the Adriatic Sea

Based on the assumptions of Borsa *et al.* (2004, the most important feature distinguishing shallow-water silver anchovies (*E. albidus sensu* Borsa *et al.* 2004) from open-water blue anchovies (*E. encrasicolus*) is bathymetry. Hence, we collected samples from shallow coastal sites (<50m) and from open-water areas (>50m). We first performed a STRUCTURE simulation to test the effect of bathymetry. Two major groups of samples were defined based on the threshold of 50 m of the bottom depth and were used as *a priori* information. Cluster 1 included SLO+NAD+ANC+PEA+KOT +BAA+CDG, and Cluster 2 included the remaining samples. Second, we verified the consistency of clustering samples based on their sampling bathymetry (as performed in STRUCTURE) using an AMOVA test. Finally, estimates of genetic differentiation between samples were used to test for large genetic differences, as expected between divergent species.

Results

Microsatellite variability

The PCR success rate was high for the 14 microsatellite loci with only 1.17% missing genotypes. Although allele dropout and stuttering were minimal, null alleles were detected in 50 of 210 global tests. Null-allele corrections with the Brookfield algorithm improved the results, but 26 of 210 tests were still significant. Ten of the significant tests were for Enja148, which we excluded from the following analyses.

Locus polymorphism varied from 11 alleles at Ee2-508 to 41 alleles at Ee2-507 (Table S3). The mean number of alleles (N_A) per sample over all loci varied from 11.4 alleles in JAB to 15.8 alleles in KOT. Allelic richness (R_S), standardized with 19 individuals per sample, varied from 8.1 (DUG) to 11.2 (CDG) (Table 1; Table S3). The smallest values of H_E and H_O were in samples NAD and RIJ, whereas the largest values were in samples SLO and KOT (Table 1; Table S3). The smallest values of H_E and H_O by locus were in Ee2-165b and Ee2-508, respectively, and the largest values were in loci Ej2 and Ee2-507, respectively (Table 2).

Significant departures from the Hardy-Weinberg (HW) proportions over all loci were found in 4 samples, MNA, MNB, BAB and ANC. F_{IS} was significant in 4 of 210 tests, and three values of F_{IS} indicated homozygote excess (Table S3). Three loci deviated significantly from HW proportions over all populations (Ej41-1, Eja17 and Ee2-508) (Table 1). Only 4 of 1170 tests (78 pairwise tests X 15 localities) showed significant linkage disequilibrium, which occurred between EJ35 x Ee2-507 and Enja83 x Ee2-507 in the NAD sample, EJ27.1 x EJ35 in the DUG sample and Ee2-507 x Eja17 in the RIJ sample.

Microsatellite genetic population structure

Overall, genetic differentiation between samples was estimated with 5 statistics. First, pairwise values of Θ_{ST} and F_{ST} over loci showed similar patterns. Θ_{ST} ranged from -0.004 (BAA *vs*. PEB) to 0.054 (NAD *vs*. JAB), and F_{ST} ranged from -0.004 to 0.058 between samples. Pairwise Θ_{ST} were significant ($P < 0.0005$) in 51 of 105 comparisons (Table 3). G_{ST} ranged from -0.002 (BAA *vs*. PEB) to 0.057 (NAD *vs*. JAB) between samples, G'_{ST} ranged from -0.002 to 0.030, and D_{EST} values ranged from -0.004 (BAA *vs*. PEB) to 0.194 (NAD *vs*. JAB). Pairwise comparisons were significant ($P < 0.0005$) in 65 of 105 tests for G_{ST} and G'_{ST} and in 66 of 105 tests for D_{EST} (Table S4).

STRUCTURE simulations of neutral loci, with and without LocPrior information, yielded two major clusters (Fig. S1A, S1B and S1E): i) DUG+JAB+RIJ (Northeast Adriatic samples) and ii) the remaining samples. However, a high probability of three major clusters emerged using only candidate outlier loci without the LocPrior information (Fig. S1C) and using all loci with the LocPrior function (Fig. 2A; Fig. S1D). These three clusters were largely represented by i) DUG+JAB+RIJ (Northeast Adriatic samples), ii) NAD (Northern Adriatic sample) and iii) the remaining samples. Other possible substructures yielding 5 and 8 clusters appeared in the simulations, but with low *q* values (Fig. S1A–S1F).

In the DAPC analysis, the first 2 of the 150 Discriminant Axes (DA) were significant. These DA's defined three major genetic Clusters: i) DUG+JAB+RIJ (Northeast Adriatic samples), ii) NAD (Northern Adriatic sample) and iii) all remaining samples (Fig. 2B). The first DA separated mainly the Northeast Adriatic

Cluster from the NAD Cluster, while the second DA largely separated those two Clusters from a third, which included the remaining samples (Fig. 2B).

A total of 38 of 531 individuals were identified as first generation migrants. The analysis of migration rates and first generation migrants between clusters revealed a higher intra-cluster migration than inter-cluster migration in the two or three genetic groups. The directionality of gene flow was largely from populations MNA and ANC (Table S5A). Estimates of Θ from the historical gene flow ranged between 0.003 in CDG to 0.046 in JAB (Table S5B). Migration rates (*M*) ranged between 18.50 ($M_{JAB \rightarrow NAD}$) and 944.01 ($M_{DUG \rightarrow PEA}$). The directionality of historical gene flow suggested a pattern that was similar to sea-surface circulation (Fig. 1B). The main sources of migrants were DUG, ANC and CDG (Table S5C).

Additional approaches resolved other features of geographic structure. First, BARRIER identified at least three major allele-frequency discontinuities, indicating a heterogeneous pattern of partial isolation between locations. These results indicated two strong barriers, one between NAD and nearby sampling areas (Fig. 1B, barrier a, dotted line) and a longitudinal barrier across the Mid Adriatic Pit that divided Adriatic anchovies into east-west groups (Fig. 1B, barrier b, solid line). In addition, a north-south barrier that divide the Mid Adriatic Pit was detected (Fig. 1B, barrier c, dashed line). This barrier was stronger in the east than in the west.

Even though differences appeared between populations, no isolation-by-distance over the length of the Adriatic was detected with Mantel's test between pairwise Θ_{ST} and geographical distances ($P > 0.05$; a = 0.0165172, b = 0.00000108) (data not shown).

Outlier loci and environment-allele frequency correlations

The methods used to identify outlier loci showed slightly different results, but together identified five candidate outlier loci among the 13 analyzed, including Ee2-91b, Ee2-407, Ee10, Ee2-507 and EJ2 (Table 4). Fdist in Lositan identified seven outlier loci: three candidate loci for directional selection (Ee10, EJ35, Ee2-507) and four for balancing selection (Ee2-91b, Ee2-407, EJ2 and Ee2-135) (Fig. S2; Table 4). BayeScan indicated highly significant probabilities for outlier candidate loci for balancing selection (negative alpha values) for all loci, with the exception of EJ35 ($P > 0.05$) (Table 4), but q values were under the 10% threshold for FDR (False Discovery Rate) correction. The Hierarchical Island Method (HIM) for clustering in STRUCTURE $K = 2$ postulated six outliers, with two loci (Ee10 and Ee2-507) as candidates for directional selection and four loci (Ee2-91b, Ee2-407, EJ2 and Ee2-165b) for balancing selection (Fig. S3A; Table 4). With $K = 3$, HIM indicated that five of these six loci (except Ee2-165b) remained significant, showing the same pattern of selection as tests with $K = 2$ (Fig. S3B; Table 4). Finally, the LnRH test yielded 42 of 1365 pairwise combinations of locations and loci with values over the ±2.58 threshold. These tests were associated to two loci: Ee2-507 for 31 tests and EJ27.1 for 11 tests (Table S6).

The Redundancy Discriminant Analysis (RDA) identified significant correlations between allelic frequencies at four outlier loci (Ee2-507, Ee2-407, Ee10 and EJ2) and four environmental variables, including temperature at sampling depth (T_{samp}), salinity at sampling depth (S_{samp}) and at 10 m (S_{10}) and dissolved oxygen concentration at 10 m ($Oxyg_{10}$). At Ee2-507, alleles 279 and 291 were positively correlated and alleles 259 and 283 were inversely correlated with S_{10} (both $P <$ 0.019) (Fig. 3A). Significant correlations appeared for Ee2-407 with

both $Oxyg_{10}$and S_{samp} ($P <$ 0.002); allele 169 varied positively and allele 147 inversely with S_{samp}, and alleles 163 and 171 were positively and allele 159 was inversely correlated with $Oxyg_{10}$ (Fig. 3B). At locus Ee10, alleles 224, 232 and 258 were positively and alleles 226 and 256 were inversely correlated with T_{samp} ($P <$ 0.003) (Fig. 3C). Finally, at EJ2, alleles 196, 200 and 210 were positively and alleles 178 and 186 were inversely correlated with T_{samp} (P < 0.033) (Fig. 3D).

Both the first and second axes of the CCA ordination of samples with four environmental variables were significant ($P < 0.05$ and $P < 0.001$, respectively). The results indicated a positive association for allele frequencies in EMB, EMC and EMD and a negative association for allele-frequencies in sample EMA with salinity S_{samp} and S_{10} (Fig. 4). Samples SLO, KOT and PEA were positively and samples SPE and CDG from the Tyrrhenian Sea were negatively associated with temperature at sampling depth (T_{samp}). Finally, SPE and CDG were directly associated with dissolved oxygen content in the first 10 m ($Oxyg_{10}$) and inversely associated with T_{samp}.

Detection of multiple species of anchovies in the Adriatic Sea

The STRUCTURE analysis using the LocPrior function to cluster sampling sites by bathymetry showed three genetic clusters, as in the other STRUCTURE analyses (Fig. S1G). In general, the pattern did not show a clear association with bathymetry as postulated by Borsa *et al.* (2004). Furthermore, the AMOVA test partitioned by ocean depth of 50 m, indicated that the largest variance was among individuals within populations (97.86% of total variance; Table S7). Although variation among populations within groups (F_{SC} = 0.0017) was significant (P < 0.05), it explained

only a small amount of the total variation (1.79%). Divergence among groups (F_{CT}= 0.0012) was not significantly different from zero ($P > 0.05$).

Discussion

The results of our survey of microsatellite DNA variability among populations show that while anchovies in the Adriatic are genetically subdivided among regions to some extent, none of these populations represent separate species, as previously proposed. The mechanisms producing divergence between populations include both demographic processes influenced by ocean currents and fronts and natural selection that leads to locally adapted populations. Before exploring these factors in more detail, we offer the following evaluation of the study. While the sample sizes used in this study were not large, they provided sufficient statistical power to detect population differences. Statistical power was enhanced by the use of a large number of microsatellite loci encoded in nuclear genes. These markers provided complementary insights into population structure and local adaptation. We additionally used several approaches to the analyses of the microsatellite data that illuminated various dimensions of population structure.

How many anchovy species in the Adriatic Sea?

Borsa (2002) and Borsa *et al.* (2004) postulated that two species of anchovy inhabit the Adriatic Sea on the basis of north-south allele-frequency gradients reported by Bembo *et al.* (1996) and on morphological parallels with two putative anchovy species in the Gulf of Lyon (Borsa 2002; Borsa *et al.* 2004). The recently described estuarine species, *E. albidus* (Borsa *et al.* 2004), resembled shallow-water anchovies

in low-saline areas of the northern Adriatic (Borsa *et al.* 2004). The larger, bluish anchovies in the Adriatic resembled European anchovies elsewhere in open, deep waters with salinities greater than 38 psu, and were postulated to be *E. encrasicolus* (Borsa *et al.* 2004). However, our analysis of nuclear markers in anchovies from throughout the Adriatic did not show genetic differentiation between coastal and deep-water populations that could be interpreted to indicate multiple species.

Nevertheless, it might be argued that the genetic differences in the northern Adriatic Sea support the occurrence of *E. albidus* in the low-saline plume of the Po River (NAD). Although these anchovies share some ecological and morphological traits with putative *E. albidus* in the Gulf of Lyon, Po River anchovies are genetically most similar to oceanic anchovies in the central Adriatic. The level of genetic differentiation between NAD and the other samples suggests a substantial genetic separation, but this level of divergence is not large enough to indicate the presence of two divergent species. The color and morphological variability observed among anchovy populations in the Adriatic Sea (and in the Gulf of Lyon) likely reflects plastic phenotypic responses to environmental variability in anchovies that are typical of anchovies in the Mediterranean Sea (Tortonese 1970).

Genetic variability among populations

The small values of F_{ST} and Θ_{ST} between Adriatic populations are typical of divergence between other population groups of anchovies or between populations of other pelagic fishes. The overall levels of genetic differentiation between populations in the Adriatic Sea were similar to those between anchovy populations in the eastern

Atlantic (Zarraonaindia *et al.* 2009; Borrel *et al.* 2012). As in many marine species, gene flow in Adriatic anchovies occurs by larval dispersal in currents, or by adult movement at low levels over long periods, and prevents the accumulation of genetic differences among populations (Waples 1998). Nevertheless, significant differences between some samples indicate that Adriatic anchovies are not entirely panmictic.

The analysis of neutral makers with STRUCTURE resolved two major population groups: populations in northeastern coastal areas (DUG, JAB and RIJ) and the remaining populations in the Adriatic and Tyrrhenian seas. The genetic uniqueness of northeastern anchovies may be due to adaptation to habitats in the numerous coves and embayments and around offshore isles (Artegiani *et al.* 1997a, 1997b; Giannoulaki *et al.* 2013). The northeastern Adriatic appears to be conducive to spawning and early larvae growth and is an area where spawners converge (Vučetić 1964).

A north-south genetic discontinuity in the northern Adriatic Sea was previously reported by Bembo *et al.* (1996), who found an allozyme-frequency gradient between northern and southern populations. Bembo *et al.* (1996) focused on the possible role of depth, defining the 40–50 m sill as the population boundary, rather than on other environmental variables. In our analyses, depth did not appear to be an important isolating variable, as we found no genetic partitioning of microsatellite variability by ocean depth in areas where samples were collected.

Instead, the oceanic front between the Istrian Peninsula and Mid Adriatic Pit may explain, in part, the north-south genetic discontinuity by limiting gene flow. The anticlockwise gyre off the Po River delta may also act as a barrier to gene flow (Artegiani *et al.* 1997a, 1997b). Although both the front and the gyre are ephemeral,

they are strongest in autumn (Artegiani *et al.* 1997a, 1997b) at the end of the anchovy spawning season when larvae and juveniles are most abundant (Morello & Arneri 2009; Giannoulaki *et al.* 2013). Larval retention by ocean fronts and gyres is a well-known isolating mechanism in other pelagic fishes (Illes & Sinclair 1982; Warner & Cowen 2002). Hence, these oceanic mechanisms may limit the movement of larvae and juveniles and prevent mixing between larvae and juveniles from neighboring spawning sites. The effect of small coastal eddies limiting the dispersals of early life-history stages was also invoked by Borrell *et al.* (2012) to explain genetic differences between nearby populations in the Bay of Biscay.

Individual assignment showed that about 10% of the fish in our samples represented first-generation migrants from other populations. However, patterns of dispersals revealed by these tests have to be interpreted with caution, because small samples sizes and finite geographical sampling introduce some randomness into the results. The results also represent only a snapshot of a dynamic process that may change over time. Nevertheless, the results indicate that dispersals between populations are asymmetrical: some populations act as 'sources' and others act as 'sinks'. The extent that populations receiving migrants depend on immigration to persist has to be evaluated with long-term abundance data. With these caveats, most migrants originated from only a few source locations (mostly MNA and ANC samples) and moved to several nearby sink locations following a pattern of dispersion that is similar to major sea-surface currents in the area. The evaluation of historical patterns of gene flow using Migrate-n suggests a similar scenario for both the magnitude and pattern of connectivity. This pattern of migration may also be related to the lack of divergence between Adriatic and Tyrrhenian anchovies because of reciprocal gene

flow between the two seas. These results are in line with a recent analysis using SNPs markers showing genetic similarities between Adriatic and Western Mediterranean population of anchovies (Zarraonaindia *et al.* 2012).

Local adaptation

While the small genetic distances between samples indicate high levels of gene flow, as expected for a pelagic species, the pattern of geographic differentiation and the associations of microsatellite loci with environmental variables indicate that gene flow may be countered by adaptation to local conditions. For example, the timing of spawning, relative to seasonal bursts of productivity can have a strong effect on year-class survival (Biktashev *et al.* 2003; Irigoien *et al.* 2009). The persistence of genetic population structure in Adriatic anchovies may indicate differential reproductive success or reduced fitness of migrants reproducing outside their natal spawning areas. If so, adaptation to local habitats explains some of the genetic population structure observed in Adriatic anchovies.

Our inferences about natural selection are based on non-coding microsatellite loci, which are unlikely to be the direct focus of selection, except when microsatellite alleles modify the functions of genes in which they are embedded (Schlötterer 2000; Ellegren 2004). Alternatively, microsatellites may show signals of selection when they are linked to genes under selection. The strength of a signal from this 'hitchhiking' effect depends on the strength of selection, recombination rates and genetic drift (Maynard Smith & Haigh 1974; Vasemägi *et al.* 2005). Our analyses identified 5 of 13 loci (38.46%) as candidate outliers, and two loci under directional selection (Ee10 and Ee2-507) (15.38%), indicating the potential role of natural

selection in driving differentiation among populations. The proportion of candidate outliers was large, but consistent with other studies that showed between 5–15% outlier loci under directional selection (Meier *et al.* 2011; Russello *et al.* 2012).

Several life-history traits can interact to facilitate local adaptation. For instance, it is well known that this species has a high skew in reproductive success, which determines the level of genetic diversity in each cohort. Sweepstakes reproduction can promote selection at local scales, especially when coupled with large demographic variation in census size. During population bottlenecks, the reduction in effective population size (N_e) can produce transient genetic sweeps that lead to locally adaptive genotypes (Charlesworth 2009). A recent study showed that the effective sizes of anchovy populations in the Adriatic are several orders of magnitude smaller than census sizes (Ruggeri *et al.* 2016). These small effective sizes appear to be in part a legacy of a severe reduction in population abundances from overharvesting in the 1980s. Small effective population sizes in anchovies suggest that genetic drift, together with repeated selective sweeps, can mold the anchovy genome, resulting in local adaptation to major environmental drivers.

To further investigate the molecular footprints of selection, we tested candidate outlier loci for associations with environmental variables suspected to be important in the physiologies and ecologies of local anchovy populations. Four of the five candidate outliers were significantly correlated with environmental variables. Interestingly, these variables, especially salinity and temperature, influence larval development, survival and growth (Morello & Arneri 2009; Giannoulaki *et al.* 2013), as well as, adult maturation and reproductive success (Zorica *et al.* 2013). Associations with salinity and temperature variation explained patterns of genetic

variation. The northeastern group of anchovies showed a positive association with higher salinities, whereas anchovies in the Po River Plume showed an inverse association with salinity. The waters in these adjoining areas showed the largest difference in salinity among the sites sampled in our study, with the Po River plume having the lowest salinity, and waters of the eastern Adriatic along the northern Croatian coast having the highest salinities in the whole Adriatic Sea (Artegiani *et al.* 1997a, 1997b). Natural selection from strong differences in salinity among some areas may explain the genetic divergence between populations in the northern Adriatic Sea, despite substantial levels of gene flow between these areas.

In the CCA analysis, a set of samples from shallow Adriatic coastal waters (SLO, KOT and PEA) were positively associated with temperature variability, whereas samples from the Tyrrhenian Sea (SPE and CDG) were inversely associated with temperature variability. This is consistent with high seasonal fluctuations in temperature in northern and western Adriatic coastal areas, in contrast to greater thermal stability in the Tyrrhenian basin (Giannoulaki *et al.* 2013; Cognetti *et al.* 2000). Temperature also appears to be an important selective factor among populations of anchovies in the Eastern Atlantic (Silva *et al.* 2014).

Adriatic anchovies are also influenced by variability in dissolved oxygen, which is inversely correlated with temperature, because oxygen saturation in water decreases at higher temperatures. Populations in the Tyrrhenian Sea (SPE and CDG) were positively associated with dissolved oxygen content, but coastal populations in shallow waters (SLO, KOT and PEA) were inversely associated with oxygen content. The dissolved oxygen content may interact with temperature to define local selective constraints. Higher water temperatures in summer generally lead to a drop

in dissolved oxygen, producing hypoxic and anoxic conditions (Giani *et al.* 2012), especially in the Adriatic Sea.

Genetic discontinuities among populations of anchovies may be related to the effects of environmental variables on the growth and survival of early developmental stages, as well as on adult reproductive success. The complex genetic patterns among populations of marine organisms have only recently been correlated with the influence of environmental variables on the genomes of several marine species (Riginos & Liggins 2013). Numerous studies show that environmental variables affect adaptive responses in marine fishes on local (European hake, Milano *et al.* 2014) and oceanic scales (Atlantic anchovies, Silva *et al.* 2014; Bluefin Tuna, Riccioni *et al.* 2013).

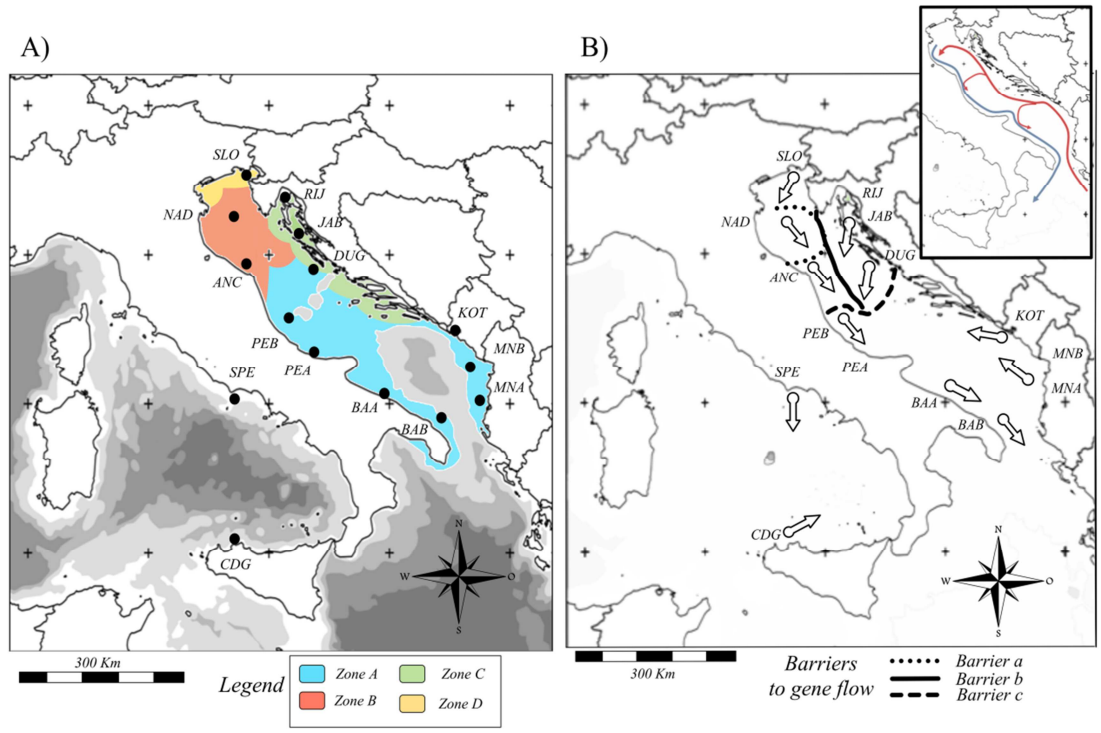
Conclusions

An important result of this study indicates that anchovies throughout the Adriatic belong to a single species, *E. encrasicolus*. Nevertheless, the individuals collected in the study area, which covered the entire Adriatic basin and part of the Tyrrhenian Sea, were partitioned into two major populations which were sometimes further sub-structured. Adriatic anchovy subpopulations show a patchy distribution of genetic variability that appears to reflect selective processes mediated by several environmental variables, including salinity, temperature and dissolved oxygen content. The relationship between environmental variables and the distribution of genetic differences among population groups is consistent with adaptation to local environmental conditions. Salinity, in particular, appears to be important in structuring northeastern Adriatic anchovy populations. Our geographical survey of

microsatellite DNA variability reveals a more complex genetic structure among Adriatic anchovy populations than was previously detected, but does not confirm the presence of a second species of *Engraulis*. Further resolution of this biocomplexity will provide a basis for understanding adaptive mechanisms and a basis for optimizing management strategies for sustainable harvests.

Figures and Tables

Fig. 1 Maps showing sample locations.



(a) Map of the Adriatic and Tyrrhenian Seas showing the sampling localities and a general representation of the four main areas described by (Artegiani *et al.* 1997a, 1997b). (b) Map showing outcomes from the BARRIER (Manni *et al.* 2004) and Migrate-n (Beerli 2006). The three lines (solid, dotted and dashed) represent major barriers to dispersal. The arrows represent the directionality of gene flow from each locality and a map (in the right up corner) of major sea-surface currents allows a direct comparison among them.

Fig. 2 Graphical outcomes of STRUCTURE simulations for K = 3 using all loci and the LocPrior function (barplots represent the individual q values) (Fig 2A). Graphical outcome of DAPC plot (Fig 2B).

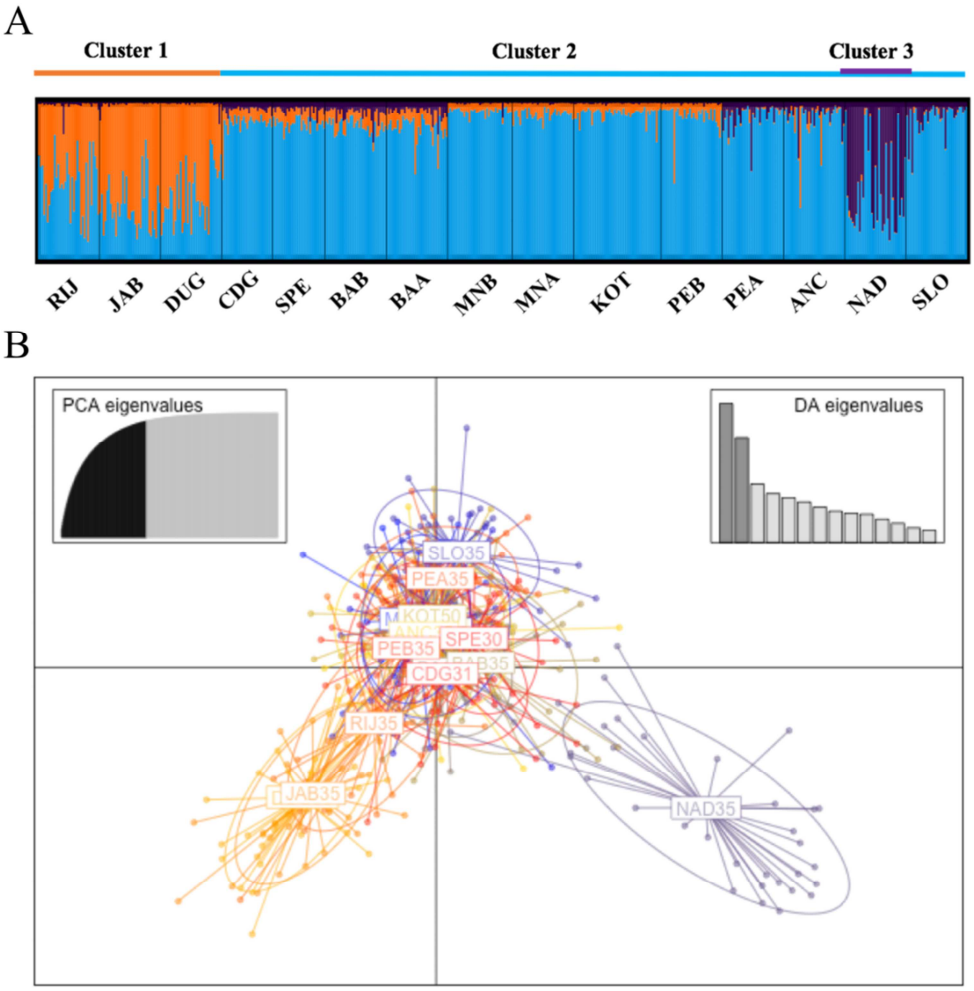
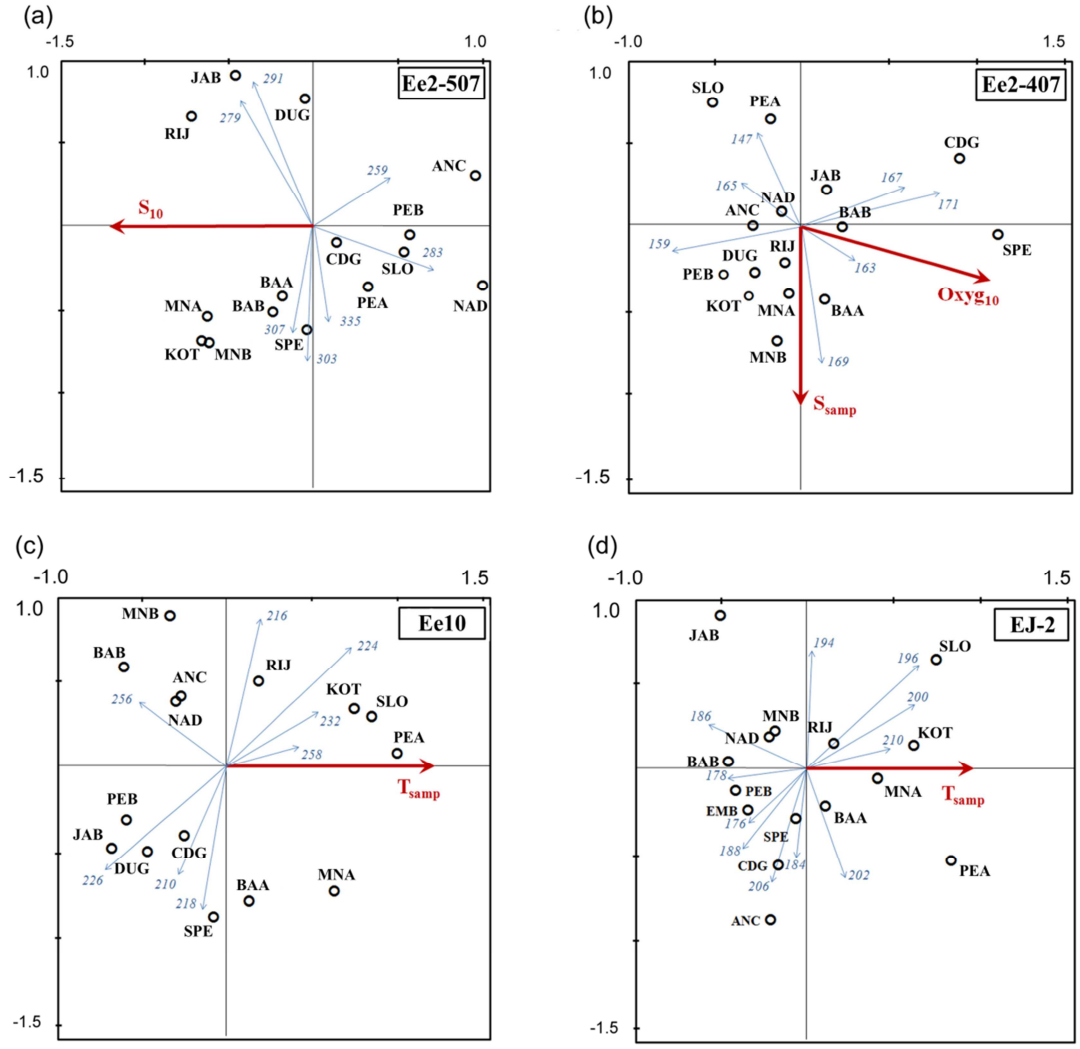
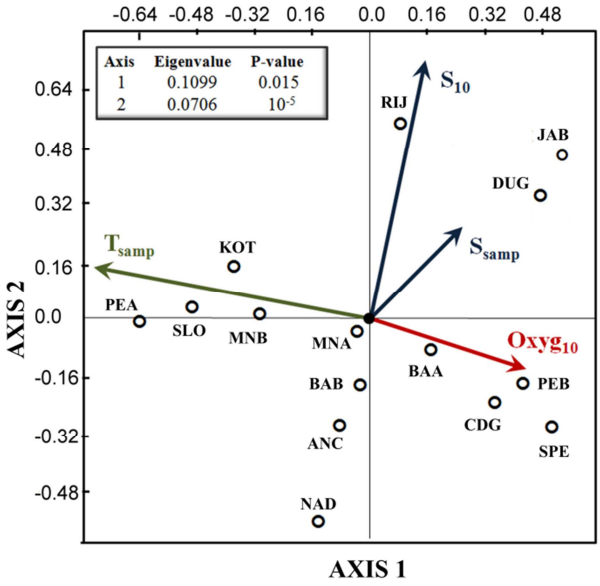


Fig. 3 Redundancy Discriminant Analysis (RDA) triplots.



Triplots show the distribution of samples (circles) relative to major significant environmental variables (dashed arrows) and microsatellite allele frequencies (solid arrows) at four candidate outlier loci. Triplots show results for Ee2-507 (A), Ee2-407 (B), Ee10 (C) and EJ2 (D), respectively.

Fig. 4 Canonical Correspondence Analysis (CCA) of microsatellites and environmental variability.



The plot shows the distribution of samples with candidate outlier loci that were correlated with the four environmental variables, *T*samp, *S*10, *S*samp, *Oxyg*10. Arrows identify the four environmental variables and their direction explain which samples are most correlated with them.

Table 1 Sample locations, dates, sample sizes and summary statistics by population.

Basin	Locality	Sample label	N	Coordinates	Sampling depth (m)	Bottom depth (m)	Sampling date (month/year)	Mean N_A	R_S	H_E	H_O
A	Slovenja-Piran	SLO	35	45°33.57'N; 13°35.64'E	12	22.5	Sep-12	14.2	9.8	0.826	0.791
A	Croatia- Rijeka	RIJ	35	45°16,53'N; 14°25,07'E	40	64	Sep-12	11.5	9.6	0.763	0.728
A	NorthernAdriatic Sea	NAD	35	44°49.97'N; 13°13.26'E	30	41	Sep-12	12.2	8.6	0.722	0.751
A	Croatia- Jablanac	JAB	35	44°41,47'N; 14°53,32'E	83	104	Sep-12	11.4	8.2	0.749	0.742
A	Croatia- DugiOtok	DUG	35	43°53,07'N; 14°55,66'E	61	75	Sep-12	12.2	8.1	0.763	0.776
A	Ancona	ANC	35	43°42.20'N; 13°37.76'E	34	38.4	Sep-12	13.5	8.4	0.818	0.752
A	Pescara	PEB	35	42°54.19'N; 14°10.31'E	58	70.8	Sep-12	13.9	9.3	0.818	0.78
A	Pescara	PEA	35	42°30.19'N; 14°18.14'E	8	24.1	Sep-12	13.9	9.8	0.811	0.781
A	BokaKotorska	KOT	50	42°25.55'N; 18°39.48'E	20	40	Jul-12	15.8	9.8	0.808	0.799
A	Montenegro	MNB	37	41°53.98'N; 19°00.96'E	75	86.7	Aug-12	13.9	9.7	0.821	0.756
A	Montenegro	MNA	35	41°49.69'N; 19°16.06'E	16	52.4	Aug-12	14	9.8	0.815	0.753
A	Bari	BAA	35	41°15.32'N; 16°34.13'E	23	31.6	Jul-12	12.8	9.2	0.809	0.758
T	Sperlonga	SPE	30	41°12.17'N; 13°23.85'E	40	80	Jun-13	12.5	10.7	0.792	0.75
A	Bari	BAB	35	41°10.60'N; 17°07.80'E	115	125.3	Jul-12	13.2	9.4	0.811	0.734
T	Castellammare del Golfo	CDG	29	38°04.81'N; 12°59.29'E	20	40	May-13	13.2	11.2	0.792	0.753

Basin, A = Adriatic Sea; T = Tyrrhenian Sea; N = Number of individuals sampled; Mean N_A = mean number of alleles; R_S = allelic richness; H_E = Expected heterozygosity by population; H_O = Observed heterozygosity.

Table 2 Summary statistics by locus.

Locus	Total N_A	Mean N_A	H_E	H_O	F_{IS}	*f* Null alleles	F_{ST}
Ee2-91b	15	8.7	0.806	0.767	0.048	0.014	0.0012
Ee2-407	38	16.5	0.86	0.805	0.064	0.023	0.0016
Ej41-1	25	11.2	0.69	0.632	**0.084**	0.030	0.0183
Ee10	37	18.5	0.838	0.778	0.072	0.025	0.0619
Ej27.1	36	22.9	0.922	0.869	0.057	0.019	0.0154
Ej35	25	12.2	0.852	0.851	0.001	-0.007	0.0341
Enja83	23	8.7	0.733	0.77	-0.050	-0.029	0.0171
Ee2-507	41	21.5	0.911	0.932	-0.023	-0.020	0.0398
Eja17	15	7.5	0.705	0.614	**0.129**	0.044	0.0096
Ej2	32	20.9	0.943	0.877	0.070	0.027	-0.001
Ee2-135	16	11.5	0.873	0.909	-0.041	-0.026	0.0048
Ee2-165b	12	5.3	0.574	0.558	0.028	0.004	0.0023
Ee2-508	11	6.7	0.672	0.522	**0.223**	0.082	0.0006
Enja148	16	5.3	0.534	0.365	**0.316**	**0.119**	0.0045

Total N_A = observed number of alleles; Mean N_A = mean number of alleles; H_E = Expected heterozygosity; H_O = Observed heterozygosity; F_{IS} = coefficient of inbreeding (bold values deviating from HW expectations); f_{Null} alleles = frequency of nulle alleles; F_{ST} = genetic differentiation estimated.

Table 3 Pairwaise multi locus estimates of θ_{ST} (below the diagonal) and F_{ST} (above the diagonal). Significant pairwise tests are in bold type.

	MNA	MNB	SLO	NAD	BAA	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG
MNA		0.010	0.015	0.033	0.000	0.003	0.009	0.005	0.020	0.026	0.018	0.009	0.003	0.005	0.001
MNB	0.009		0.010	0.035	0.013	0.009	0.007	0.005	0.035	0.037	0.017	0.006	0.017	0.023	0.018
SLO	0.014	0.010		0.037	0.016	0.009	0.004	0.013	0.041	0.047	0.021	0.005	0.019	0.030	0.030
NAD	**0.030**	**0.035**	0.035		0.026	0.020	0.035	0.032	0.051	0.058	0.041	0.030	0.031	0.038	0.029
BAA	-0.002	0.012	**0.015**	**0.023**		0.003	0.013	0.005	0.013	0.018	0.015	0.010	-0.004	0.005	0.000
BAB	0.002	**0.010**	0.009	**0.019**	0.002		0.005	0.003	0.021	0.028	0.013	0.003	0.005	0.008	0.007
KOT	0.009	0.006	0.005	**0.033**	0.011	0.005		0.008	0.026	0.031	0.011	0.002	0.016	0.022	0.017
ANC	0.003	0.004	0.012	0.028	0.004	0.003	0.007		0.019	0.026	0.016	0.007	0.005	0.014	0.009
DUG	0.018	**0.036**	**0.038**	**0.046**	0.011	**0.020**	**0.025**	0.017		-0.002	0.015	0.032	0.016	0.024	0.022
JAB	**0.025**	**0.038**	**0.046**	**0.054**	0.016	**0.027**	**0.031**	**0.026**	-0.002		0.013	0.037	0.023	0.027	0.028
RIJ	0.017	0.017	**0.020**	**0.038**	0.014	0.011	0.010	0.016	**0.016**	0.016		0.016	0.020	0.260	0.021
PEA	0.008	0.005	**0.004**	**0.029**	**0.009**	0.004	0.002	0.006	**0.031**	**0.038**	**0.015**		0.017	0.017	0.014
PEB	0.002	**0.017**	**0.019**	**0.028**	-0.004	0.004	**0.015**	0.004	**0.015**	**0.022**	**0.019**	**0.017**		0.008	0.006
SPE	0.003	**0.022**	**0.028**	**0.035**	0.003	0.007	**0.020**	**0.011**	**0.023**	**0.026**	0.024	**0.017**	0.007		-0.001
CDG	-0.001	0.017	**0.029**	**0.026**	-0.002	0.006	**0.016**	0.006	**0.020**	**0.027**	**0.020**	**0.014**	0.006	-0.002	

Table 4 Observed F_{ST} values from Lositan (Beamount *et al.* 1996; Antao *et al.* 2008), BayeScan (Foll & Gaggiotti 2008) (plus q values) and the Hierarchical Island Model (HIM) (Excoffier *et al.* 2009) to test for non-neutral loci.

Locus name	**Obs F_{ST} Lositan**	**Obs F_{ST} BayeScan**	**Obs F_{ST} HIM K=2**	**Obs F_{ST} HIM K=3**
Ee2-91b	0.0026**	0.0130*** (0.000)	0.0021*	0.0017**
Ee2-407	0.0023***	0.0033*** (0.000)	0.0016*	0.0015**
Ej41-1	0.0216	0.0164*** (0.000)	0.0195	0.0233
Ee10	0.0661***	0.0269*** (0.000)	0.0698**	0.0659***
Ej27.1	0.0167	0.0103*** (0.000)	0.0400	0.0293
Ej35	0.0343**	0.0300 (0.000)	0.0448	0.0427
Enja83	0.0163	0.0205*** (0.000)	0.0299	0.0243
Ee2-507	0.0397***	0.0193*** (0.000)	0.0771***	0.0623***
Eja17	0.0111	0.0152*** (0.000)	0.0250	0.0162
Ej2	0.0001*	0.0027*** (0.000)	0.0003**	0.0004*
Ee2-135	0.0046**	0.0062*** (0.000)	0.0039	0.0040*
Ee2-165b	0.0036	0.0084*** (0.000)	0.0018*	0.0080
Ee2-508	0.0059	0.0185*** (0.000)	0.0099	0.0044

Level of significance: *P<0.05; **P<0.01; ***P<0.001. Loci showed in bold type represent those resulted significant in all selection tests.

References

Antao T., Lopes A., Lopes R.J., Beja-Pereira A., Luikart G. (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323–327.

Artegiani A., Bregant D., Paschini E., Pinardi N., Raicich F., Russo A. (1997) The Adriatic Sea general circulation. Part I: air-sea interactions and water mass structure. *Journal of Physical Oceanography*, **27**, 1492–1514.

Artegiani A., Bregant D., Paschini E., Pinardi N., Raicich F., Russo A. (1997) The Adriatic Sea general circulation. Part II: Baroclinic circulation structure. *Journal of Physical Oceanography*, **27**, 1515–1532.

Bassam B.J., Caetano-Anollés G., Gresshoff P.M. (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analytical Biochemistry*, **196**, 80–83.

Beaumont M.A., Nichols R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**, 1619–1626.

Beerli P. (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.

Bembo D.G., Carvalho G.R., Cingolani N., Arneri E., Giannetti G., Pitcher T.J. (1996) Allozymic and morphometric evidence for two stocks of the European anchovy *Engraulis encrasicolus* in Adriatic waters. *Marine Biology*, **126**, 529–538.

Biktashev V.N., Brindley J., Horwood J.W. (2003) Phytoplankton blooms and fish recruitment rate. *Journal of Plankton Research*, **25**, 21–33.

Bird C.E., Karl S.A., Smouse P.E., Toonen R.J. (2011) Detecting and measuring genetic differentiation. In: Koenemann S., Held C., Schubart C. (Eds), *Crustacean Issues: Phylogeography and Population Genetics in Crustacea*. CRC Press, Boca Raton, FL, USA: 31–55.

Borrell Y.J., Piñera J.A., Sánchez Prado J.A., Blanco G. (2012) Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES Journal of Marine Science*, **69**, 1357–1371.

Borsa P., Collet A., Durand J.D. (2004) Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean, *Comptes Rendus Biologies*, **327**, 1113–1123.

Borsa P. (2002) Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). *Biological Journal of the Linnean Society*, **75**, 261–269.

Brookfield J.F.Y. (1996) A simple method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Buljan M., Zore-Armanda M. (1974) Oceanographical properties of the Adriatic Sea. *Oceanography and Marine Biology: An Annual Review*, **14**, 11–98.

Chapuis M.P., Estoup A. (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evoultion*, **24**, 621–631.

Charlesworth B. (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.

Chistiakov D.A., Hellemans B., Volckaert F.A.M. (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, **255**, 1–29.

Chiu T.S., Lee Y.J., Huang S.W., Yu H.T. (2002) Polymorphic microsatellite markers for stock identification in Japanese anchovy (*Engraulis japonica*). *Molecular Ecology Notes*, **2**, 49–50.

Cognetti G., Lardicci C., Abbiati M., Castelli A. (2000) The Adriatic Sea and the Tyrrhenian Sea. In: Sheppard CRC (Ed), *Seas at the millennium: an environmental evaluation: 1. Regional chapters: Europe, The Americas and West Africa*. Elsevier Science Ltd., London: 267–284.

Dempster A.P., Laird N.M., Rubin D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, **39**, 1–38.

Earl D.A., von Holdt B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Ellegren H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**: 435–445.

Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L., Hofer T., Foll M. (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Excoffier L., Lischer H.E.L. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Falush D., Stephens M., Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Foll M., Gaggiotti O.E. (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977–993.

Gaggiotti O.E., Bekkevold D., Jørgensen H.B.H., Foll M., Carvalho G.R., Andrè C. *et al.* (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution*, **63**, 2939–2951.

Giani M., Djakovac T., Degobbis D., Cozzi S., Solidoro C., Fonda Umani S. (2012) Recent changes in the marine ecosystems of the northern Adriatic Sea. *Estuarine, Coastal and Shelf Science*, **115**, 1–13.

Giannoulaki M., Iglesias M., Tugores M.P., Bonanno A., Patti B., De Felice A. *et al.* (2013) Characterizing the potential habitat of European anchovy *Engraulis encrasicolus* in the Mediterranean Sea at different life stages. *Fisheries Oceanography*, **22**, 69–89.

Goudet J. (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html

Hammer Ø., Harper D.A.T., Ryan P.D. (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontolia Electronica*, **4**, 9.

Hedgecock D., Pudovkin A.I. (2011) Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, **87**, 971–1002.

Hedrick P.W. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hilborn R., Quinn T.P., Schindler D.E., Rogers D.E. (2003) Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences*, **100**, 6564–6568.

Illes T.D., Sinclair M. (1982) Atlantic herring: stock discreteness and abundance. *Science*, **215**, 627–633.

Irigoien X., Fernanders J.A., Grosjean P., Denis K., Albaina A., Santos M. (2009) Spring zooplankton distribution in the Bay of Biscay from 1998 to 2006 in relation with anchovy recruitment. *Journal of Plankton Research*, **31**, 1–17.

Jombart T., Devillard S., Balloux F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jost L. (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kauer M.O., Dieringer D., Schlotterer C. (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of Drosophila melanogaster. *Genetics*, **165**, 1137–1148.

Kelly R.P., Palumbi S.R. (2010) Genetic structure among 50 species of the Northeastern Pacific rocky intertidal community. *PloS One*, **5**, e8594.

Kopelman N.M., Mayzel J., Jakobsson M., Rosenberg N.A., Mayrose I. (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179–1191.

Lamichhaney S., Martinez Barrio A., Rafati N., Sundström G., Rubin C.J., Gilbert E.R. *et al.* (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, **109**, 19345–19350.

Landi M., Garoia F., Piccinetti C., Tinti F. (2005) Isolation of polymorphic microsatellite loci from the European anchovy, *Engraulis encrasicolus*. *Molecular Ecology Notes*, **5**: 266–268.

Lin L., Zhu L., Liu S.F., Su Y.Q., Zhuang Z.M. (2011) Polymorphic microsatellite loci for the Japanese anchovy *Engraulis japonicus* (Engraulidae). *Genetics and Molecular Research*, **10**, 764–768.

Manel S., Holderegger R. (2013) Ten year of landscape genetics. *Trends in Ecology and Evolution*, **28**, 614–621.

Manni F., Guérard E., Heyer E. (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". *Human Biology*, **76**, 173–190.

Maynard Smith J., Haigh J. (1974) The hitch-hiking effect of a favourable gene. *Genetics Research*, **23**, 23–35.

Meier K., Hansen M.M., Bekkevold D., Skaala Ø., Mensberg K.L.D. (2011) An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta* L.): footprints of selection at microsatellite DNA loci. *Heredity*, **106**, 488–499.

Milano I., Babbucci M., Cariani A., Atanassova M., Bekkevold D., Carvalho G.R. *et al.* (2014) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, **23**, 118–135.

Morello E.B., Arneri E. (2009) Anchovy and sardine in the Adriatic Sea-an ecological review. *Oceanography and Marine Biology: An Annual Review*, **47**, 209–256.

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, **70**, 3321–3323.

Olsen E.M., Knutsen H., Gjøsæter J., Jorde P.E., Knutsen J.A., Stenseth N.C. (2008) Small-scale biocomplexity in coastal Atlantic cod supporting a Darwinian perspective on fisheries management. *Evolutionary Applications*, **1**, 524–533.

Paetkau D., Slade R., Burden M., Estoup A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Pakaki V., Magoulas A., Kasapidis P. (2009) New polymorphic microsatellite loci for population studies in the European anchovy, *Engraulis encrasicolus* (L.). *Molecular Ecology Resources*, **9**, 1406–1409.

Peakall R., Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

Piry S., Alapetite A., Cornuet J.M., Paetkau D., Baudouin L., Estoup A. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.

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

Rannala B., Mountain J.L. (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Acadamy of Sciences*, **94**, 9197–9201.

Riccioni G., Stagioni M., Landi M., Ferrara G., Barbujani G., Tinti F. (2013) Genetic structure of bluefin tuna in the mediterranean sea correlates with environmental variables. *PLoS One*, **8**, e80105.

Rice W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Riginos C., Liggins L. (2013) Seascape genetics: populations, individuals, and genes marooned and adrift. *Geography Compass*, **7**, 197–216.

Rousset F. (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Ruggeri P., Splendiani A., Di Muri C., Fioravanti T., Santojanni A., Leonori I. *et al.* (2016) Coupling demographic and genetic variability from archived collections of European anchovy (*Engraulis encrasicolus*). *PLoS One*, **11**, e0151507.

Russello M.A., Kirk S.L., Frazer K.K., Askey P.J. (2012) Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications*, **5**, 39–52.

Ruzzante D.E., Mariani S., Bekkevold D., André C., Mosegaard H., Clausen L.A.W. *et al.* (2006) Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proceedings of the Royal Society of London B: Biological Sciences*, **273**, 1459–1464.

Santojanni A., Arneri E., Bernardini V., Cingolani N., Di Marco M., Russo A. (2006) Effects of environmental variables on recruitment of anchovy in the Adriatic Sea. *Climate Research*, **31**, 181–193.

Schlötterer C. (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**, 365–371.

Selkoe K.A., Henkler C.M., Gaines S.D. (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, **9**, 363–377.

Selkoe K.A., Toonen R.J. (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, **436**, 291–305.

Selkoe K.A., Watson J.R., White C., Ben Horin T., Iacchei M., Mitarai S. *et al.* (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patters in three temperate reef species. *Molecular Ecology*, **19**, 3708–3726.

Silva G., Lima F.P., Martel P., Castilho R. (2014) Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. *Proceedings of the Royal Society of London B: Biological Sciences*, **281**, 20141093.

Taggart J.B., Hynes R.A., Prodöuhl P.A., Ferguson A. (1992) A simplified protocol for routine total DNA isolation from salmonid fishes. *Journal of Fish Biology*, **40**, 963–965.

Ter Braak C.J.F., Šmilauer P. (2002) CANOCO Reference Manual and CanoDraw for Windows user's Guide–Software for Canonical Community Ordination (Version 4.5): Microcomputer Power, Ithaca, New York.

Tortonese E. (1970) Fauna d'Italia. Vol. X. Osteichthyes, part 1. (Calderini Ed). Bologna, Italy: 565.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M., Shipley P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vasemägi A., Nilsson J., Primmer C.R. (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution*, **22**, 1067–1076.

Vučetić T. (1964) On the spawning of the anchovy (*Engraulis encrasicolus* L.) in the open waters of the Adriatic Sea (in Croatian). *Acta Adriatica*, **11**, 277–284.

Waples R.S. (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438–450.

Warner R.R., Cowen R.K. (2002) Local retention of production in marine populations: evidence, mechanisms, and consequences. *Bulletin of Marine Science*, **70**, 245–249.

Weir B.S., Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

White C., Selkoe K.A., Watson J., Siegel D.A., Zacherl D.C., Toonen R.J. (2010) Ocean currents help explain population genetic structure. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 1685–1694.

Zarraonaindia I., Iriondo M., Albaina A., Pardo M.A., Manzano C., Grant W.S. *et al.* (2012) Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). *PloS One*, **7**, e42201.

Zarraonaindia I., Pardo M.A., Iriondo M., Manzano C., Estonba A. (2009) Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. *ICES Journal of Marine Science*, **66**, 2176–2182.

Zorica B., Vilibić I., Čikeš Keč V., Šepić J. (2013) Environmental conditions conducive to anchovy (*Engraulis encrasicolus*) spawning in the Adriatic Sea. *Fisheries Oceanography*, **22**, 32–40.

CHAPTER III

Ruggeri P., Splendiani A., Di Muri C., **Fioravanti T.**, Santojanni A., Leonori I., De Felice A., Biagiotti I., Carpi P., Arneri E., Nisi Cerioni P., Giovannotti M., Caputo Barucchi V. (2016) Coupling Demographic and Genetic Variability from Archived Collections of European Anchovy (*Engraulis encrasicolus*). *PLoS ONE*, **11** (3), e0151507. doi:10.1371/journal.pone.0151507

Coupling Demographic and Genetic Variability from Archived Collections of European Anchovy (*Engraulis encrasicolus*)

Abstract

It is well known that temporal fluctuations in small populations deeply influence evolutionary potential. Less well known is whether fluctuations can influence the evolutionary potentials of species with large census sizes. Here, we estimated genetic population parameters from as survey of polymorphic microsatellite DNA loci in archived otoliths from Adriatic European anchovy (*Engraulis encrasicolus*), a fish with large census sizes that supports numerous local fisheries. Stocks have fluctuated greatly over the past few decades, and the Adriatic fishery collapsed in 1987. Our results show a significant reduction of mean genetic parameters as a consequence of the population collapse. In addition, estimates of effective population size (N_e) are much smaller than those expected in a fishes with large population census sizes (N_c). Estimates of N_e indicate low effective population sizes, even before the population collapse. The ratio N_e/N_e ranged between 10^{-6} and 10^{-8}, indicating a large discrepancy between the anchovy gene pool and population census size. Therefore, anchovy populations may be more vulnerable to fishery effort and environmental change than previously thought.

Introduction

Population dynamics are driven by a complex set of ecological and evolutionary variables that mold population demography on several spatial and temporal scales (Schoener 2011). Unfortunately, a temporal dimension is not always considered in

the assessment of population structure, because historical samples are not alwaysavailable, or because of the high costs of retrospective laboratory analysis (reviewed in Nielsen & Hansen 2008). Among vertebrates, populations of bony fishes are most often examined for historical trends because of the availability of historical tissues (Speller *et al.* 2012), which are sometimes available from archeological excavations (Speller *et al.* 2012) and from archived fish scales and otoliths used for aging by fishery managers (Campana & Thorrold 2001). The extraction of DNA from these archived collections (archived DNA) offers the opportunity to investigate the temporal dynamics of fish populations (Nielsen & Hansen 2008).

Several key variables influencing the evolutionary potential of fish populations and conservation status can be evaluated using archived DNA (Habel *et al.* 2015). This approach has successfully detected the loss of genetic variability as a consequence of natural fluctuations in population census size (N_c) (Nielsen & Hansen 2008; Speller *et al.* 2012; Habel *et al.* 2014). Many marine fishes with large N_c are thought to be sheltered from rapid genetic collapse because of large population sizes, but recent genetic studies cast doubt on this conclusion (Ruggeri *et al.* 2012; O'Leary *et al.* 2013). Temporal genetic analyses of archived samples have also demonstrated that fishing can lead to the loss of allelic richness and heterozygosity in over-exploited marine populations, indicating that overharvesting can drive the loss of evolutionary potential (Pinsky & Palumbi 2014; Allendorf *et al.* 2014; Marty *et al.* 2015; Bonanomi *et al.* 2015). The analysis of archived DNA can provide estimates of effective population size (N_e), "the size of an ideal population (*i.e.*, with discrete generations, random mating, and constant population sizes) that would undergo the

same amount of genetic drift, measured by the rate of loss of heterozygosity, as the actual population" (Gaggiotti & Vetter 1999). N_e is a key conservation parameter, which represents the "genetic currency" of evolutionary potential of a population (Nielsen & Hansen 2008; Hare *et al.* 2011). N_e measures the extent of genetic erosion by genetic drift and, together with life history traits of a species, can predict future population viability (Hare *et al.* 2011).

Among marine species, small pelagic fish periodically experience rapid fluctuations in N_c (Grant & Bowen 1998; Ganias *et al.* 2014), but the mechanisms leading to the rapid loss of individuals are still uncertain (Alheit *et al.* 2014; Van Beveren *et al.* 2014). Many stocks of small pelagic fishes are the targets of intensive fishing throughout the world. Unfortunately, how fishing activities influence local fluctuations in N_c and the effect on the evolutionary dynamics of local populations are not fully understood (Pinsky & Palumbi 2014). The Mediterranean area is one of the most exploited in the world with over 80% of its stocks are facing overexploited (Colloca *et al.* 2013). In recent years, European anchovy (*Engraulis encrasicolus*, hereafter "anchovy") stocks were over-exploited in two Spanish GSAs (Geographical Sub-Areas) (GSA 1 and GSA 6) (Colloca *et al.* 2013), and signs of overfishing were found in the northern Adriatic Sea (GSA 17) (Carpi *et al.* 2015). Adriatic anchovy populations have experienced strong fluctuations over the past four decades (Santojanni *et al.* 2003, 2006; Leonori *et al.* 2011, 2012). A decline, beginning in 1978, led to a stock collapse in 1987, with a recent partial recovery (Morello & Arneri 2009). The 1987 stock collapse was preceded by low recruitment in 1986 and 1987 (Azzali *et al.* 2002). Recruitment failure has been suggested for the anchovy

stock collapse, but the effects of intense fishing pressure cannot be excluded (Santojanni *et al.* 2006a, 2006b).

Recent genetic studies of small pelagic fishes other than anchovies show that temporal variation in N_e and genetic diversity is often lower than previously believed (Ruggeri *et al.* 2012; Laurent & Planes 2007). These studies suggest that both fishing and other factors, such as environment shifts and interactions with other species, can lead to the loss of genetic diversity (Pinski & Palumbi 2014). The evaluation of genetic variables can help to understand the status and factors influencing the short-term standing variation of European anchovies in the Mediterranean Sea. The goal of this study is to assess temporal genetic variation in European anchovies with a survey of microsatellite DNA extracted from archived otoliths collected over the past four decades at two sites in the Adriatic Sea. The results were used to address two questions: i) whether the large fluctuations in N_c experienced by Adriatic anchovies produced a critical loss of genetic diversity that could lead to a loss in evolutionary potential; and ii) whether genetic population structure changed in response to the demographic fluctuations.

Materials and Methods

Samples studied

We used 408 archived anchovy specimens, each providing two otoliths. Samples were provided by the ISMAR–CNR of Ancona, Italy (Istituto di Scienze Marine–Consiglio Nazionale delle Ricerche) and were collected 1978–2000. After removal from individuals for age determination, otoliths were stored individually in plastic tubes at room temperature. We analyzed a time series of samples from two spawning

grounds, the northern Adriatic (Chioggia) from 1978, 1987, 1994 and 2000, and the middle-southern Adriatic (Vieste) from 1985, 1987 and 1989 (Fig. 1). Samples from these years were chosen to include the biomass fluctuations starting in the mid 1970s. This objective was easily fulfilled for samples from Chioggia, because of a more complete time series. Unfortunately, the Vieste time series was available for only 1985–1989. We used otoliths collected between May and September to include individuals collected only during the spawning period. These time series were compared with 73 contemporary fish collected in 2010 in surveys off Chioggia and Vieste. The number of specimens used for each sampling year and location appears in Table S1.

Ethics statement

Ethical procedures were not required to manipulate specimens in this study. Tissues used in this study were recovered from previously collected otoliths and from Italian commercial catches.

Genomic DNA extraction

Genomic DNA was extracted from tissue residuals on otolith following the method described in (Hutchinson *et al.* 1999). Each set of otoliths was incubated at 55°C (up to 5 hours) in 500 µL of digestion buffer (100 mmol·L^{-1} Tris–HCl, pH 8.0; 100 mmol·L^{-1} NaCl; 1 mmol·L^{-1} EDTA, pH 8.0; 0.5% SDS). EDTA and SDS concentrations in the digestion buffer were reduced according to Cuveliers *et al.* (2009). Fin or caudal muscle tissues from contemporary samples were used for DNA extraction by standard phenol–chloroform procedures described in Taggart *et al.* (1992).

PCR amplification and genotyping

Samples were screened at 7 microsatellite loci that yielded short alleles less than 200 bp, a feature enhancing the ability to amplify degraded DNA from archived tissues. Primers for 6 of these loci were obtained from the European anchovy genome described by Landi *et al.* (2005) (Ee10) and Pakaki *et al.* (2009) (Ee2-508, Ee2-165b, Ee2-135, Ee2-407 and Ee2-91b). Primers for the seventh locus (Eja183) was described by Lin *et al.* (2011) in the Japanese anchovy (*Engraulis japonicus*) genome (Table S2). However, new primers were needed to obtain molecular sizes smaller than those produced by the original primers. Hence, primer sequences for these loci, except Ee2-91b and Ee2-135, were re-designed using the online program Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/) (Rozen & Skaletsky 2000) from their NCBI (National Center for Biotechnology Information) clone sequence references (Table S2). Attempts to develop additional primers producing shorter alleles to increase the number of loci failed because sequence constraints did not allow the design of new primers.

PCR conditions were optimized for all loci using touchdown amplification. The PCR reaction mixture contained approximately 5–10 ng genomic DNA, 0.25 U Taq DNA polymerase (MyTaq, Bioline), 0.5 $\mu mol \cdot L^{-1}$ of each primer, and 1× MyTaq (Bioline) Reaction buffer (15 $mmol \cdot L^{-1} MgCl_2$, 1.25 $mmol \cdot L^{-1}$ of each dNTP, plus stabilizers and enhancers) in a total volume of 10 µL. Each PCR plate included a blank control, including an extra control from each DNA extraction. PCR products were separated on a 2% (w/v) agarose gel and stained with GelRed™ (Biotium, Inc.) to check for size and PCR specificity. Subsequently, these products were run on a 5% denaturing polyacrylamide sequencing gel and visualized by a silver staining protocol

(Benbouza *et al.* 2006). Genotyping procedures were described in Peakall & Smouse (2012).

Accuracy of extraction and genotyping archived DNA

Archived DNA is particularly prone to genotyping errors due to i) its degraded nature increases the risks for allele dropout and null alleles and ii) potential contamination from exogenous DNA. To address these issues, a standardized protocol was used for DNA extraction and genotyping. First, archived DNAs were extracted in an ancient-DNA room equipped with sterilized laboratory utensils regularly sterilized and supplies dedicated to analyzing archived materials. Secondly, DNA extractions were performed in clusters of individuals from the same locality and year. These precautions helped to avoid potential contamination of DNA between individuals from different years and sampling sites. To further reduce sample contamination, no contemporary anchovy samples were processed while working on archived samples.

Genotyping errors occur primarily during PCR amplification by increasing the number of DNA copies from non-target individuals. Hence, PCRs were carried out in a separate ancient-DNA room, furnished with PCR workstations, set of pipettes, reagents and thermal cyclers used exclusively to amplify archived DNA. PCRs were carried out in single tubes and individuals were amplified in clusters belonging to the same sampling site and year. Additional attention was paid to the number of individuals amplified per PCR; less than 1/4 of the capacity of each thermal cycler was used to maximize the space between tubes. Thermal cyclers were sterilized by UV light and by cleaning with 10–20% bleach before PCR.

Contamination was tested by screening PCR products from two microsatellite loci, Ee2-91b and Ee2-165b. These loci show clear allelic patterns and were used to eliminate individuals that showing multiple alleles indicating contamination (Table S1). Faint genotypes from uncontaminated individuals were initially entered as missing data. DNA from these individuals was amplified three times using a progressively stringent PCR protocol. This was accomplished by increasing the concentration of Taq Polymerase in each reaction (0.05 µl, 0.1 µl, 0.2 µl on a final volume of 10 µl) and by increasing the annealing temperature by 0.5°C per PCR. Genotypes were not clearly detectable after these three trials were excluded from the dataset. A set of 10 individuals were randomly selected from each site and genotyped again as quality controls.

Finally, the probability of identity, P_{ID}, and probability of siblings, $P_{ID(sib)}$, were used to test whether contamination biased genotype quality of our dataset. P_{ID}, the probability that two unrelated individuals in a dataset will have the same multilocus genotype by chance, and $P_{ID(sib)}$, an analogous that takes inbreeding into account, were calculated with GenAlEx 6.502 (Peakall & Smouse 2012). Since we analyzed years and sampling locations independently, we assumed that DNA contamination would only be by individuals in the same sample. Therefore, cross-contaminated DNA should share more common genotypes than non-contaminated DNA. Thresholds for detecting contaminated samples were set at $P_{ID} > 0.001–0.0001$ and $P_{ID(sib)} > 0.01$, as proposed by Waits *et al.* (2001).

Statistical treatment of data

We estimated genetic diversities from genotypic and allelic frequencies only after checking the quality of the genotypes (amplification success rate and genotype consistency) and the occurrence of null alleles and other genotyping errors (allele dropout and stutter peaks) using MICRO-CHECKER 2.2.1 (Van Oosterhout *et al.* 2004). Loci affected by null alleles were corrected following the Brookfield algorithm (Brookfield 1996). Mean number of alleles observed for each locus (N_A), allelic richness (R_S), observed (H_o) and expected (H_e) heterozygosities, and the inbreeding coefficient (F_{IS}) were estimated with FSTAT 2.9.3 (Goudet 2001). Linkage disequilibrium between loci was tested with a Monte Carlo Markov Chain (MCMC) test executed by 1000 batches of 2000 iterations each, using Genepop 4.0.10 (Rousset 2008). A sequential Bonferroni adjustment of P-values was applied to account for an increase in type-I error from multiple comparisons (Rice 1989). Outlier loci were detected by departures from the neutral expectations with *fdist* (Beaumont & Nichols 1996) implemented in Lositan (Antao *et al.* 2008) and a LnRH method developed for microsatellites (Kauer *et al.* 2003) (File S1).

Temporal variation in genetic diversity was quantified using estimates of expected (H_e) and observed (H_o) heterozygosity and the expected number of alleles (N_A). Since estimates of genetic diversity can be biased by differences in sample sizes, the software POPTOOLS 3.1.0 (Hood 2009) (http://www.cse.csiro.au/poptools) was used to standardize and analyze temporal trends in these variables. Estimates were standardized by generating 1000 'real' samples (n = 24) with sampling without replacement for each year and location (Chioggia and Vieste). A total of 9000 “real” samples were generated. In addition, 9000 “randomized” samples were generated pooling together the samples from the Chioggia and Vieste time series. Statistical

significances of temporal trends were estimated by calculating the slope (b) and the Pearson's correlation coefficient (r) of linear regressions of H_e, H_o, and N_A against years. The r and b estimates of "real" parameters for each year in both the Chioggia and Vieste time series were compared with randomized values, using MCMC chains of 1000 iterations. We evaluated the significance of trends in H_e, H_o, and N_A over the entire time-series and between consecutive years.

Effective population size (N_e) was estimated using a moment-based temporal method and a coalescent likelihood-based method. These methods are generally robust and are often used on historical time series (Hare *et al.* 2011). Both temporal methods use changes in allele frequencies between two samples separated by a known number of generations. Since anchovies become mature at one year of age (Morello & Arneri 2009), one generation per year was assumed. Following (Waples & Yokota 2007), N_e was also estimated over the whole timeframe of each time series to detect consistency with estimates obtained from intermediate temporal timeframes. NE-ESTIMATOR 2.01 (Do *et al.* 2014) was used to estimate moment-based N_e values. We implemented the following procedures: i) we used Plan II sampling in which individuals were removed after sampling, ii) we used the Pollak formula (Pollak 1983) for computing the standardized variance in allele frequency (F_K) to estimate N_e values, iii) we estimated 95% confidence intervals from jackknifing, and iv) we used a threshold of 0.01 as the lowest allele frequency considered to estimate N_e values, because allele frequencies were seldom lower than 0.01. The TM3 software (Berthier *et al.* 2002) was used to obtain coalescent likelihood-based estimates of N_e. TM3 uses a Bayesian approach based on coalescent theory and MCMC simulations to generate a posterior distribution of N_e values. In addition, we used

maximum N_e values (Ne MAX) and the number of generations between consecutive samples (T) as priors in the TM3 simulations. To test for consistency between simulations, three independent runs were performed using Ne MAXes equal to 1000, 5000, and 10,000. Simulations were performed with 20,000 MCMC iterations and were conducted separately for Chioggia and Vieste samples.

Estimates of N_e from the moment-based temporal method were used to provide temporal trends in N_e/N_c ratios from both time series. Estimates of N_c were obtained from annual acoustic surveys (MEDIAS Project, http://www.medias-project.eu; Fig. 2) (Leonori *et al.* 2011, 2012) for stocks in the northern and middle-southern Adriatic Sea.

Datasets were tested for genetic population bottleneck signatures with a heterozygosity-excess based method implemented in BOTTLENECK 1.2 (Piry *et al.* 1999). This test assumes that during a strong reduction in population size, allele numbers decline faster than H_e. Since expected heterozygosity at mutation–drift equilibrium (H_{eq}) is calculated from the number of observed alleles, H_e becomes larger than $H_{eq,}$ leading to heterozygosity excess (H_{exc}) (Cornuet & Luikart 1996). Values of H_{exc} were tested with Wilcoxon's signed rank tests (Luikart & Cornuet, 1998), a powerful and robust test when the dataset contains few (< 20) polymorphic loci (Piry *et al.* 1999). We used 1000 iterations and three mutational models: infinite allele model (IAM), stepwise mutation model (SMM), and a two phase mutation model (TPM) with 95% single-step mutations and 5% multistep mutations, as recommended in Piry *et al.* (1999).

In addition, purported bottlenecks were verified by simulating bottlenecked populations using BOTTLESIM 2.6 (Kuo & Jansen 2003). Genotypic data for CH78

and VI85 were used independently as founder populations to simulate bottleneck events lasting 32 and 25 generations, respectively. Population declines were tested using 1000 iterations with the following criteria: dioecious organisms with random mating system, balanced sex ratio (1:1), two years of expected lifespan for the species, given its high natural mortality (Carpi *et al.* 2015), one year at first reproduction (Morello & Arneri 2009), and 90% generational overlap. The moment-based N_e estimates from NE-ESTIMATOR were used as analogues of population size. Simulated values of H_e and N_A were plotted and compared with observed H_e and N_A estimates to provide evidence of a population bottleneck.

Finally, we estimated the degree of geographic and temporal variability among samples first with pairwise values of θ_{ST} (Weir & Cockerham 1984), calculated with FSTAT and second with a Bayesian approach, implemented in STRUCTURE 2.3.2.1 (Pritchard *et al.* 2000; Falush *et al.* 2003). The second approach estimates the number of populations (K) under Hardy–Weinberg expectations and linkage equilibrium. The most probable number was tested using priors ranging from $K = 1$ to $K = 6$, under an admixture model and with correlated allele frequencies. Ten independent runs were performed for each Kusing an MCMC of 500,000 iterations after a burn-in of 50,000 iterations.

Results

Missing genotypes accounted for 7.51% of the overall dataset. A pool of 90 individuals were randomly chosen (10 individuals per sampling year and site) to re-genotype 7 of the screened loci (a total of 630 control genotypes were produced). We were unable to PCR-amplify 2.3% of control genotypes (15 of 630), but the

remaining 615 genotypes were consistent with their first molecular size detection. There was a lack of allele dropout and stuttering throughout the entire dataset. The occurrence of null alleles was detected in 16 of 63 tests, and after the Brookfield correction (Brookfield 1996) 6 of 63 remained significant. Since the Brookfield correction improved our dataset, and no loci showed a systematic presence of null alleles, the entire corrected dataset was used for subsequent statistical analyses.

P_{ID} values ranged from 7.731 x 10^{-9} (CH87) to 2.854 x 10^{-7} (CH94), and $P_{ID(sib)}$ values ranged from 1.133 x 10^{-3} (CH87) to 2.743 x 10^{-3} (CH94) (S3 Table). Both values were lower than expected thresholds for unrelated or unbiased multilocus genotypes, suggesting enough power and sufficient quality of the data.

The observed number of alleles ranged from 7 (Ee2-165m) to 38 (Ee2-407m) across samples. Mean N_A ranged from 9.143 (CH10) and 12.000 (VI89 and VI10), whereas R_S ranged from 6.982 (CH94) to 8.769 (VI87). Mean H_e ranged from 0.703 (CH94) to 0.767 (CH87 and CH10), while mean H_o varied from 0.590 (CH87) to 0.746 (CH00) (S4 Table). There were 4 of 63 significant F_{IS} values (critical P < 0.00079). Two deviations were related to the Ee2-165m locus and led to heterozygote excesses in samples CH94 and CH00, while the remaining two were related to the Ee2-407m locus and suggested heterozygote deficiency in CH87, and VI87 (Table S4). Significant heterozygote deficits were found in samples CH78, CH87, and VI10 (Table S4), and for the Ee2-407m locus overall. No evidence for linkage disequilibrium was detected.

The overall temporal trends in H_e, H_o, and N_A of the Chioggia time series indicated a significant overall decrease in H_e (r = -0.053, P < 0.001; b = -0.001, P > 0.05), a significant overall increase in H_o (r = 0.666, P < 0.01; b = 0.002, P < 0.05), and a

non-significant overall increase in N_A (r = 0.402, P > 0.05; b = 0.022, P > 0.05) (Fig. 3a, 3b and 3c). Specifically, a highly significant decreasing trend was observed in H_e between CH87 and CH94 (Fig 3a; r = -1.000, P < 0.001; b = -0.113, P < 0.001), whereas significant increasing trends were observed in H_e between CH78–CH87 (Fig 3a; r = 1.000, P < 0.05; b = 0.038, P < 0.05) and between CH00-CH10 (Fig 3a; r = 1.000, P < 0.01; b = 0.007, P < 0.001), as well as, in N_A between CH00–CH10 (Fig 3c; r = 1.000, P < 0.01; b = 0.105, P < 0.001). The Vieste time series showed a non-significant overall decrease in H_e (r = -0.474, P > 0.05; b = -0.001, P > 0.05), a non-significant overall increase in H_o (r = 0.481, P > 0.05; b = 0.001, P > 0.05) and a non-significant overall increase in N_A (r = 0.640, P > 0.05; b = 0.034, P > 0.05) (Fig. 3d, 3e and 3f). Uniquely, the VI87–VI89 trend showed a significant increase in N_A (Fig. 3f; r = 1.000, P < 0.05; b = 0.605, P < 0.05).

Moment-based estimators yielded N_e values between 103.50 and 416.70 from the Chioggia time series (Table 1) and values between 68.00 and 891.20 from the Vieste time series (Table 1). Estimates from the overall timeframes showed N_e values of 597.70 in Chioggia and 666.70 in Vieste (Table 1).

The TM3 coalescent likelihood-based method showed N_e between 161.15 and 2373.85 for the Chioggia time series (Table 2) and between 134.17 and 2883.85 for the Vieste time series (Table 2). For both sets of estimates, the overall N_e estimates were consistent with the magnitudes of shorter timeframes estimates.

The results of the BOTTLENECK analysis using the IAM mutation model showed significant tests for heterozygosity excesses (H_{exc}) in every sample for both "one-tail" and "two-tailed" Wilcoxon's tests (Table 3). In contrast, tests using the SMM

and TPM mutation models were not significance (Table 3). The Shift-Mode tests revealed normal L-shaped allele frequency distributions in all the samples.

Simulated population bottlenecks indicated the likely occurrence of bottlenecks in both the Chioggia and Vieste time series, based on comparisons between simulated and observed expected heterozygosity (H_e) values. Similar values between simulated and observed H_e were found for CH78, CH94, and CH00 in the Chioggia time series (Fig. 4a), whereas H_e values for CH87 and CH10 were outside the confidence interval (Fig. 4a). In the Vieste time series, simulations showed similar simulated and observed H_e values for VI85, VI87, and VI89, whereas VI10 was placed outside the confidence interval (Fig. 4b). Observed values of N_A were consistent with simulated N_A for CH78 and CH87 in the Chioggia time series and for VI85 and VI87 in the Vieste time series.

Estimates of θ_{ST} revealed 9 significant tests among 36 total pairwise comparisons (Table 4). These significant comparisons were largely between localities and years (*i.e.* significant tests were CH87-VI85, CH94-VI85, CH94-VI89, CH00-VI85, CH00-VI89, CH10-VI89) (Table 4). Genetic variation varied little with time within each time series. In fact, in the Chioggia time series only the CH94-CH00 comparison was statistically significant, whereas two tests (VI85-VI89 and VI85-VI10) were significant in the Vieste time series (Table 4).

STRUCTURE analysis showed a lack of genetic structure between and within the Chioggia and Vieste time series. The greatest likelihood indicated a single population, $K = 1$ [LnP(K) = -10843.28] (Fig. S1).

Outcomes from the neutrality tests are available in File S1 and Table S5 and S6.

Discussion

The extent of genetic variability estimated from both time series revealed the effectiveness of using archived DNA for population genetics studies. Otolith collections represent valuable sources of material for historical comparisons of genetic variation among locations or temporal variation within locations. Major challenges with the use of archived tissues are the reliability of results in light of potential DNA contamination and the level of polymorphism of the molecular marker. In the first case, we took considerable precautions to avoid DNA contamination. Genotypes were consistent between initial genotyping and re-genotyping some numerous individuals, indicating genotype quality. The amount of missing data was similar with that in other studies in which archived DNA from otoliths was used (Cuveliers *et al.* 2011; Schaerlaekens *et al.* 2011; Ruggeri *et al.* 2012; Jasper *et al.* 2013). In addition, the comparison between polymorphisms detected in the oldest and newest samples confirmed no age-related failure in genotyping. A new set of PCR primers for seven microsatellite loci yielding products < 200 bp produced a high degree of amplification success, with amplification success was between 68.75% and 100%. The number of loci genotyped in this study is similar to the mean number of loci typed in other archived DNA studies (Cuveliers *et al.* 2011; Schaerlaekens *et al.* 2011; Ruggeri *et al.* 2012; Jasper *et al.* 2013).

Loss of genetic variability as a consequence of demographic collapse

Temporal trends in genetic variables are concordant with abundance trends in the Chioggia and Vieste time series. However, the Chioggia time series showed greater changes in genetic variables than the Vieste time series. In the Chioggia time series, a significant reduction in H_e appeared after the population collapse in 1987 and lasted

until 1994. This observation was additionally confirmed by a decrease in N_A from 1978 to 1994 and supports the idea of different responses of H_e and N_A to demographic declines (Nei *et al.* 1975). After 2000, both H_e and N_A increased in the Chioggia time series in association with a growing population biomass in northern Adriatic Sea (Leonori *et al.* 2011). Values of H_e also declined in the middle-southern Adriatic basin (Vieste), but this drop was not statistically significant. However, values of N_A increased significantly after 1989. These results clearly show the influence of demographic fluctuations on genetic variability of Adriatic anchovies, which has also been observed in other marine populations (Hauser *et al.* 2002; Ruggeri *et al.* 2012). These results add to a growing number of studies showing a large contrast between the genetic effective population size and census in marine species (Hauser & Carvalho 2009).

Even though a significant reduction in H_e in northern Adriatic anchovies is consistent with a bottleneck in N_e and with a recent 10 to 20-fold drop in fishery biomass (Carpi *et al.* 2015; Santojanni *et al.* 2003; Leonori *et al.* 2011), the results from the heterozygosity excess test did not support this scenario. This outcome may be due to two mechanisms. First, the loss of genetic diversity after demographic decline may not have been strong enough to produce genetic profiles leading to heterozygosity excess. This weak-effect hypothesis, however, is inconsistent with the temporal drop in H_e in response to large declines in population abundance (Allendorf *et al.* 2013). Second, the lack of significance may have been due to low statistical power of the test to detect heterozygosity excess with our dataset. Peery *et al.* (2012) reviewed 703 populations belonging to 116 vertebrate species and concluded that even in species with strong population declines these tests often failed to detect bottlenecks

(Peery *et al.* 2012). The low statistical power of bottleneck tests results in many cases from little temporal variation in N_e (Peery *et al.* 2012), and this is consistent with our observations of a relatively small range of N_e among samples from Chioggia from 1978 to 2000, as well as that among samples from Vieste from 1985 to 1989.

The use of demographic simulations with BOTTLESIM showed that most of the observed H_e values between 1978 and 2000 in Chioggia and between 1985 and 1989 in Vieste, may be consistent with the loss of genetic variability as a consequence of a bottleneck. Observed N_A values between 1978 and 1987 in Chioggia and between 1985 and 1987 in Vieste were close to simulated values. These simulations also indicated a decrease in genetic diversity in the late 1980s that is consistent with bottlenecks in population size.

Reduced genetic variability as a consequence of a bottleneck, especially in the northern Adriatic Sea, is also indicated by significant deviations from Hardy-Weinberg expectations for Chioggia in 1978 and 1987, as a consequence of homozygote excess. These genetic signals support the hypothesis that a reduction in genetic variability began earlier than the decline in effective population size. Although unlikely in small pelagic fishes, other marine species characterized by high fecundities and a high variance in reproductive success (sweepstakes recruitment) often show signs of inbreeding (Hoarau *et al.* 2005; Hedgecock & Pudovkin 2011; O'Leary *et al.* 2013).

Evolutionary potential and effective population sizes in anchovies

The MCMC methods used to evaluate temporal variability in N_e in the Chioggia and the Vieste time series were characterized by good converges for this parameter,

which gives some confidence in the results. We observed low levels of effective population sizes (N_e) of only 100s of fish in the pre-collapse and early post-collapse periods in both historical time series. These results indicate that the reductions in genetic diversity were largely driven by genetic drift (Hare *et al.* 2011) and that genetic diversity dropped before the population collapse in 1987. Estimates of N_e from pre-collapse years were as low as those for 1987. Our global estimates of N_e for Adriatic anchovy indicate values of N_e between one hundred and a few thousand fish, values that have previously been reported in other small pelagics, such as European sardines (Ruggeri *et al.* 2012; Gaggiotti & Vetter 1999; Laurent & Planes 2007), and in other marine organisms characterized by high fecundity, type III survivorship and sweepstakes recruitment (Pinsky & Palumbi 2014; Hauser *et al.* 2002; Hoarau *et al.* 2005). These small effective population sizes challenge the idea that marine populations are invulnerable and inexhaustible resources by virtue of their huge census sizes (Pinsky & Palumbi 2014; Gaggiotti & Vetter 1999).

Estimates of N_e/N_c in Adriatic anchovies was between 10^{-6} and 10^{-8} orders of magnitude and are smaller than those typical of many marine species ($N_e/N_c \geq 10^{-5}$) (Hare *et al.* 2011). These small ratios indicate that the large variance in reproductive success can lead to the demographic instability. Although the primary intent of this article was not to provide advice to management, it is nevertheless important to compare our temporal estimates of N_e with conservation guidelines provided by the revised 50/500 rule (Frankham *et al.* 2014). The 50/500 rule recommends that populations be maintained at sizes larger than 50 individuals over the short term and at least 500 individuals over the long term to avoid inbreeding and the loss of genetic diversity, and at sizes larger than 500–1000 to indefinitely retain evolutionary

potential (reviewed in Frankham *et al.* 2014). In our results, anchovy pre- and immediately post-population collapse showed estimates of N_e that come close to those recommended by 50/500 rules, suggesting that severe population declines place Adriatic anchovies at risk of losing evolutionary potential in the absence of stable connectivity among anchovy subpopulations in the Adriatic basin.

Temporal population genetic integrity

Despite fluctuations in genetic variables in response to severe population declines, anchovies in the Adriatic Sea generally showed little genetic heterogeneity both between northern and middle-southern samples and among temporal samples in these areas. We detected a lack of genetic differentiation and a single population comprising all the Chioggia and Vieste samples. The occurrence a single genetic stock is also confirmed by a microsatellite study of samples from throughout the Adriatic basin (Ruggeri *et al.* in preparation). The lack of temporal variability confirms that this genetic composition has been maintained over at least the past four decades. Similar evidence was found in other temporal studies of small pelagics (Larsson *et al.* 2010; Ruggeri *et al.* 2012) and other commercially exploited demersal fishes (Cuveliers *et al.* 2011; Poulsen *et al.* 2006). The increasing trends in H_e, N_A and N_e over time indicates that both localities have recovered from a population bottleneck, but that the recovery appears to have started earlier in Vieste (from 1989) than in Chioggia (from 1994). The relative geographic genetic homogeneity of anchovy populations in the Adriatic (Ruggeri *et al.* in preparation) may indicate that migration between subpopulations is important for the recovery of genetic diversity in local populations. In this way, migration between the Chioggia

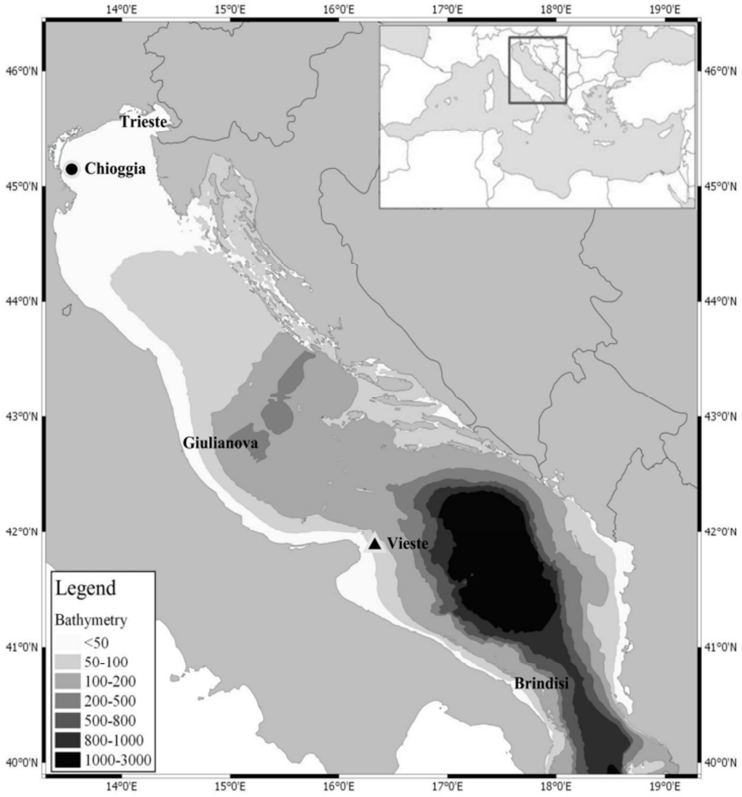
and Vieste populations likely led to the recovery of most of the genetic diversity lost in the Chioggia population over about seven generations.

Conclusions

This study confirms the utility of archived DNA to address evolutionary and fishery-related problems. Our results indicate that, contrary to the common assumption that abundant species, such as anchovies, are not immune to the loss of genetic variability, because sweepstakes recruitment can lead to small effective population sizes. The effective population sizes of Adriatic anchovies are several orders of magnitude smaller than census population sizes. This large discrepancy indicates that anchovy populations are more vulnerable to fishery pressures and environmental change than previously thought, and these small effective sizes must be taken into account in the management of fishery harvests. Our results also indicate that temporal genetic data provides an assessment of the impact of demographic disturbances on the persistence of local populations.

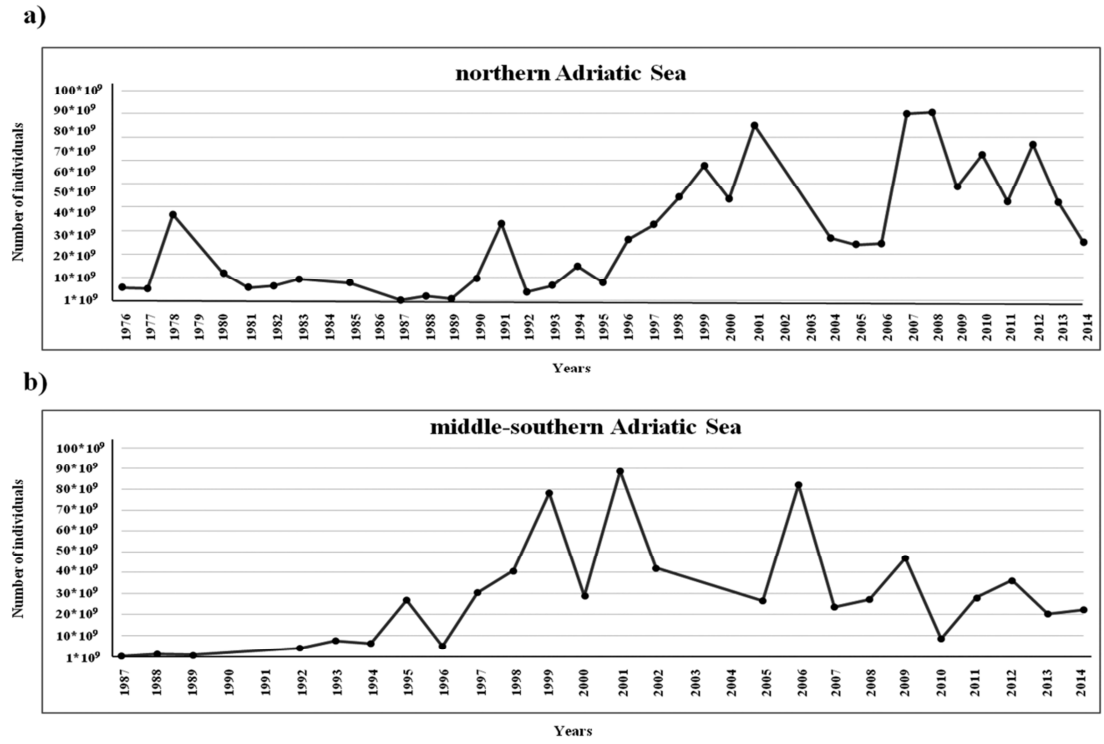
Figures and Tables

Fig. 1 Sampling area.



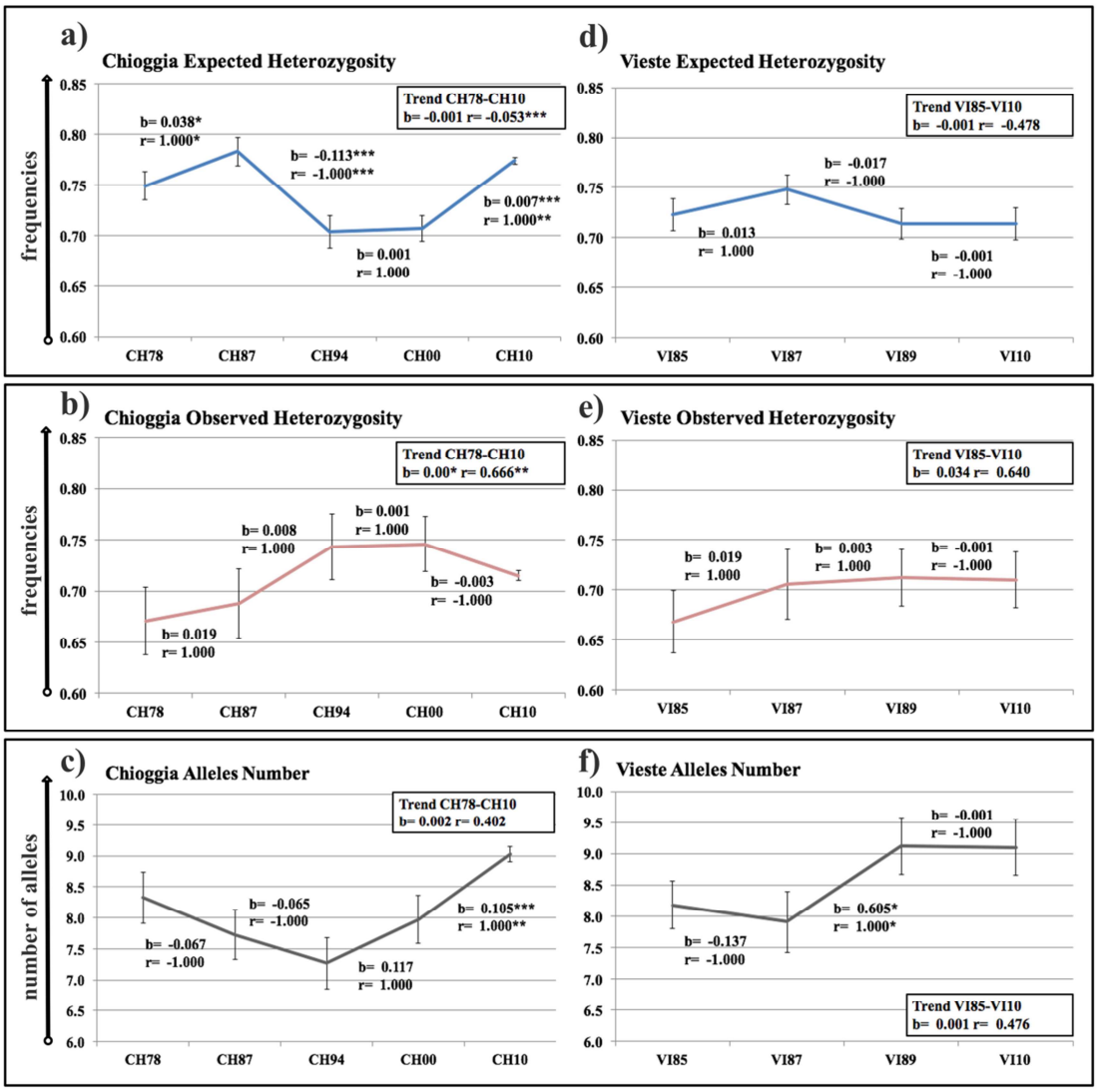
The solid circle indicates the location of the Chioggia time series between 1978 and 2010. The solid triangle indicates the locations of the Vieste time series between 1985 and 2010.

Fig. 2 Graphical representation of temporal trends of anchovy population abundance (expressed as number of individuals) obtained from annual MEDIAS acoustic surveys data (black dots).



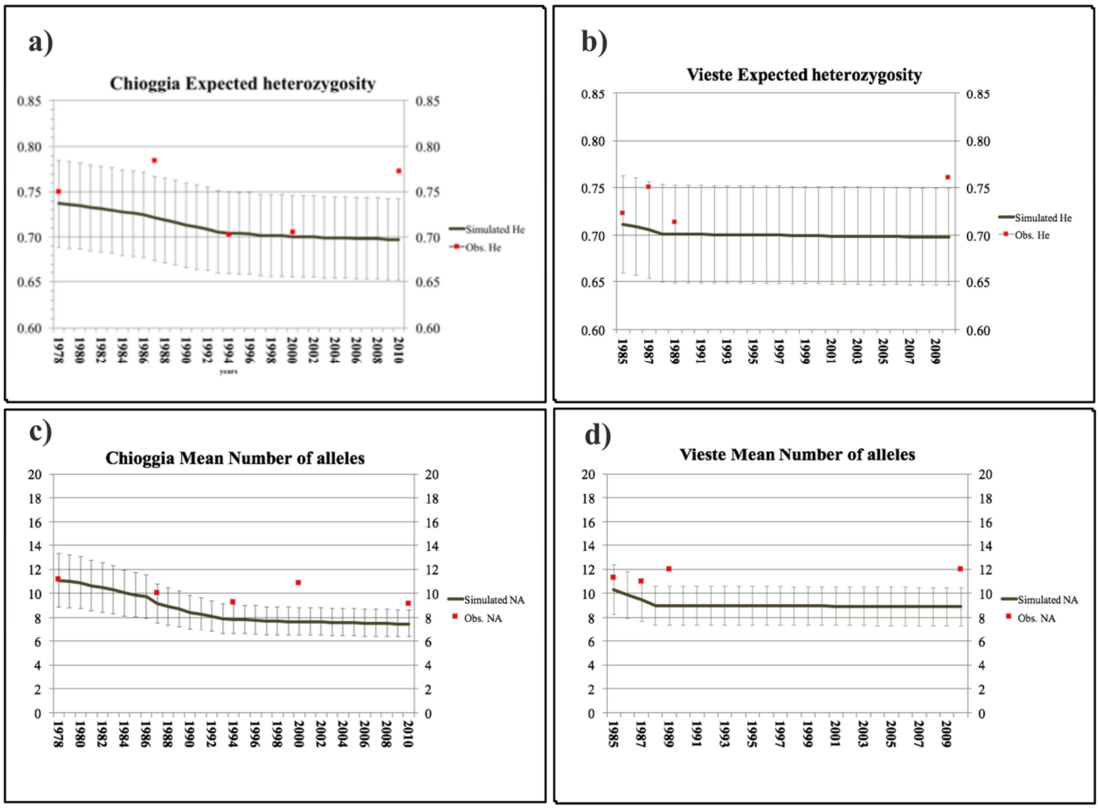
(a) trend of abundance in the northern Adriatic Sea. (b) trend of abundance in the middle-southern Adriatic Sea.

Fig. 3 Temporal changes in estimates of expected (*He*) and observed (*Ho*) heterozygosities, and mean number of alleles (*NA*).



Panels (a) (*He*), (b) (H_O) and (c) (N_A) depict temporal trends at Chioggia, and panels (d) (*He*), (e) (H_O) and (f) (N_A) depict temporal trends at Vieste. Standard deviation intervals are provided for each sample, estimated from 1000 resampled samples per year and standardized at 24 individuals. Statistical significances of *b* (slope) and *r* (Pearson's regression coefficient) indicate global trends (into the box to the left of each panel) and trends from consecutive temporal samples (above or below each timeframe considered). Level of significance: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Fig. 4 BOTTLESIM simulation.



(a) Trend (solid line) of *He* and (b) *N*A simulated values for a population (represented by CH78) experiencing a bottleneck for over 32 generations (years). Red squares represent the observed (a) *He* and (b) N_A values in CH78, CH87, CH94, CH00, and CH10. (c) Trend (solid line) of *He* and (d) *N*A simulated values for a population (represented by VI85) that went through a bottleneck for over 25 generations (years). Red squares represent the observed (c) *He* and (d) *N*A values in VI85, VI87, VI89, and VI10. Light grey bars are standard errors for simulated values.

Table 1 Temporal moment-based effective genetic size estimates.

Timeframe	N_G	N_e	Lower 95% CI	Upper 95% CI	N_e/N_c ratio
CH78-87	9	167.00	104.00	284.50	$1.90*10^{-8}$ CI ($1.18*10^{-8}$-$3.23*10^{-8}$)
CH87-94	7	103.50	56.00	197.30	$8.26*10^{-8}$ CI ($4.47*10^{-8}$-$1.57*10^{-7}$)
CH94-00	6	221.70	104.00	646.10	$1.12*10^{-8}$ CI ($5.20*10^{-9}$-$3.27*10^{-8}$)
CH00-10	10	416.90	171.10	3749.00	$1.05*10^{-8}$ CI ($4.31*10^{-9}$-$9.45*10^{-8}$)
CH78-10	32	597.70	342.40	1199.50	$1.50*10^{-7}$ CI ($8.59*10^{-8}$-$3.01*10^{-7}$)
VI85-87	2	68.00	23.60	588.30	NA
VI87-89	2	75.00	27.30	526.30	$3.10*10^{-7}$ CI ($1.13*10^{-7}$-$2.18*10^{-6}$)
VI89-10	21	891.20	444.20	2496.40	$1.47*10^{-7}$ CI ($7.35*10^{-8}$-$4.13*10^{-7}$)
VI85-10	25	666.70	247.50	2864.90	$3.89*10^{-7}$ CI ($1.44*10^{-7}$-$1.67*10^{-6}$)

N_G = number of generations between a pair of samples. *N*e = effective genetic size estimate. Lower and upper 95% CI represents a 95% Confidence Interval (CI) for *N*e estimated for each pair of samples. *N*e/*N*c ratio = ratio between effective population size (estimated from temporal moment-based method) and census size (obtained from MEDIAS acoustic survey data; Fig. 2). N_A = not available data

Table 2 Temporal TM3 coalescent-based effective population size estimates.

		Ne MAX = 1000			Ne MAX = 5000			Ne MAX = 10000		
Timeframe	N_G	N_e	Lower 95% CI	Upper 95% CI	N_e	Lower 95% CI	Upper 95% CI	N_e	Lower 95% CI	Upper 95% CI
CH78-87	9	**282.79**	148.51	464.09	258.31	163.58	523.98	270.47	145.12	460.43
CH87-94	7	**161.15**	96.14	260.52	149.69	92.47	272.03	145.78	90.48	226.06
CH94-00	6	817.52	314.40	1000.00	1000.77	257.99	5000.00	**680.22**	0.00	10000.00
CH00-10	10	984.40	387.64	1000.00	1162.30	323.83	5000.00	**2373.85**	118.03	10000.00
CH78-10	32	950.69	529.09	1000.00	**1026.87**	493.44	4370.12	1086.84	419.73	4805.98
VI85-87	2	**182.23**	66.81	777.09	149.76	0.00	577.55	153.22	0.00	560.15
VI87-89	2	**134.17**	67.96	383.19	119.02	33.26	456.84	250.03	0.00	1591.39
VI89-10	21	991.57	752.78	1000.00	4675.32	1202.63	5000.00	**2883.85**	977.89	10000.00
VI85-10	25	990.87	823.19	1000.00	1740.50	812.00	5000.00	**1612.00**	890.08	5760.81

N_G = number of generations between a pair of samples. *N*e = effective population size estimate. Lower and Upper 95% CI = represents a 95% Confidence Interval (CI) for *N*eestimated for a pair of samples. Ne MAX = maximum prior effective population size defined in coalescent-based simulations. Values in bold represent the most likely coalescent estimates based on higher 95% CI limit per simulation

Table 3 BOTTLENECK 1.2 results.

	IAM			TPM			SMM		
Samples	H_{def}	H_{exc}	H_{exc}-H_{def}	H_{def}	H_{exc}	H_{exc}-H_{def}	H_{def}	H_{exc}	H_{exc}-H_{def}
CH78	0.9921	**0.0117**	**0.0234**	0.4687	0.5937	0.9375	0.2890	0.7656	0.5781
CH87	0.9882	**0.0195**	**0.0390**	0.2343	0.8125	0.4687	0.1875	0.8515	0.3750
CH94	0.9960	**0.0078**	**0.0156**	0.7656	0.2890	0.5781	0.2890	0.7656	0.5781
CH00	0.9960	**0.0078**	**0.0156**	0.9609	0.0546	0.1093	0.8515	0.1875	0.3750
CH10	1.0000	**0.0039**	**0.0078**	0.5937	0.4687	0.9375	0.4687	0.5937	0.9375
VI85	1.0000	**0.0039**	**0.0078**	0.8515	0.1875	0.3750	0.6562	0.4062	0.8125
VI87	0.9960	**0.0078**	**0.0156**	0.4687	0.5937	0.9375	0.2890	0.7656	0.5781
VI89	1.0000	**0.0039**	**0.0078**	0.6562	0.4062	0.8125	0.4687	0.5937	0.9375
VI10	1.0000	**0.0039**	**0.0078**	0.6562	0.4062	0.8125	0.1875	0.8515	0.3750

IAM = Infinite Allele Model; TPM = Two Phase Mutation Model; SMM = Stepwise Mutation Model. Hdef = One-tailed heterozygosity deficiency test; Hexc = One-tailed heterozygosity excess test; Hexc–Hdef = Two-tailed heterozygosity deficiency or excess test. Values in bold are significant P-values ($P < 0.05$).

Table 4 Pairwise multilocus estimates of θ_{ST}.

	CH78	**CH87**	**CH94**	**CH00**	**CH10**	**VI85**	**VI87**	**VI89**
CH87	0.0169							
CH94	0.0061	0.0198						
CH00	0.0154	0.016	**0.0048**					
CH10	0.0078	0.0069	0.0078	0.0074				
VI85	0.018	**0.0238**	**0.0107**	**0.0197**	0.0079			
VI87	0.0048	0.0217	0.0176	0.0233	0.0149	0.0186		
VI89	0.0139	0.022	**0.0078**	**0.0157**	**0.0189**	**0.0053**	0.0209	
VI10	0.0077	0.009	0.0049	0.0104	-0.0021	**0.0111**	0.0175	0.0052

Bold values are significant after a sequential Bonferroni correction (Rice 1989) for 36 multiple tests ($P < 0.0014$).

References

Alheit J., Licandro P., Coombs S., Garcia A., Giráldez A., Garcia Santamaría M.T. *et al.* (2014) Atlantic Multidecadal Oscillation (AMO) modulates dynamics of small pelagic fishes and ecosystem regime shifts in the eastern North and Central Atlantic. *Journal of Marine Systems*, **133**, 88–102.

Allendorf F.W., Berry O., Ryman N. (2014) So long to genetic diversity, and thanks for all the fish. *Molecular Ecology*, **23**, 23–25.

Allendorf F.W., Luikart G.H., Aitken S.N. (2013) Conservation and the Genetics of Populations. 2nd Ed. Wiley-Blackwell Publishers, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK.

Antao T., Lopes A., Lopes R.J., Beja-Pereira A., Luikart G. (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323–327.

Azzali M., De Felice A., Cosimi G., Luna M., Parmiggiani F. (2002) The state of the Adriatic Sea centered on the small pelagic fish populations. *Marine Ecology-An Evolutionary Perspective*, **23**, 78–91.

Beaumont M.A., Nichols R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**, 1619–1626.

Benbouza H., Jacquemin J.M., Baudoin J.P., Mergeai G. (2006) Optimization of reliable, fast, cheap and sensitive silver staining method to detect SSR markers in

polyacrylamide gels. *Biotechnologie, Agronomie, Société et Environnement*, **10**, 77–81.

Berthier P., Beaumont M.A., Cornuet J.M., Luikart G. (2002) Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*, **160**, 741–751.

Bonanomi S., Pellissier L., Overgaard Therkildsen N., Berg Hedeholm R., Retzel A., Meldrup D. *et al.* (2015) Archived DNA reveals fisheries and climate induced collapse of a major fishery. *Scientific Reports*, **5**, 15395.

Brookfield J.F.Y. (1996) A simple method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Campana S.E., Thorrold S.R. (2001) Otoliths, increments and elements: keys to comprehensive understanding of fish populations? *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 30–38.

Carpi P., Santojanni A., Donato F., Colella S., Čikeš Keč V., Zorica B. *et al.* (2015) A joint stock assessment for the anchovy stock of the northern and central Adriatic Sea: comparison of two catch-at-age models. *Scientia Marina*, **79**, 57–70.

Colloca F., Cardinale M., Maynou F., Giannoulaki M., Scarcella G., Jenko K. *et al.* (2013) Rebuilding Mediterranean fisheries: a new paradigm for ecological sustainability. *Fish and Fisheries*, **14**, 89–109.

Cornuet J.M., Luikart G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Cuveliers E.L., Bolle L.J., Volckaert F.A.M., Maes G.E. (2009) Influence of DNA isolation from historical otoliths on nuclear-mitochondrial marker amplification and age determination in an overexploited fish, the common sole (*Solea solea* L.). *Molecular Ecology Resources*, **9**, 725–732.

Cuveliers E.L., Volckaert F.A.M., Rijnsdorp A.D., Larmuseau M.H.D., Maes G.E. (2011) Temporal genetic stability and high effective population size despite fisheries-induced life-history trait evolution in the North Sea sole. *Molecular Ecology*, **20**, 3555–3568.

Do C., Waples R.S., Peel D., Macbeth G.M., Tillett B.J., Ovenden J.R. (2014) NEESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, **14**, 209–214.

Falush D., Stephens M., Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Frankham R., Bradshaw C.J.A., Brook B.W. (2014) Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, **170**, 56–63.

Gaggiotti O.E., Vetter R.D. (1999) Effect of life history strategy, environmental variability, and overexploitation on the genetic diversity of pelagic fish populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 1376–1388.

Ganias K., Somarakis S., Nunes C. (2014) *Reproductive potential in Biology and Ecology of Sardines and Anchovies*. Taylor & Francis (Eds). CRC Press, Boca Raton, Florida, USA.

Goudet J. (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available: http://www.unil.ch/izea/softwares/fstat.html

Grant W.S., Bowen B.W. (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity*, **89**, 415–426.

Habel J.C., Husemann M., Finger A., Danley P.D., Zachos F.E. (2014) The relevance of time series in molecular ecology and conservation biology. *Biological Reviews of the Cambridge Philosophical Society*, **89**, 484–492.

Hare M.P., Nunney L., Schwartz M.K., Ruzzante D.E., Burford M., Waples R.S. *et al.* (2011) Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology*, **25**, 438–449.

Hauser L., Adcock G.J., Smith P.J., Ramirez J.H.B., Carvalho G.R. (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences*, **99**, 11742–11747.

Hauser L., Carvalho G.R. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333–362.

Hedgecock D., Pudovkin A.I. (2011) Sweepstakes Reproductive Success in Highly Fecund Marine Fish and Shellfish: A Review and Commentary. *Bulletin of Marine Science*, **87**, 971–1002.

Hoarau G., Boon E., Jongma D.N., Ferber S., Palsson J., Van Der Veer H.W. *et al.* (2005) Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (*Pleuronectes platessa* L.). *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 2125–2132.

Hood G. (2009) PopTools v. 3.1.0 Available: http://www.cse.csiro.au/poptools.

Hutchinson W.F., Carvalho G.R., Rogers S.I. (1999) A nondestructive technique for the recovery of DNA from dried fish otoliths for subsequent molecular genetic analysis. *Molecular Ecology*, **8**, 893–894.

Jasper J.R., Habicht C., Moffitt S., Brenner R., Marsh J., Lewis B. *et al.* (2013) Source-sink estimates of genetic introgression show influence of hatchery strays on wild chum salmon populations in Prince William Sound, Alaska. *PlosOne*, **8**, e81916.

Kauer M.O., Dieringer D., Schlötterer C. (2003) A microsatellite variability screen for positive selection associated with the 'Out of Africa' habitat expansion of *Drosophila melanogaster*. *Genetics*, **165**, 1137–1148.

Kuo C.H., Janzen F.J. (2003) Bottlesim: a bottleneck simulation program for long-lived species with overlapping generations. *Molecular Ecology Notes*, **3**, 669–673.

Landi M., Garoia F., Piccinetti C., Tinti F. (2005) Isolation of polymorphic microsatellite loci from the European anchovy, *Engraulis encrasicolus*. *Molecular Ecology Notes*, **5**, 266–268.

Larsson L.C., Laikre L., André C., Dahlgren T.G., Ryman N. (2010) Temporally stable genetic structure of heavily exploited Atlantic herring (*Clupea harengus*) in Swedish waters. *Heredity*, **104**, 40–51.

Laurent V., Planes S. (2007) Effective population size estimation on *Sardina pilchardus* in the Bay of Biscay using a temporal genetic approach. *Biological Journal of the Linnean Society*, **90**, 591–602.

Leonori I., De Felice A., Campanella F., Biagiotti I., Canduci G. (2011) Assessment of Small Pelagic Fish Biomass in the Western Adriatic Sea by Means of Acoustic Methodology. Marine research at CNR. Vol. DTA/06. 2019–2029. Fishery and Sea Resources 2011. Brugnoli E., Cavarretta G., Mazzola S., Trincardi F., Ravaioli M., Santoleri R. (Eds). Roma, ISSN 2239-5172, http://dta.cnr.it/

Leonori I., Tičina V., De Felice A., Vidjak O., Grubišić L., Pallaoro A. (2012) Comparisons of two research vessels' properties in the acoustic surveys of small pelagic fish. *Acta Adriatica*, **53**, 389–398.

Lin L., Zhu L., Liu S.F., Su Y.Q., Zhuang Z.M. (2011) Polymorphic microsatellite loci for the Japanese anchovy *Engraulis japonicus* (Engraulidae). *Genetics and Molecular Research*, **10**, 764–768.

Luikart G., Cornuet J.M. (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*, **12**, 228–237.

Marty L., Dieckmann U., Ernande B. (2015) Fisheries-induced neutral and adaptive evolution in exploited fish populations and consequences for their adaptive potential. *Evolutionary Applications*, **8**, 47–63.

Morello E.B., Arneri E. (2009) Anchovy and Sardine in the Adriatic Sea—An Ecological Review. *Oceanography and Marine Biology: An Annual Review*, **47**, 209–256.

Nei M., Maruyama T., Chakraborty R. (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Nielsen E.E., Hansen M.M. (2008) Waking the dead: the value of population genetic analyses of historical samples. *Fiah and Fisheries*, **9**, 450–461.

O'Leary S.J., Hice L.A., Feldheim K.A., Frisk M.G., McElroy A.E., Fast M.D. *et al.* (2013) Severe inbreeding and small effective number of breeders in a formerly abundant marine fish. *Plos One*, **8**, e66126.

Pakaki V., Magoulas A., Kasapidis P. (2009) New polymorphic microsatellite loci for population studies in the European anchovy, *Engraulis encrasicolus* (L.). *Molecular Ecology Resources*, **9**, 1406–1409.

Peakall R., Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **28**, 2537–2539.

Peery M.Z., Kirby R., Reid B.N., Stoelting R., Doucet-Bëer E., Robinson S. *et al.* (2012) Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology*, **21**, 3403–3418.

Pinsky M.L., Palumbi S.R. (2014) Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology,* **23**, 29–39.

Piry S., Luikart G., Cornuet J.M. (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Pollak E. (1983) A new method for estimating the effective population size from allele frequency changes. *Genetics*, **104**, 531–548.

Poulsen N.A., Nielsen E.E., Schierup M.H., Loeschcke V., Grønkjær P. (2006) Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). *Molecular Ecology*, **15**, 321–331.

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

Rice W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Rousset F. (2008) Genepop'007: a complete reimplementation of the Genepop software for windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Rozen S., Skaletsky H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (Eds), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ: 365-386.

Ruggeri P., Splendiani A., Bonanomi S., Arneri E., Cingolani N., Santojanni A. *et al.* (2012) Temporal genetic variation as revealed by a microsatellite analysis of European sardine (*Sardina pilchardus*) archived samples. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**, 1698–1709.

Santojanni A., Arneri E., Barry C., Belardinelli A., Cingolani N., Giannetti G. *et al.* (2003) Trends of anchovy (*Engraulis encrasicolus*, L.) biomass in the northern and central Adriatic Sea. *Scientia Marina*, **67**, 327–340.

Santojanni A., Arneri E., Bernardini V., Cingolani N., Di Marco M., Russo A. (2006) Effects of environmental variables on recruitment of anchovy in the Adriatic Sea. *Climate Research*, **31**, 181–193.

Santojanni A., Cingolani N., Arneri E., Belardinelli A., Giannetti G., Colella S. *et al.* (2006) Use of an exploitation rate threshold in the management of anchovy and sardine stocks in the Adriatic Sea. *Biologia Marina Mediterranea*, **13**, 98–111.

Schaerlaekens D.G., Dekker W., Wickström H., Volckaert F.A.M., Maes G.E. (2011) Extracting a century of preserved molecular and population demographic data from archived otoliths in the endangered European eel (*Anguilla anguilla* L.). *Journal of Experimental Marine Biology and Ecology*, **398**, 56–62.

Schoener T.W. (2011) The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science*, **331**, 426–429.

Speller C.F., Hauser L., Lepofsky D., Moore J., Rodrigues A.T., Moss M.L. *et al.* (2012) High potential for using DNA from ancient herring bones to inform modern fisheries management and conservation. *Plos One*, **7**, e51122.

Taggart J.B., Hynes R.A., Prodöuhl P.A., Ferguson A. (1992) A simplified protocol for routine total DNA isolation from salmonid fishes. *Journal of Fish Biology*, **40**, 963–965.

Van Beveren E., Bonhommeau S., Fromentin J.M., Bigot J.L., Bourdeix J.H., Brosset P. *et al.* (2014) Rapid changes in growth, condition, size and age of small pelagic fish in the Mediterranean. *Marine Biology*, **161**, 1809–1822.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M., Shipley P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Waits L.P., Luikart G., Taberlet P. (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology*, **10**, 249–256.

Waples R.S., Yokota M. (2007) Temporal estimates of effective population size in species with overlapping generations. *Genetics*, **175**, 219–233.

Weir B.S., Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

CHAPTER IV

*Splendiani A., *__Fioravanti T.__, Giovannotti M., Negri A., Ruggeri P., Olivieri L., Nisi Cerioni P., Lorenzoni M., Caputo Barucchi V. (2016) The Effects of Paleoclimatic Events on Mediterranean Trout: Preliminary Evidences from Ancient DNA. *PLoS ONE*, **11** (6), e0157975. doi:10.1371/journal.pone.0157975

** These authors contributed equally to this work*

The Effects of Paleoclimatic Events on Mediterranean Trout: Preliminary Evidences from Ancient DNA

Abstract

In this pilot study for the first time, ancient DNA has been extracted from bone remains of *Salmo trutta*. These samples were from a stratigraphic succession located in a coastal cave of Calabria (southern Italy) inhabited by humans from upper Palaeolithic to historical times. Seven pairs of primers were used to PCR-amplify and sequence from 128 to 410 bp of the mtDNA control region of eleven samples. Three haplotypes were observed: two (ADcs-1 and MEcs-1) already described in rivers from the Italian peninsula; one (ATcs-33) belonging to the southern Atlantic clade of the AT *Salmo trutta* mtDNA lineage (*sensu* Bernatchez). The prehistoric occurrence of this latter haplotype in the water courses of the Italian peninsula has been detected for the first time in this study. Finally, we observed a correspondence between frequency of trout remains and variation in haplotype diversity that we related with ecological and demographic changes resulting from a period of rapid cooling known as the Younger Dryas.

Introduction

The Pleistocene (from 2.58 to 0.0117Ma, Cohen *et al.* 2013) is a geological epoch characterized by repeated glacial-interglacial cycles that caused drastic environmental changes and impacted the distribution range and the genetic diversity of species and populations (Hewitt 1996). During glacial periods, the expansion of

ice sheets forced north-temperate species to move southward and survive in warmer refugia, the same species moved northwards in interglacial phases to recolonize their previous habitat. This process lead to "southern richness and northern purity" in north-temperate species biodiversity (Hewitt 1996, 2000). Species unable to shift their distribution range or adapt to the new environmental conditions became extinct at local or global scale (de Bruyn *et al.* 2011; Bellard *et al.* 2012).

Despite freshwater ecosystems cover only 0.8% of the Earth's surface, they host slightly less than one-half of the fish species currently known (Nelson 1994). This abundance of species in a so limited and confined space makes freshwater biodiversity particularly sensitive to various threats such as: i) climate changes; ii) overexploitation; iii) pollution; iv) flow modification; v) habitat degradation; and vi) invasion of exotic species (Dudgeon *et al.* 2006). One major threat is represented by the rapid and continuous global warming phase our planet is going through (IPCC 2014) that could lead to an inexorable decline of biodiversity over the 21st century. Probably, freshwater species will be the most affected (Sala *et al.* 2000; Pereira *et al.* 2010). Indeed, according to Xenopoulos *et al.* (2005), over 75% of fish biodiversity will be lost in the next 70 years in riverine systems due to reduced discharge. Besides the alteration of river's flow, climate change will threaten the survival of freshwater fish species because of the increase in water temperature and the subsequent reduction of dissolved oxygen (Ficke *et al.* 2007). Among freshwater fishes, salmonids are probably the organisms most sensitive to the severe alterations in the environmental parameters mentioned above (Jonsson & Jonsson 2011). Therefore, to understand how salmonids reacted to past climate changes is important to predict

how these fishes will respond to the future impact of climate changes on freshwater ecosystems.

Climatic fluctuations and environmental changes that occurred during the Pleistocene seem to have influenced the present distribution range of many salmonids (Waples *et al.* 2008). Therefore, the *Salmo trutta* species complex could be considered a good model to study how paleoclimatic changes affected biological and ecological traits of these freshwater fishes. This species complex is distributed largely across Eurasia and North Africa with anadromous populations occurring throughout its distribution range, except in the Mediterranean basin (Elliot 1994; Snoj *et al.* 2002). Several mitochondrial DNA (mtDNA) studies showed that Pleistocene successions of glacial and interglacial periods determined the conditions for the evolution of six *S. trutta* lineages: Atlantic (AT), Danubian (DA), Mediterranean (ME), Adriatic (AD), *marmoratus* (MA) and Duero (DU) lineages (Bernatchez 2001; Suárez *et al.* 2001; Cortey *et al.* 2009).

Genetic characterization of Pleistocene salmonids could provide crucial information concerning phylogeographic issues and temporal change in demographic and life history characteristics (Consuegra *et al.* 2002; Turrero *et al.* 2012). Over the last 30 years, great developments in the study of DNA from historical and archaeological samples have been made (Hagelberg *et al.* 2015). This led to use ancient DNA (aDNA) as a useful tool to perform retrospective monitoring of genetic biodiversity (Leonard 2008). The major studies in this field of research are focused on response of mammal populations during Pleistocene-Holocene climatic change (Dalén *et al.* 2007; Campos *et al.* 2010; Foote *et al.* 2013), especially to assess the causes of Quaternary megafauna extinctions (Lorenzen *et al.* 2011). Much less has been done

to reveal past climate change effects on fish species (Consuegra *et al.* 2002; Turrero *et al.* 2012). Indeed, aDNA studies on fish species were centered mainly on taxonomic identification of archaeological remains, without considering the role of climate changes in fish micro-evolutionary pathways (Arndt *et al.* 2003; Yang *et al.* 2004; Speller *et al.* 2005; Pagès *et al.* 2009; Halfmann *et al.* 2015).

Fishing and consumption of freshwater species are known for Europe during prehistoric times, when opportunistic Upper-Paleolithic hunter-gatherers fed mainly on large mammals, but integrated their diet also with birds, mollusks and fishes (Mannino *et al.* 2011). Among these alternative food sources freshwater fishes were the most common, especially salmonids, as demonstrated by archaeological excavations (Muñoz & Casadevall 1997; Albertini & Tagliacozzo 2005; Van Neer *et al.* 2007; Russ & Jones 2009). Therefore, trout and/or salmon bone remains accumulated in archaeological sites represent valuable records to study how past climate oscillations affected fish populations (*e.g.*, Turrero *et al.* 2013). In the sedimentary succession in the *Grotta del Santuario della Madonna* (Praia a Mare, Cosenza, Italy, GSM henceforth), described by Cardini (1970), *Salmo trutta* remains were found along a stratigraphic succession spanning from the Upper Paleolithic to the Mesolithic (Durante 1978). The distinctive vomer bones allowed the identification of these remains as belonging to *S. trutta*. In addition, the coastal position of the cave and the bone size led to hypothesize that these *S. trutta* remains belonged to anadromous individuals (Durante 1978). According to the literature (Scarciglia *et al.* 2009), the sedimentary succession in GSM displays a continuous series of remains referable to the presence of a human settlement from the late Pleistocene to the Holocene. In this succession, three different climatic phases can be

recognised: i) the Bølling-Allerød interstadial, a warm period at the end of the Pleistocene (*c.* 14,700–12,700 cal yr BP), ii) the Younger Dryas (YD) stadial, a period that lasted 1,300 (±70) yr, characterized by cold climatic conditions and droughts, which occurred between approximately 12,800 and 11,600 cal yr BP, and iii) the Preboreal oscillation of the early Holocene (*c.* 11,500 cal yr BP).

In this study, bone remains of trout were selected from the stratigraphic succession of GSM. aDNA was extracted from these remains and the variation of a portion of the mtDNA control region (D-loop) was analysed in order to verify if genetic diversity changes occurred in the period of time encompassed by the stratigraphic succession analysed. The mtDNA genetic diversity observed was then compared with *S. trutta* abundance (Durante 1978) and paleoclimatic reconstructions.

Materials and Methods

Salmo trutta bone samples used in this study were collected between 1957 and 1970 during the excavation of GSM (Cardini 1970) and provided to us by the *Istituto Italiano di Paleontologia Umana (Is.I.P.U.).* A total of 12 brown trout sub-fossil bone remains were collected for this study. This brown trout bone collection is housed at the storage facilities of *Is.I.P.U.*, [Piazza Ruggero Bonghi 2, 03012 Anagni (FR), Italy]. The permit for studying and using the above samples was issued by Fabio Parenti of the Scientific Comittee of Istituto Italiano di Paleontologia Umana, on March 14th, 2014. This cave is located on the slopes of Monte Vingiolo, Praia a Mare village (Cosenza, Italy), at *c.* 50 m above sea level and not far from the mouth of the Noce River (Fig. 1). The excavation exposed a stratigraphic succession about

12 m thick that contained several faunal and artifact remains. The finding shows that GSM was occupied by humans at least from the late Upper Paleolithic (around 14,200 cal yr BP) to historical times. This stratigraphic succession was divided into ten distinct archaeological layers which correspond to different periods when the cave was inhabited, as reported by Alessio *et al.* (1966, 1967), from top to bottom: I, Roman period; II, Bronze age; III, Eneolithic; IV, period not identified; V, late Neolithic; VI, VII and VIII, Middle Neolithic; IX, Mesolithic and X, Upper Paleolithic. Each layer was subdivided into stratigraphical spits (or sections) numbered with Arabic numbers from 0 (top) to 72 (bottom) and, for some of them, samples of charcoal, charred bones or shells were radiocarbon dated by Alessio *et al.* (1966, 1967). To establish the correct temporal occurrence of our samples, radiocarbon dates estimated for sections from which the samples were obtained were calibrated using the CalPal software (Weninger & Jöris 2008), CalCurve: CalPal2007HULU (for charcoal and charred bones) and INTCAL 04 Marine (Hughen *et al.* 2004) (for marine shells). This work was focused only on the Upper Paleolithic layer (X) and a part of the Mesolithic one (IX) in which a lot of *S. trutta* remains (8,290 specimens) were found, especially concentrated in the sections from 65 to 39 (Durante 1978).

A total of 12 *Salmo trutta* bone samples (Fig. 2; Table 1) were selected from eight stratigraphical sections. During laboratory analyses the bone remains were divided into two groups (six bones per group) and then, for each group, DNA extraction, PCR amplification and sequencing were performed in two distinct periods of time. This was done to test the validity of results. Either half of a large bone or a whole small bone was processed for aDNA extraction using a modified silica-spin column

method (Yang *et al.* 2004). The initial physical-chemical decontamination step was modified as proposed by Speller *et al.* (2005), and the bone powder from each sample was subdivided to have three extraction replicas. After the DNA extraction, a 633 bp fragment of the mitochondrial DNA, including 85 bp of tRNA Pro-gene and 548 bp of mtDNA D-loop, was amplified using six overlapping pairs of primers (Fig. 3; Table 2). An additional primer set (StCR-7) was designed to amplify a consecutive but not overlapped D-loop region with a mutation characteristic of the South European Atlantic clade of *S. trutta* (*i.e.* AT3-3 clade *sensu* Cortey *et al.* 2009).

Each mtDNA fragment was amplified for every sample in a separate 25 µl PCR reaction containing 1X PCR buffer, 1.5 mM $MgCl_2$, 0.08 mM of each dNTP, 0.48 µM of each primer, 4 U Platinum Taq DNA Polymerase (Invitrogen) and 3 µl of DNA extract. PCR was performed in a BIO RAD T100 Thermal Cycler with a 7 min initial denaturation at 94°C, followed by 60 cycles at 94°C for 20 s (denaturation), 53.9°C for 30 s (annealing), 72°C for 40 s (extension), and a final extension at 72°C for 7 min. All amplicons were visualized on 2% agarose gels to evaluate success and quality of amplification. PCR products were purified by EXOSAP and Sanger sequenced with both forward and reverse primers using an ABIPRISM 3730XL DNA sequencer. To validate our results, PCR amplification and sequencing were performed twice, from different extraction replicas of each sample. To avoid contamination, DNA extraction and PCR preparation were performed in a dedicated laboratory (Cooper & Poinar 2000; Pääbo *et al.* 2004), in which modern trout DNA had never been introduced. Furthermore, the setting of the laboratory and the preparation of all working phases followed many of the recommendations from Knapp *et al.* (2012) to prevent DNA contamination. Each step was performed in two

different laboratory hoods in order to spatially separate extraction and PCR preparation phases. Each sample was manipulated and processed individually from the DNA extraction to sequencing with no overlapping steps in samples handling. Other precautions to prevent human-DNA contamination were adopted like wearing full body disposable suits, laboratory dedicated shoes, face masks and gloves. All working surfaces were washed with 10% bleach and UV-lighted for 20–30 min before and after any work session. Dedicated pipettes, filtered pipette tips, laboratory plasticware, glassware and reagents, were irradiated with UV in the same way as the working surfaces. In order to avoid the use of thermal cycler already used for PCR amplifications of modern trout DNA, a separate room for PCR was provided with a brand new thermal cycler that was cleaned before every amplification with 10% bleach and, after some minutes, rinsed with pure sterile water. To detect any possible contamination, extraction and PCR-negative controls were performed in all cases (Cooper & Poinar 2000).

The sequences of all fragments were assembled to give a partial mtDNA control region sequence. The assembled sequences were first analysed by BLAST (Altschull *et al.* 1990) to obtain preliminary taxonomic identifications. Subsequently, aDNA sequences obtained in this study were aligned with six modern *Salmo trutta* sequences found in GenBank using CLUSTALW (Thompson *et al.* 1994; Larkin *et al.* 2007). The classification into the main brown trout mtDNA lineages was based i) on the diagnostic nucleotide state of character observed in polymorphic sites and ii) a haplotype network based on the 410 bp mitochondrial DNA data set composed of 125 modern (see Table S1) and our three ancient haplotypes using the TCS method (Clement *et al.* 2000) in PopART v1.7 (Leigh & Bryant 2015).

Results

Genetic analyses

After PCR amplification and sequencing of seven fragments a total of 410 bp at the 5' end of the mtDNA control region was successfully reconstructed for six samples out of 12. For five samples, we were able to amplify from three to six of the seven fragments, while for one sample we failed in all PCR attempts (Table S2). The samples with the highest number of missing amplified fragments were the most recent one (V44) and the two oldest (TP65 and Pr65) (see Table 3 and Table S2). The worst amplification success was observed for fragment 2a. In this case the use of a new primer set (StCR-2c) allowed the amplification of a region shorter than 2a (Fig. 3; Table S2).

The BLAST search performed for all the sequences obtained showed that all the bone remains analysed here belong to *S. trutta* individuals, consistent with morphological identification of bone samples carried out by Durante (1978). The comparison with modern brown trout D-loop sequences found in GenBank highlighted the presence of nine variable positions, which allowed the identification of three different haplotypes already described in the literature: MEcs1, ADcs1 (Cortey *et al.* 2004) and ATcs-33 (Cortey *et al.* 2009) (Table 3). The oldest sample TP65 was identified as the ATcs-33 haplotype on the basis of the state of character observed at the nucleotide positions 15,775, 15,809 and 16,193 of the sequence assembled from the five segments amplified. The classification of the sample V44 was more difficult due to the presence of only three variable sites (positions 16,052, 16,053 and 16,066) which allowed its attribution to the Atlantic lineage but not to a

specific haplotype. However, due to the observation of the sole haplotype ATcs-33 in the rest of the Atlantic samples we also attributed this haplotype to the sample V44.

The above classification was also confirmed by a statistical parsimony haplotype network (Fig. 4). In fact, the three haplotypes detected from ancient trout were classified into three different clusters. The nucleotide sequences obtained by the specimens V62, M62 and M57 were identical to the partial haplotype MEcs-1. In addition, the haplotype MEcs-1 occupied a central position in a star like structure characterizing the ME lineage. In a similar manner, the specimens PT57 showed a nucleotide sequence identical to the partial haplotype ADcs-1. Also in this case, we detected this haplotype in the centre of a star like figure for the AD lineage. Finally, the specimens Pr50, M50, M48 and V46, showed the same nucleotide composition of the partial haplotype ATcs-33. In the parsimony network this latter haplotype represented the basal haplotype of a star-like cluster composed by the haplotypes already classified by Cortey *et al.* (2009) into the southern Atlantic clade AT3-3.

Correlation of DNA data to Greenland Ice-Core (GRIP)

The CalPal calibration (Table 4) allowed to refine the age of the interval studied (10,850 and 7,555 years BP, not calibrated) between 12,820 and 8,340 cal yr BP. The correlation between our DNA data and the paleoclimatic reconstruction based on the δ18O of the Greenland Ice-Core (GRIP) is shown in Fig. 5. The comparison with the paleo-temperature shows that our samples encompass three different climatic phases: i) the Bølling/Allerød (warm and wet); ii) the Younger Dryas (cold and dry, hereafter YD); and iii) the Preboreal oscillation (a climate reversal) (*e.g.*, Fisher *et al.* 2002), which mark the transition from Pleistocene to Holocene. In particular, in the oldest

sample (TP65, about 12,820 cal yr BP), which falls within an interval characterized by a transition from warm temperature (as evidenced by low $\delta^{18}O$ values, related to the Bølling/Allerød interstadial) to the start of the cold YD, we observed the haplotype ATcs-33 (Table 4; Fig. 5). In more recent strata, we observed the haplotypes MEcs-1 (samples: V62, M62 and M57) and ADcs-1 (PT57) (Table 4; Fig. 5). The time interval of these strata correlated to the YD cold stadial (about 12,800–11,600 cal yr BP) and probably to the Preboreal Oscillation, a climate reversal due to the melt water flow originated after the abrupt drainage of glacial Lake Agassiz (Fisher *et al.* 2002) which occurred about 11,335 cal yr BP, and lasted until 10,750 cal yr BP. Finally, from about 10,120 cal yr BP to the disappearance of trout remains (layers IX and VIII, sections 45–46 and 40–41, about 9,760 and 8,340 cal yr BP) we detected only the haplotype ATcs-33 (samples: Pr50, M50, M48, M47, V46 and V44) (Table 4; Fig. 5). This last interval identifies the beginning of the warmer Holocene climate persisting nowadays.

Discussion

Validation of the results and criticisms

The main criticisms to this study might relate to: i) the small sample size and ii) the scarcity and degradation of extracted DNA from archaeological samples. As for the first issue, obtaining samples from which aDNA can be extracted is challenging, especially because during laboratory analysis the samples must be destroyed. Consequently, the limited availability of larger sample sizes is unavoidable in such studies. As for the second issue, we retain our results reliable as: i) we avoided contamination during DNA extraction and PCR by mean of a specific

decontamination protocol (see material and methods), ii) we obtained comparable haplotypes from different samples of the same layer and from different extraction replicas of the same samples, iii) we observed the haplotype ATcs-33 never observed before in the modern samples analysed in our laboratories (in case of contamination we would expect to find the haplotype ATcs-1, or similar domestic haplotypes) and iv) the absence of new mutations and/or chimera sequences suggests that the sequence analysis presented in this study is not altered by the typical PCR errors generated by aDNA lesions.

mtDNA diversity of the Paleolithic Mediterranean trout

The analysis of genetic diversity of mtDNA extracted from skeletal remains of trout revealed the presence of three different brown trout phylogenetic lineages: ME, AD and AT. In particular, for the lineages ME and AD we observed the basal haplotypes (MEcs-1 and ADcs-1) of both lineages in agreement with the wide geographical distribution of these mtDNA variants shown by previous studies (Bernatchez 2001; Cortey *et al.* 2004; Snoj *et al.* 2011). According to Cortey *et al.* (2004), the wide geographic distribution of MEcs-1 and ADcs-1 and their location at central position in star-like phylogenies fit well with a sudden demographic expansion for both lineages. As an example, paleontological data from the Mediterranean area show that salmonids varied their range by activating migratory tactics in colder climatic periods. In particular, the study of fossil remains from Iberian Peninsula suggested that during glacial intervals brown trout occupied river portions in lowland areas currently dominated by thermophilic fish species (Muñoz & Casadevall 1997). This example suggests a similar scenario during the Pleistocene to explain the natural

presence of the AT lineage in southern Italy. In fact, the haplotype ATcs-33, belonging to the so called "southern Atlantic clade" AT3-3 (Suárez *et al.* 2001), and already observed in Mediterranean rivers (Spain, North Africa and Sicily; Cortey *et al.* 2009; Snoj *et al.* 2011; Fruciano *et al.* 2014), was detected in the present study. This haplotype also shows a central position in a star-like phylogenetic network and a pattern of nucleotide diversity indicating a relatively recent origin, about 100,000 years ago (Snoj *et al.* 2011). We therefore suggest that the beginning of the last glacial interval represented an important opportunity for this pioneer clade to expand its range in the Mediterranean basin. The recovery of this Atlantic haplotype among Upper Paleolithic remains of GSM shows that the expansion of the southern clade AT3-3 was not limited to the rivers of South-eastern Sicily; instead it moved northwards in the Mediterranean basin.

The results of this study also show that different haplotypes occur in different sections and layers of the stratigraphic succession. The oldest samples here analysed (about 12,820 cal yr BP) were characterized by the presence of haplotype ATcs-33. The haplotype composition changed between about 12,800–11,600 cal yr BP when only the "Mediterranean" haplotypes MEcs-1 and ADcs-1 were found. Finally, in the most recent layers (about 10,120–8,340 cal yr BP), we detected only haplotype ATcs-33 (Fig. 5). The changes in the haplotype composition observed (from ATcs-33 to MEcs-1 and ADcs-1, and *vice versa*) along the stratigraphic succession coincide with the change in the abundance pattern of *S. trutta* remains described by Durante (1978). In fact, the most abundant deposit of bone remains coincide with the observation of ME and AD haplotypes that in turn is related to the sudden onset of the YD (Fig. 5). However, the limited number of specimens analysed in this study do

not allow us to outline strong and conclusive hypothesis concerning the paleo-ecologic dynamics occurring on *S. trutta* populations from southern Italian peninsula. Therefore, future analyses of a larger number of samples will allow to make sound inferences on the paleo-ecologic dynamic occurring on *S. trutta* populations from southern Italian Peninsula.

Concluding remarks

This is the first aDNA study performed on brown trout sub-fossil bone remains. Here, we analysed remains dating back to the Pleistocene-Holocene transition (Cardini, 1970), and we observed a probably correspondence between the frequency of *S. trutta* in the deposit and variation in mitochondrial genetic diversity. Although the scarce number of bones analysed call for caution on results interpretation we tentatively relate this variation in the genetic pattern to changes in the ecological and demographic characteristics resulting after the abrupt and short climatic event known as the "Younger Dryas" (Pleistocene/Holocene transition). The results discussed highlight how the aDNA from sub-fossil remains of salmonids can provide crucial information to link population processes with the climatic changes.

If the results of this study suggest that trout survived in the Mediterranean basin during the alternate climatic phases of Pleistocene, then we have to seriously consider the current deterioration in environmental conditions mediated by humans. Nowadays, the survival chances for modern and/or future populations to severe warming periods could be even more challenging than before because of anthropogenic factors, such as the introduction of alien taxa, freshwater habitat alteration and excessive water abstraction (*e.g.*, Clavero *et al.* 2010; Almodòvar *et al.*

2012; Filipe *et al.* 2013). This study points out the importance of the archaeological DNA data to understand the complex dynamics of colonization that characterized many Holarctic species in response to climate changes which occurred during the Pleistocene. The need for an improvement of our knowledge on how species respond to climatic changes is particularly important for freshwater fishes and in particular for salmonids, especially if we take into account the current global warming and the enormous socioeconomics interests surrounding the management of local trout populations (*e.g.*, Blanchet & Dubut 2012).

Figures and Tables

Fig. 1 Geographic position of Grotta del Santuario della Madonna.

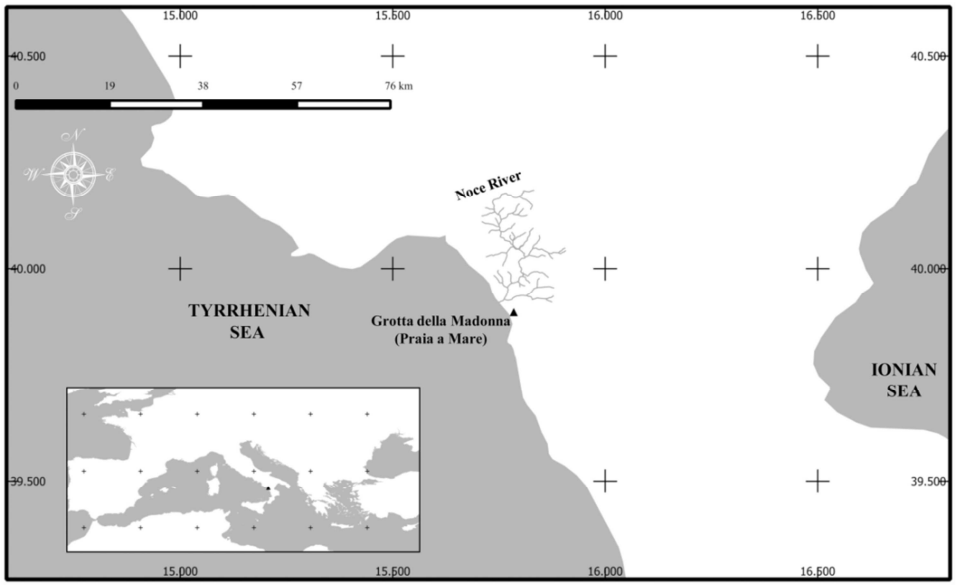


Fig. 2 *Salmo trutta* sub-fossil samples analysed in this study.

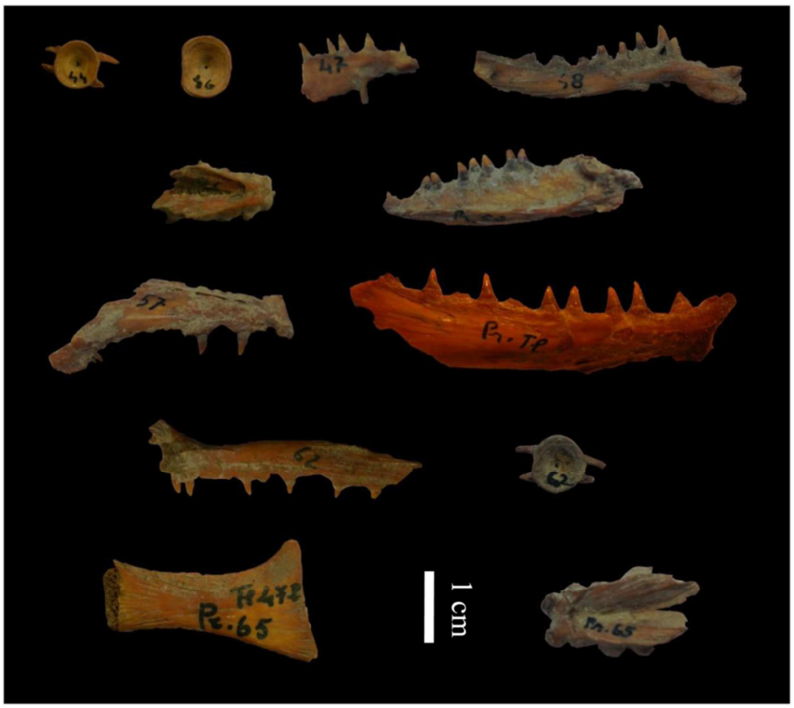
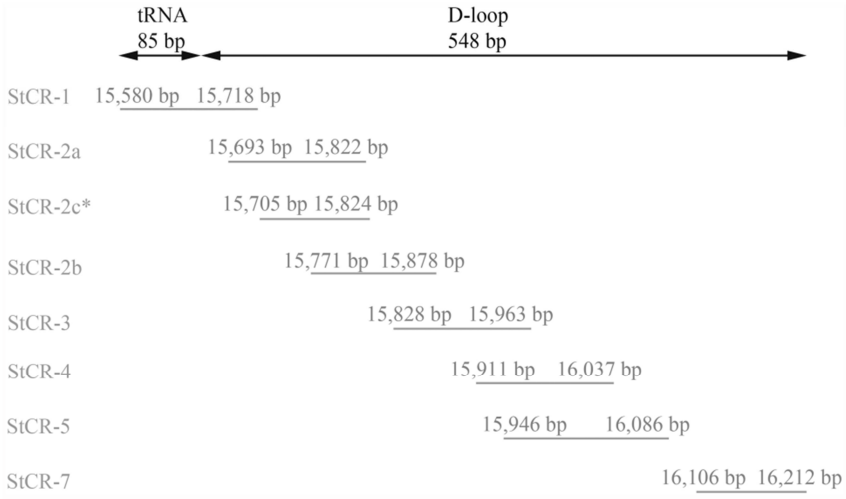
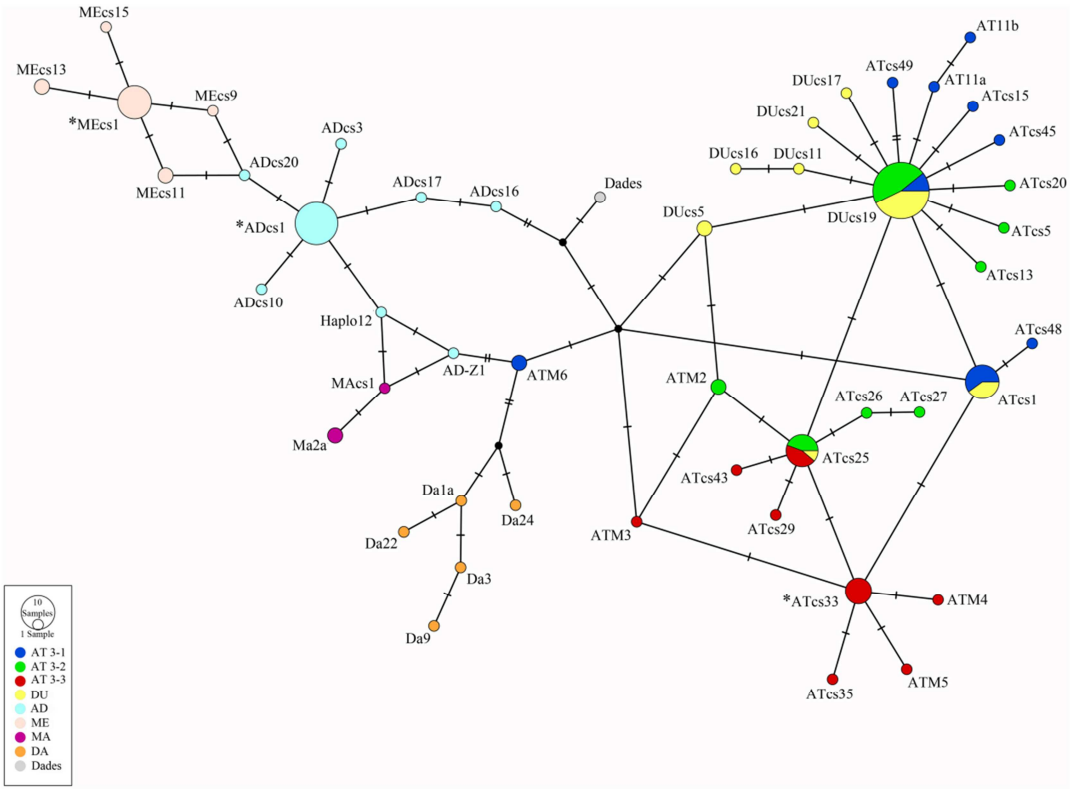


Fig. 3 Primer pairs designed and used in this study.



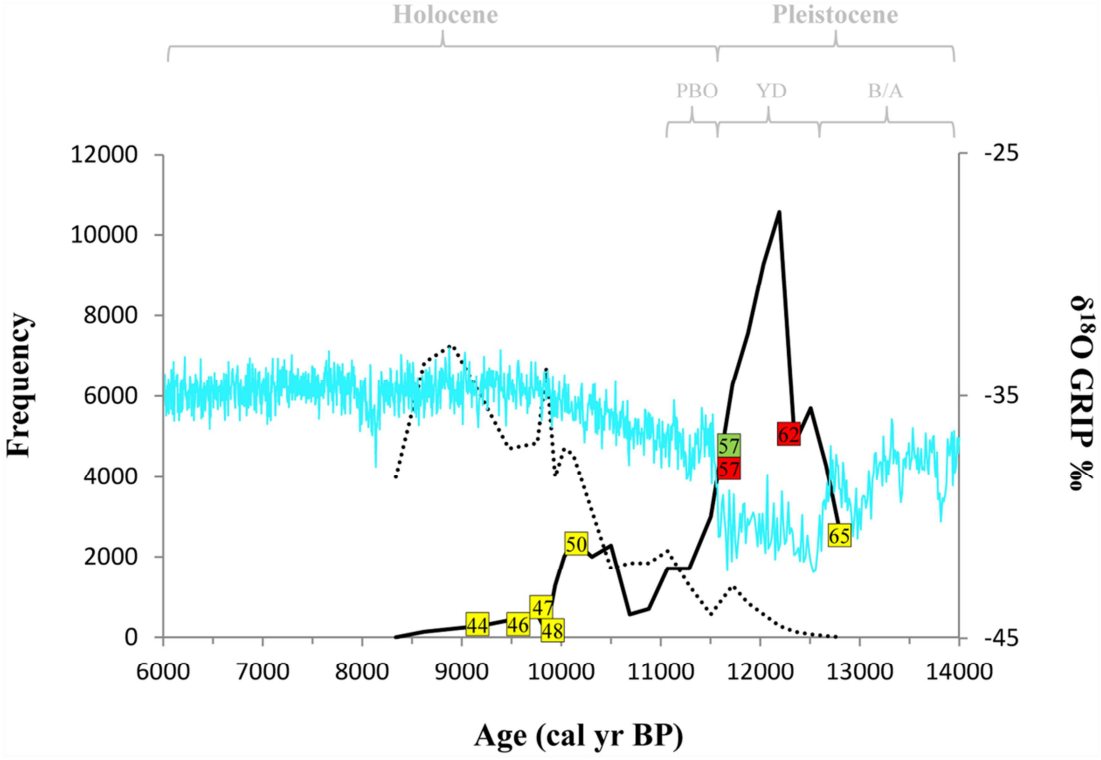
Primer pairs were designed on the basis of the brown trout complete mitochondrial genome present in GenBank (Accession no. AM910409.1) to amplify a 633 bp fragment of the mtDNA, including 85 bp of tRNA Pro-gene and 548 bp of mtDNA D-loop. *StCR-2c primer pair was designed to substitute for StCR-2a in the cases in which the amplification with the latter pair of primers was not successful.

Fig. 4 Statistical parsimony haplotypes network.



Haplotype network showing the relationships between modern haplotypes and haplotypes observed in the archaeological samples analysed in the present study (indicated by asterisks).

Fig. 5 Relationship between abundance of remains, mitochondrial haplotypes observed in this study and Pleistocene-Holocene climatic fluctuations.



The continuous line in the graph indicates the frequency of Salmo trutta remains and the dotted line indicates the frequency of Patella and Monodonta remains in all sections of the stratigraphic succession of GSM (approximate reconstruction from [40]). Mitochondrial haplotypes observed in this study are indicated by coloured squares: ATcs-33 in yellow, ADcs-1 in green and MEcs-1 in red. This information is related with the δ18O (‰) climate proxies of the GRIP ice core record represented by a blue line (dataset from: http://www1.ncdc.noaa.gov/pub/data/paleo/icecore/greenland/summit/grip/isotopes/gripd18o.txt).

Table 1 Information for *Salmo trutta* bone samples analysed.

Layer	Section	Sample ID	Laboratory ID	Sample type
First set of samples analysed				
IX	44	44	V44	Vertebra
IX	46	46	V46	Vertebra
X	50	50	M50	Glossohyal
X	57	Pr-Tl57	PT57	Dentary
X	62	62	M62	Maxilla
X	65	Tl472 Pr-65	TP65	Ceratohyal
Second set of samples analysed				
IX	47	47	M47	Maxilla
X	48	48	M48	Maxilla
X	50	Pr50	Pr50	Dentary
X	57	57	M57	Maxilla
X	62	62	V62	Vertebra
X	65	Pr65	Pr65	Glossohyal

Table 2 Details regarding primer pairs used in this study.

Primer		Sequence 5' to 3'	Product length (bp)
StCR-1	**F**	AACCCTCCCTAGTGCTCAGA **(20)**	139
	R	TATAACATTGGGTTAGCAAGGTACA **(25)**	
StCR-2a	**F**	TTGTACCTTGCTAACCCAATGTT **(23)**	130
	R	GGGGTTAAATTCACTAATGTTGA **(23)**	
StCR-2b	**F**	AGCATGTGAGTAGTACATCAT **(21)**	108
	R	TGGTTATTATCACGTGTTTTGCT **(23)**	
StCR-2c	**F**	AACCCAATGTTATACTACATCTAT **(24)**	120
	R	GAGGGGTTAAATTCACTAATGTTGA **(25)**	
StCR-3	**F**	CATCAGCACTAACTCAAGGTTTACA **(25)**	136
	R	TGCTGAAAGTTGGTGGGTAA **(20)**	
StCR-4	**F**	TCAATAAAACTCCAGCTACACG **(23)**	127
	R	TGTCCCTGACCATCAATAAGAG **(22)**	
StCR-5	**F**	ACCCACCAACTTTCAGCATC **(20)**	141
	R	GGAACCAAATGCCAGGAATA **(20)**	
StCR-7	**F**	AAGAAACCACCCCCTGAAAG **(20)**	107
	R	GGGAACCCTATGCATATAA **(19)**	

Table 3 Alignment result.

	15,690	15,775	15,809	15,859	15,925	16,052	16,053	16,066	16,193
ATcs-1 (AF273086)	**T**	**T**	**G**	**A**	**G**	**C**	**T**	**T**	**C**
ATcs-3 (AF274574)	**.**	**.**	**.**	**.**	**.**	**.**	**.**	**.**	**T**
ATcs-25 (EF530487)	**.**	**.**	**A**	**.**	**.**	**.**	**.**	**.**	**T**
ATcs-33 (EF530495)	**.**	**C**	**A**	**.**	**.**	**.**	**.**	**.**	**.**
V44 (KX022957)	-	-	-	-	-	.	.	.	-
V46 (KX022958)	.	C	A	.	.	.	.	.	.
M47 (KX022959)	-	C	A	.	.	.	.	.	.
M48 (KX022960)	.	C	A	.	.	.	.	.	.
Pr50 (KX022961)	.	C	A	.	.	.	.	.	.
M50 (KX022962)	.	C	A	.	.	.	.	.	.
ADcs-1 (AY836330)	**C**	**.**	**.**	**.**	**C**	**T**	**C**	**C**	**T**
PT57 (KX022963)	C	-	.	.	C	T	C	C	T
MEcs-1 (AY836350)	**C**	**.**	**A**	**C**	**.**	**T**	**C**	**C**	**T**
M57 (KX022964)	C	.	A	C	.	T	C	C	T
M62(KX022965)	C	.	A	C	.	T	C	C	T
V62(KX022966)	C	.	A	C	.	T	C	C	T
TP65(KX022967)	.	C	A	-	-	.	.	.	.
Pr65	-	-	-	-	-	-	-	-	-

Diagnostic sites are numbered with reference to the brown trout complete mitochondrial genome (GenBank accession no. AM910409.1) and were obtained by aligning our ancient D-loop sequences against reference sequences from GenBank (in bold). GenBank accession numbers are in brackets. Dashes (-) indicate missing data due to sequencing failure.

Table 4 Radiocarbon dates and CalPal calibration results

Layer	Section	Sample type	Sample name	AGE yr BP	AGE cal yr BP	95% range cal BP
VIII	40-41	Charcoal	R-285	7,555 ± 85	**8,340**	8,520-8,160
IXa	45-46	Charred bones	R-188	9,070 ± 80	**10,250**	10,450-10,050
IXb	45-46	Charcoal	R-187	8,735 ± 80	**9,760**	10,040-9,480
IXb	45-46	Insoluble residue of R-187	R-187α	8,875 ± 85	**9,960**	10,280-9,640
X	49-50a	Charcoal	R-286	9,020 ± 125	**10,120**	10,500-9,740
X	49-50b	Burnt bones	R-287	9,035 ± 100	**10,140**	10,460-9,820
X	49-50c	*Porchus turbinatus*	R-288	8,600 ± 120	**9,180**	9,480-8,880
X	49-50d	*Patella* shells	R-288A	9,800 ± 140	**10,700**	11,100-11,300
X	54-55a	Charcoal	R-289	10,300 ± 100	**12,140**	12,640-11,640
X	54-55b	Burnt bones	R-290	9,750 ± 100	**11,070**	11,410-10,730
X	54-55c	Shells marine molluscs(mainly *Patella* shells)	R-291	10,450 ± 100	**11,530**	11,950-11,110
Xa	57-58	Charred bones	R-186	10,030 ± 90	**11,570**	11,950-11,190
Xb	57-58	Charcoal	R-185	10,120 ± 70	**11,720**	12,120-11,320
X	64-65	Charcoal	R-292	10,850 ± 100	**12,820**	13,000-12,640
X	71-72	Charcoal	R-293	12,100 ± 150	**14,200**	14,820-13,580

Details and radiocarbon dates of samples from different layers and sections of the stratigraphic succession from GSM *e.g.*, (Alessio *et al.* 1966, 1967). In bold, the results of radiocarbon dates calibration using CalPal software (Weninger & Jöri 2008), CalCurve: CalPal 2007 HULU or INTCAL 04 Marine (Hughen *et al.* 2004).

References

Albertini D., Tagliacozzo A. (2005) I resti di pesce provenienti dai livelli preistorici di Riparo Cogola. *Preistoria Alpina*, **40**, 187–191.

Alessio M., Bella F., Bachechi F., Cortesi C. (1966) University of Rome Carbon 14 Dates IV. *Radiocarbon*, **8**, 401–404.

Alessio M., Bella F., Bachechi F., Cortesi C. (1967) University of Rome Carbon 14 Dates V. *Radiocarbon*, **9**, 354–357.

Almodóvar A., Nicola G.G., Ayllón D., Elvira B. (2012) Global warming threatens the persistence of Mediterranean brown trout. *Global Change Biology*, **18**, 1549–1560.

Altschull S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Arndt A., Van Neer W., Hellemans B., Robben J., Volckaert F., Waelkens M. (2003) Roman trade relationships at Sagalassos (Turkey) elucidated by ancient DNA of fish remains. *Journal of Archaeological Sciences*, **30**, 1095–1105.

Bellard C., Bertelsmeier C., Leadley P., Thuiller W., Courchamp F. (2012) Impacts of climate change on the future of biodiversity.*Ecology Letters*, **15**, 365–377.

Bernatchez L. (2001) The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution*, **55**, 351–379.

Blanchet S., Dubut V. (2012) 'Back to the future': How archaeological remains can describe salmon adaptation to climate change. *Molecular Ecology*, **21**, 2311–2314.

Campos P.F., Kristensen T., Orlando L., Sher A., Kholodova M.V., Götherström A. *et al.* (2010) Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene. *Molecular Ecology*, **19**, 4863–4875.

Cardini L. (1970) Praia a Mare. Relazione degli scavi 1957–1970 dell'Istituto Italiano di Paleontologia Umana. *Bullettino di Paletnologia Italiana*, **79**, 31–59.

Clavero M., Hermoso V., Levin N., Kark S. (2010) Geographical linkages between threats and imperilment in freshwater fish in the Mediterranean Basin. *Diversity and Distribution*, **16**, 744–754.

Clement M., Posada D., Crandall K. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Cohen K., Finney S., Gibbard P., Fan J. (2013) The ICS international chronostratigraphic chart. *Episodes*, **36**, 199–204.

Consuegra S., García de Leániz C., Serdio A., González Morales M., Straus L., Knox D. *et al.* (2002) Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular Ecology*, **11**, 2037–2048.

Cooper A., Poinar H.N. (2000) Ancient DNA: do it right or not at all. *Science*, **289**, 1139–1139.

Cortey M., Pla C., García-Marín J. (2004) Historical biogeography of Mediterranean trout. *Molecular Phylogenetics and Evolution*, **33**, 831–844.

Cortey M., Vera M., Pla C., García-Marín J.L. (2009) Northern and Southern expansions of Atlantic brown trout (*Salmo trutta*) populations during the Pleistocene. *Biological Journal of the Linnean Society*, **97**, 904–917.

Dalén L., Nyström V., Valdiosera C., Germonpré M., Sablin M., Turner E. *et al.* (2007) Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox. *Proceedings of the National Academy of Sciences*, **104**, 6726–6729.

de Bruyn M., Hoelzel A.R., Carvalho G.R., Hofreiter M. (2011) Faunal histories from Holocene ancient DNA. *Trends in Ecology & Evolution*, **26**, 405–413.

Dudgeon D., Arthington H., Gessner M.O., Kawabata Z.I., Knowler D.J., Lévêque C. *et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, **81**, 163–182.

Durante S. (1978) Note on *Salmo trutta* L. in the Pleistocene of Praia a Mare, (Souther Italy). *Quaternaria*, **20**, 117–121.

Elliot J.M. (1994) *Quantitative ecology and the brown trout*. Oxford University Press, Oxford.

Ficke A.D., Myrick C.A., Hansen L.J. (2007) Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries*, **17**, 581–613.

Filipe A.F., Markovic D., Pletterbauer F., Tisseuil C., Wever A., Schmutz S. *et al.* (2013) Forecasting fish distribution along stream networks: brown trout (*Salmo trutta*) in Europe. *Diversity and Distribution*, **19**, 1059–1071.

Fisher T.G., Smith D.G., Andrew J.T. (2002) Preboreal oscillation caused by a glacial Lake Agassiz flood. *Quaternary Science Reviews*, **21**, 873–878.

Foote A.D., Kaschner K., Schultze S.E., Garilao C., Ho S.Y.W., Post K. *et al.* (2013) Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. *Nature Communications*, **4**, 1677.

Fruciano C., Pappalardo A.M., Tigano C., Ferrito V. (2014) Phylogeographical relationships of Sicilian brown trout and the effects of genetic introgression on morphospace occupation. *Biological Journal of the Linnean Society*, **112**, 387–398.

Hagelberg E., Hofreiter M., Keyser C. (2015) Ancient DNA: the first three decades. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **370**, 1660.

Halffman C.M., Potter B.A., McKinney H.J., Finney B.P., Rodrigues A.T., Yang D.Y., Kemp B.M. (2015) Early human use of anadromous salmon in North America at 11,500 y ago. *Proceedings of the National Academy of Sciences*, **112**, 12344–12348.

Hewitt G.M. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt G.M. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hughen K.A., Baillie M.G.L., Bard E., Bayliss A., Beck J.W., Bertrand C. *et al.* (2004) Marine04 Marine Radiocarbon Age Calibration, 0–26 cal kyr BP. *Radiocarbon*, **46**, 1059–1086.

IPCC 2014: *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Pachauri R.K., Meyer L.A. (Eds). IPCC, Geneva, Switzerland: 151.

Jonsson B., Jonsson N. (2011) *Climatic Effects on Atlantic Salmon and Brown Trout. Ecology of Atlantic Salmon and Brown Trout*. Springer: 473-515.

Knapp M., Clarke A.C., Horsburgh K.A., Matisoo-Smith E.A. (2012) Setting the stage-building and working in an ancient DNA laboratory. *Annals of Anatomy-Anatomischer Anzeiger*, **194**, 3–6.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Leigh J.W., Bryant D. (2015) popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110–1116.

Leonard J.A. (2008) Ancient DNA applications for wildlife conservation. *Molecular Ecology*, **17**, 4186–4196.

Lorenzen E.D., Nogués-Bravo D., Orlo L., Weinstock J., Binladen J., Marske K.A. *et al.* (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, **479**, 359–364.

Mannino M.A., Di Salvo R., Schimmenti V., Di Patti C., Incarbona A., Sineo L. *et al.* (2011) Upper Palaeolithic hunter-gatherer subsistence in Mediterranean coastal environments: an isotopic study of the diets of the earliest directly-dated humans from Sicily. *Journal of Archaeological Science*, **38**, 3094–3100.

Muñoz M., Casadevall M. (1997) Fish remains from Arbreda Cave (Serinyà Girona), northeast Spain, and their palaeoecological significance. *Journal of Quaternary Science*, **12**, 111–115.

Nelson J.S. (1994) Fishes of the World, 3rd edition. New York, Chichester, Brisbane, Toronto, Singapore, John Wiley & Sons, Inc.

Pääbo S., Poinar H., Serre D., Jaenicke-Després V., Hebler J., Rohland N. *et al.* (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics*, **38**, 645–679.

Pagès M., Desse-Berset N., Tougard C., Brosse L., Hänni C., Berrebi P. (2009) Historical presence of the sturgeon *Acipenser sturio* in the Rhône basin determined by the analysis of ancient DNA cytochrome b sequences. *Conservation Genetics*, **10**, 217–224.

Pereira M., Leadley P.W., Proença V., Alkemade R., Scharlemann J.P.W., Fernez-Manjarrés J.F. *et al.* (2010) Scenarios for global biodiversity in the 21st century. *Science*, **330**, 1496–1501.

Russ H., Jones A.K. (2009) Late Upper Palaeolithic fishing in the Fucino Basin, central Italy, a detailed analysis of the remains from Grotta di Pozzo. *Environmental Archaeology*, **14**, 155–162.

Sala O.E., Chapin F., Stuart A., Juan J., Berlow E., Bloomfield J. *et al.* (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

Scarciglia F., Robustelli G., Tiné V., Russa M.F.L., Abate M., Pezzino A. (2009) The role of human impacts and Holocene climate change in the Santuario della Madonna Cave (Calabria). *Méditerranée*, **112**, 137–143.

Snoj A., Marčeta B., Sušnik S., Melkič E., Meglič V., Dovč P. (2002) The taxonomic status of the "sea trout" from the north Adriatic Sea, as revealed by mitochondrial and nuclear DNA analysis. *Journal of Biogeography*, **29**, 1179–1185.

Snoj A., Marić S., Bajec Sušnik S., Berrebi P., Janjani S., Schöffmann J. (2011) Phylogeographic structure and demographic patterns of brown trout in North-West Africa. *Molecular Phylogenetics and Evolution*, **61**, 203–211.

Speller C.F., Yang D.Y., Hayden B. (2005) Ancient DNA investigation of prehistoric salmon resource utilization at Keatley Creek, British Columbia, Canada. *Journal of Archaeological Sciences*, **32**, 1378–1389.

Suárez J., Bautista J.M., Almodóvar A., Machordom A. (2001) Evolution of the mitochondrial control region in Palaearctic brown trout (*Salmo trutta*) populations: the biogeographical role of the Iberian Peninsula. *Heredity*, **87**, 198–206.

Thompson J.D., Higgins D.G., Gibson T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Turrero P., Horreo J., Garcia-Vazquez E. (2012) Same old *Salmo*? Changes in life history and demographic trends of North Iberian salmonids since the Upper Palaeolithic as revealed by archaeological remains and beast analyses. *Molecular Ecology*, **21**, 2318–2329.

Turrero P., Horreo J.L., López B., Pola I.G., Arbizu M., García Vázquez E. (2013) Chronological changes in Upper Paleolithic fisheries revealed by museum archival material. *Palaios*, **28**, 228–232.

Van Neer W., Wouters W., Germonpré M. (2007) Fish remains from three Upper Palaeolithic cave deposits in southern Belgium. *Anthropologica et Praehistorica*, **118**, 5–22.

Waples R.S., Pess G.R., Beechie T. (2008) Evolutionary history of Pacific salmon in dynamic environments. *Evolutionary Applications*, **1**, 189–206.

Weninger B., Jöris O. (2008) A 14C age calibration curve for the last 60 ka: the Greenland-Hulu U/Th timescale and its impact on understanding the Middle to Upper Paleolithic transition in Western Eurasia. *Journal of Human Evolution*, **55**, 772–781.

Xenopoulos M.A., Lodge D.M., Alcamo J., Märker M., Schulze K., Van Vuuren D.P. (2005) Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biol*ogy, **11**, 1557–1564.

Yang D.Y., Cannon A., Saunders S.R. (2004) DNA species identification of archaeological salmon bone from the Pacific Northwest Coast of North America. *Journal of Archaeological Science*, **31**, 619–631.

APPENDIX

Supplementary materials CHAPTER I

(See CHAPTER I for References)

Figures and Tables

Fig. S1 GENELAND software simulation results. A) Trend of the most probable number of clusters detected along the MCMC iterations performed with GENELAND (Guillot *et al.* 2005a). B). Histograms explain the number of simulation that showed the maximum probability with our genetic data for a population structure composed from 1 to 7 clusters.

A B

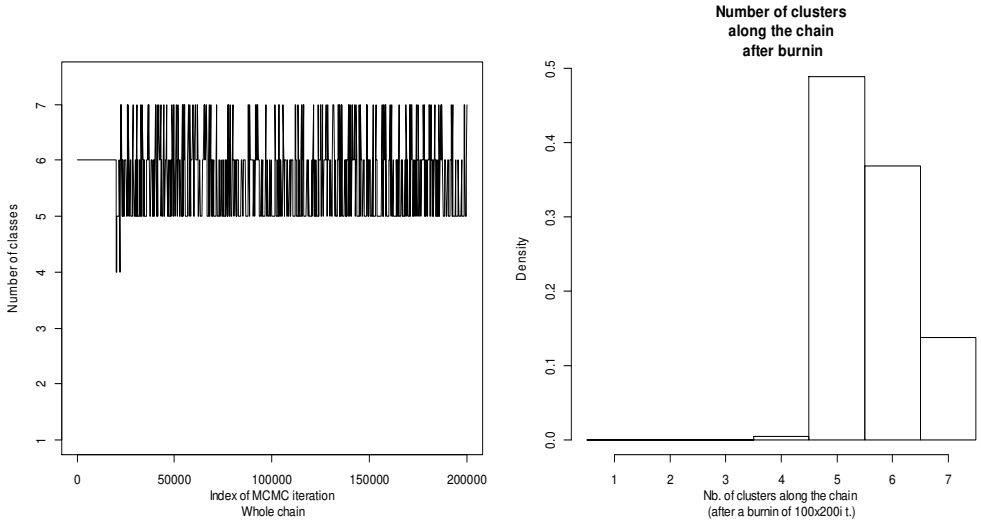


Fig.S2 STRUCTURE software simulation ΔK results. A) The most probable number of cluster estimated on the basis of all (neutral and non-neutral) loci analyzed. B) The most probable number of cluster estimated on the basis of all neutral loci analyzed.

A **B**

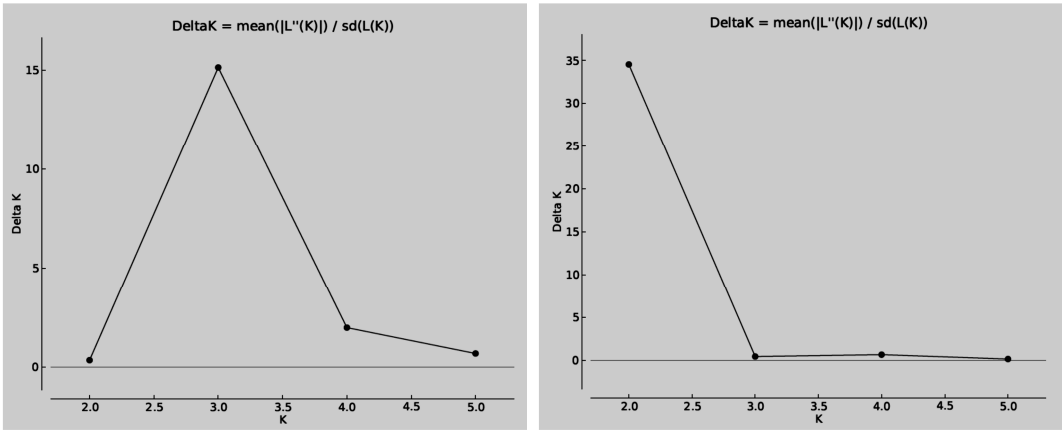
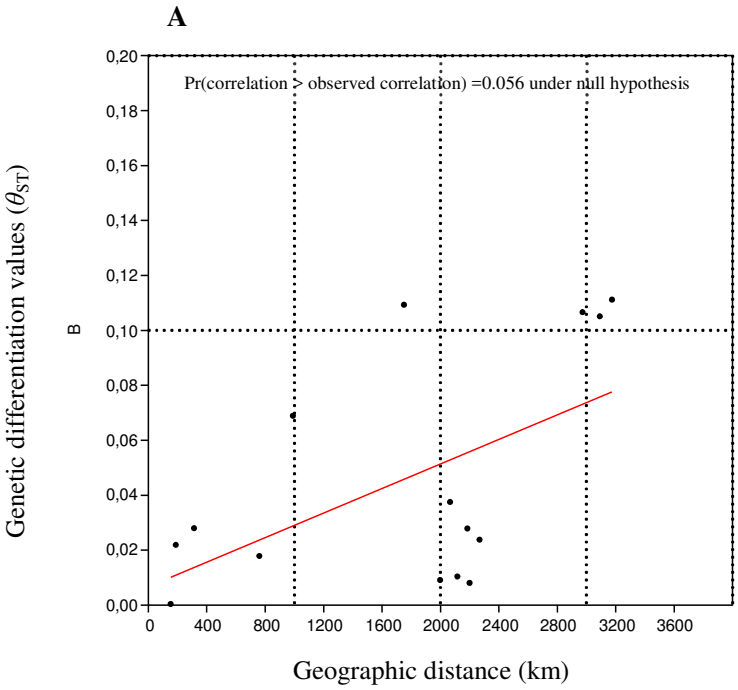


Fig.S3 Mantel Test. Graphical output of Mantel tests carried out using θ_{ST} (Fig.S3A) and ρR_{ST} (Fig.S3B).



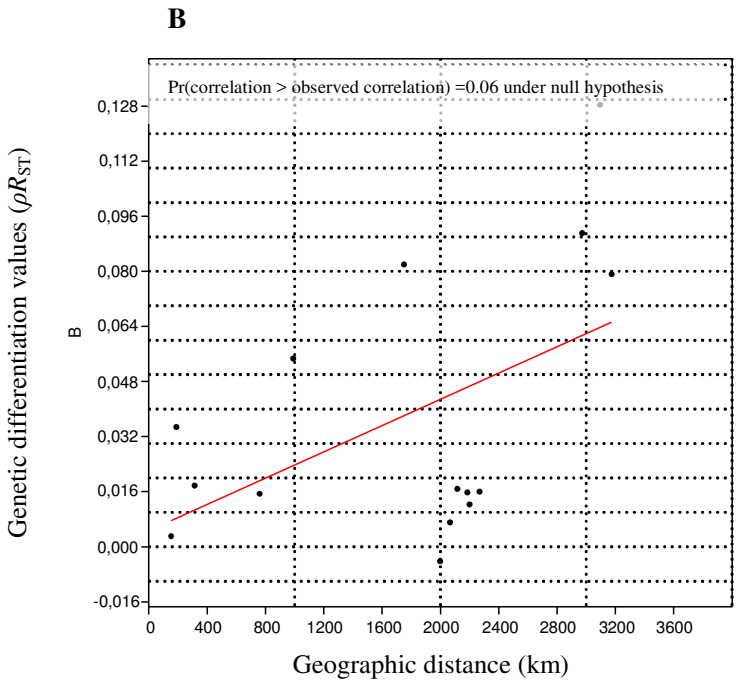


Table S1 Powsim simulation results.

	Fisher's exact test		Chi squared	
Fst	**all loci**	**neutral loci**	**all loci**	**neutral loci**
0,0000	0,0670	0,0580	0,0320	0,0290
0,0001	0,0810	0,0450	0,0550	0,0830
0,0002	0,0800	0,0710	0,1190	0,1100
0,0005	0,1990	0,1910	0,1830	0,1510
0,0010	0,4690	0,4030	0,4840	0,4160
0,0020	0,8920	0,8390	0,9340	0,9020
0,0050	1,0000	1,0000	1,0000	1,0000
0,0100	1,0000	1,0000	1,0000	1,0000

The table S1 explains the frequencies for each F_{ST} value from 0.0000 to 0.0100 where our data set was able to identify genetic differences between samples, comparing data from all loci (11 microsatellite loci) and from seven neutral loci. The tests were comparatively carried out applying both Fisher's exact test and Chi squared test.

Supplementary materials CHAPTER II

(see CHAPTER II for References)

File S1

Microsatellite PCR protocols

PCR conditions were optimized for the 14 microsatellite loci using a touchdown amplification profile. The PCR profileincluded an initial denaturation step at 92°C for 5 min, followed by 10 cycles of 92°C/30 s denaturation, 40 s annealing (see Table S1 for starting annealing temperature) and 72°C/50 s extension. The starting annealing temperature was reduced of 0.5°C/cycle. Additional 25 cycles were performed with annealing temperature fixed at 5°C less than the starting annealing temperature. A step of final elongation at 72°C for 7 min was performed.

Three multiplex PCR mixes were optimized as follows:

- multiplex A contained loci Ee2-91b, Ee2-407, EJ41.1 and Ee10 and was performed in a total volume of 20 μL.

- multiplex B1 contained loci EJ27.1, EJ35, Enja83 and Ee2-507 and was performed in a total volume of 20 μL.

- multiplex B2 contained loci EJ2, Ee2-135 and Eja17 and was performed in a total volume of 15 μL.

The PCR reactions for non-labelled loci (Enja148, Ee2-508 and Ee2-165b) were performed individually in a total volume of 10 μL, each.

PCR reaction mixture contained approximately 1-5 $ng \cdot \mu L^{-1}$of genomic DNA, 0.025 $U \cdot \mu L^{-1}$ of Taq DNA polymerase (MyTaq DNA Polymerase, Bioline), 0.25 μM of

each primer, 1× MyTaq Reaction Buffer (5mM dNTPs, 15 mM $MgCl_2$, stabilizers and enhancers). A blank control was included in each PCR plate.

Microsatellite Genotyping Procedures

Labelled loci were analyzed by means of an Automated sequencer ABI-PRISM3130xl Genetic Analyzer (Applied Biosystems) using LIZ500 as internal standard. The genotypes were obtained using the software PeakScanner™ v.1.0 (freely available from Applied Biosystems, http://www.appliedbiosystems.com/peakscanner.html). Over 10% of the global dataset were re-genotyped to assure consistency of results. Raw genotyped data were refined using the binning procedure implemented in Flexib in (Amos *et al.* 2007). Non-labelled loci were genotyped using a 5% Polyacrylamide (Acrylamide-Bis-Acrylamide 19:1) denaturing sequencing gel and visualized by the silver staining protocol proposed by Bassam *et al.* (1991). The allele sizes were defined by comparison withthe reaction sequence of pGEM-3Zf (+) (Promega sequencing kit). Reproducibility of the results were verified rerunning on each gel a subset of randomly chosen individualsthat represented about 10% of total samples. In addition, genotyping error rates for non-labelled loci were limitedby applying the allele ladder method as proposed by LaHood *et al.* (2002). The allele ladder is a pool of PCR products from multiple individuals and is representative of all or many ofthe alleles encountered in a given species for a particular locus (Moran *et al.* 2006). This method reduces the risks of misaligning alleles and consequently increases the accuracyof genotyping.

Microsatellite data Power Analysis

A simulation method implemented in an extended version of POWSIM 4.1 software (Ryman & Palm 2006) was used to assess the statistical power of the dataset, for detection of population differentiation. Tests were carried out testing the 15 localities using default parameters (1000 dememorizations, 100 batch and 1000 iterations per batch) and using several combinations of population divergence time (t) with an effective population size (N_e) of 1000 individuals (Ne/t: 1000/2, 1000/5; 1000/10, 1000/20.1, 1000/40.4; 1000/50.6, 1000/102.6). Each F_{ST} tested were simulated with 1000 replicates. The choice of testing for a fixed N_e of 1000 individuals was related to the detection of effective population size in European anchovy that are set in the low thousands of individuals (Ruggeri *et al.* 2016). Different set of markers were comaperd: 1) all loci (13 microsatellite loci), ii) nine neutral loci and iii) 4 candidate outliers. The tests were comparatively carried out applying both Fisher's exact test and Chi squared testHere below the Table with results:

	Fisher's exact test			Chi squared		
Fst	**all loci**	**neutral loci**	**outliers**	**all loci**	**neutral loci**	**outliers**
0.0005	0.4700	0.3100	0.2100	0.4300	0.2100	0.2100
0,0010	0.8600	0.6800	0.5600	0.7900	0.5300	0.5500
0,0025	1.0000	0.9900	1.0000	1.0000	0.9800	0.9800
0.0050	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0200	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0250	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.0500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The table explains the frequencies for each F_{ST} value from 0.0005 to 0.0500. The usage of all loci demonstrated to be adequate to detect F_{ST} values as low as 0.0010 in 79% or 86% of the replicates. Using exclusively neutral markers the power decreased (as expected) for the same F_{ST} vales to 53% and 68% of the replicates. The use of

exclusively candidate outliers (despite only 4 loci) allowed enough power to detect F_{ST} of at least 0.025.

Bibliographical support to File S1

Amos W., Hoffman J.I., Frodsham A., Zhang L., Best S., Hill A.V.S. (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes,* **7**, 10-14.

Bassam B.J., Caetano-Anollés G., Gresshoff P.M. (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analytical Biochemistry*, **196**, 80-83.

LaHood E.S., Moran P., Olsen J., Grant W.S., Park L.K. (2002) Microsatellite allele ladders in two species of Pacific salmon:preparation and field-test results. *Molecular Ecology Notes*, **2**, 187–190.

Moran P., Teel D.J., LaHood E.S., Drake J., Kalinowski S. (2006) Standardizing multi-laboratory microsatellite data in Pacific salmon: an historical view of the future. *Ecology of Freshwater Fish*, **15**, 597–605.

Ryman N., Palm S. (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, **6**, 600-602.

Ruggeri P., Splendiani A., Di Muri C., Fioravanti T., Santojanni A., Leonori I. *et al.* (2016) Coupling demographic and genetic variability from archived collections of European anchovy (*Engraulis encrasicolus*). *PLoS ONE*, **11**, e0151507.

Figures and Tables

Fig. S1A STRUCTURE Simulation based on the use of 13 loci (neutral plus candidate outliers) without the LocPrior function. a) Plot of values for simulated mean Logarithmic Probabilities from each K. b) Plot of best K based on Evanno method. c) Barplots of best K outcomes.

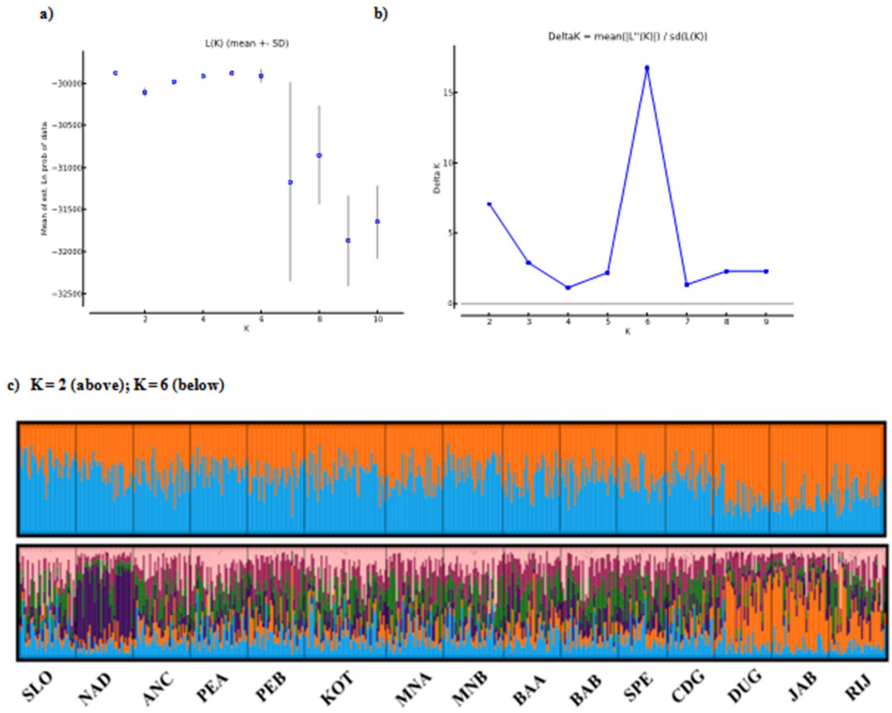


Fig. S1B STRUCTURE Simulation based on the use of 9 neutral loci without the LocPrior function. a) Plot of values for simulated mean Logarithmic Probabilities from each K. b) Plot of best K based on Evanno method. c) Barplots of best K outcomes.

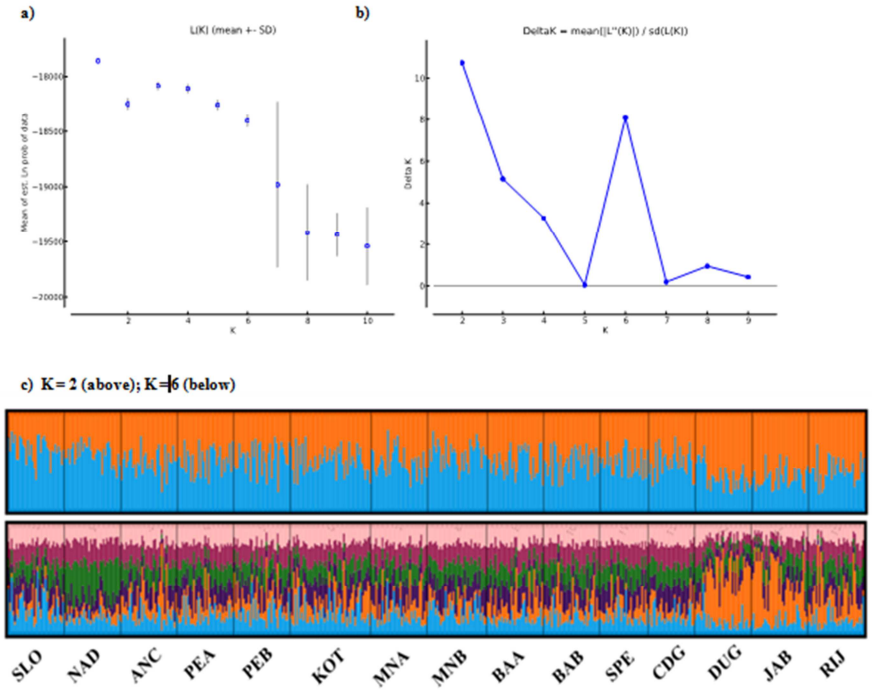


Fig. S1C STRUCTURE Simulation based on the use of 4 candidate outlier loci without the LocPrior function.

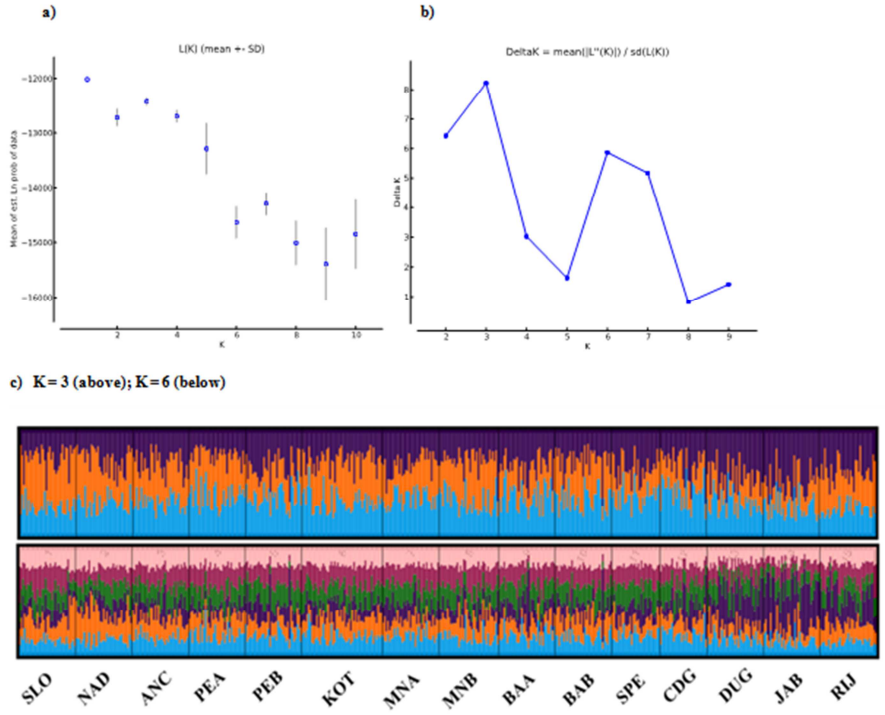


Fig. S1D STRUCTURESimulation based on the use of 13 loci (neutral plus candidate outliers) with the LocPrior function. a) Plot of values for simulated mean Logarithmic Probabilities from each K. b) Plot of best K based on Evanno method. c) Barplots of best K outcomes.

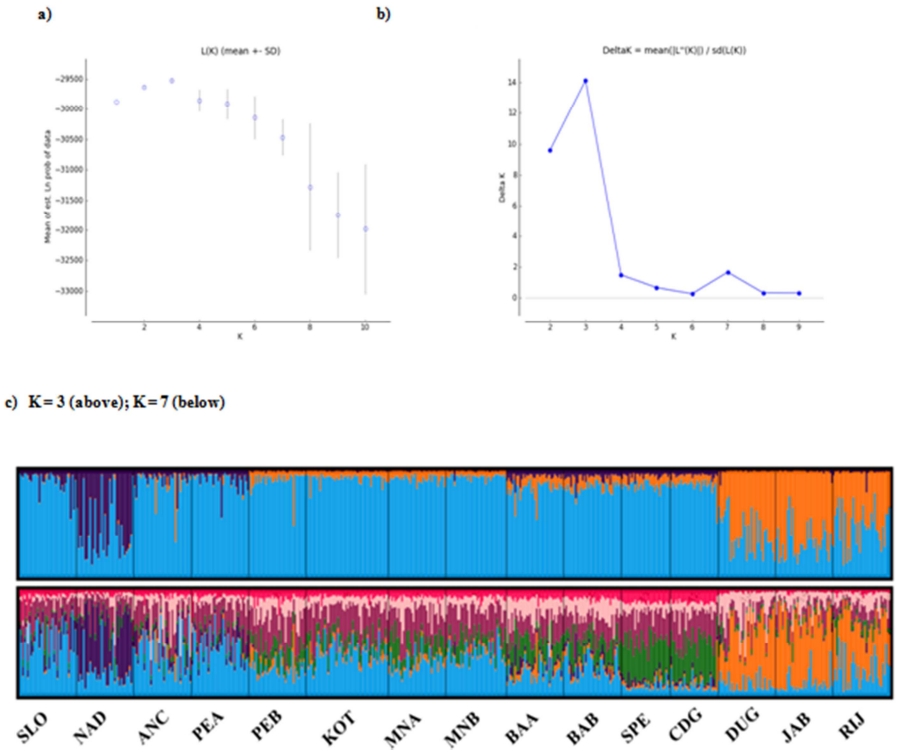


Fig. S1E STRUCTURE Simulation based on the use of 9 neutral loci with the LocPrior function. a) Plot of values for simulated mean Logarithmic Probabilities from each K. b) Plot of best K based on Evanno method. c) Barplots of best K outcomes.

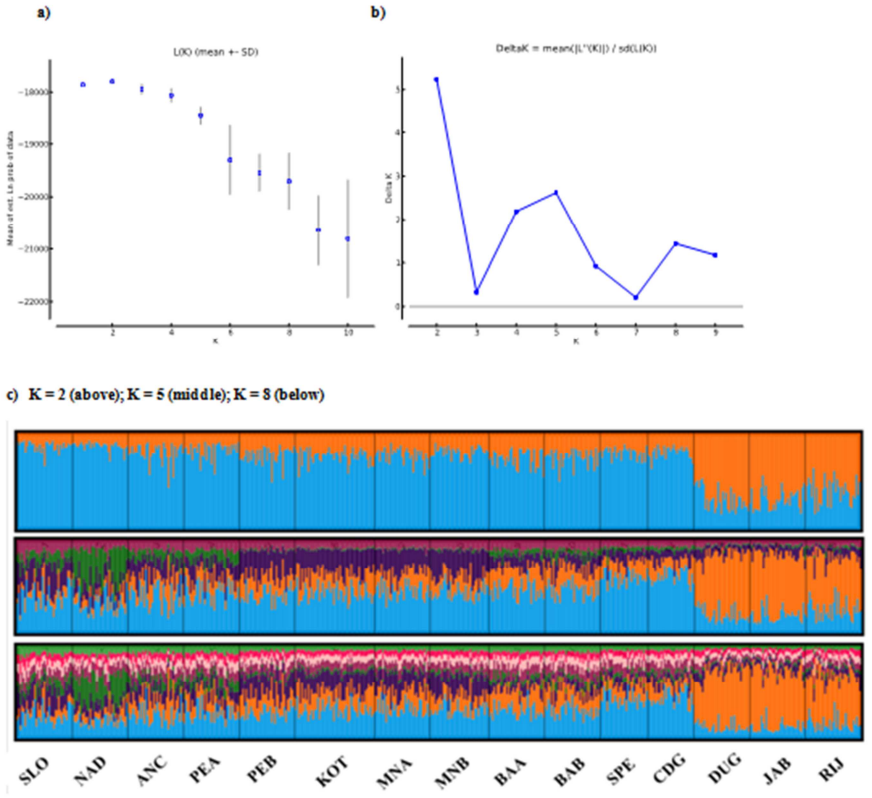


Fig. S1F STRUCTURE Simulation based on the use of 4 candidate outlier loci with the LocPrior function. a) Plot of values for simulated mean Logarithmic Probabilities from each K. b) Plot of best K based on Evanno method. c) Barplots of best K outcomes.

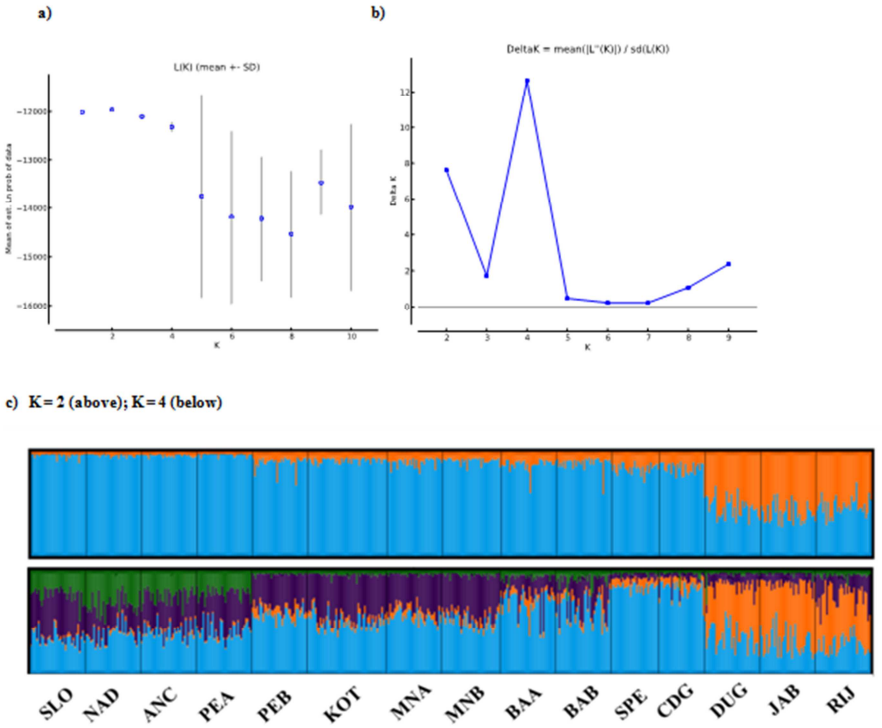
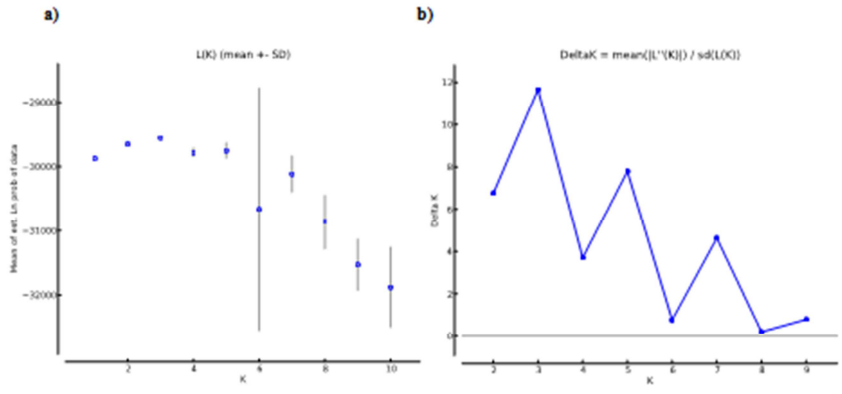
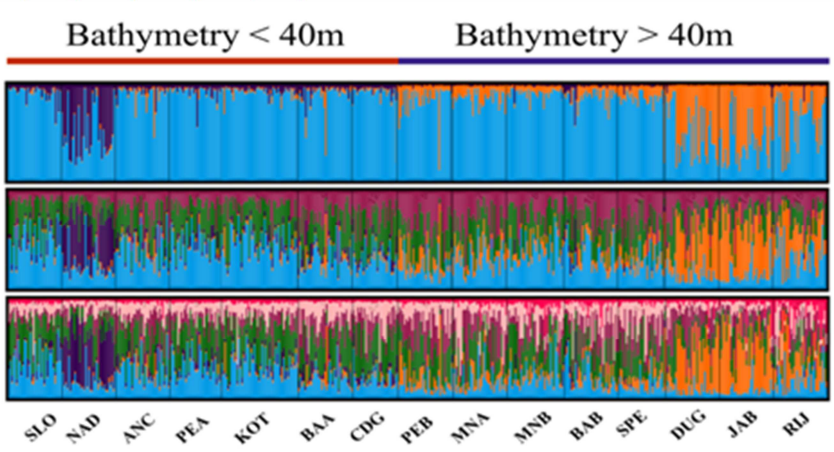


Fig. S1G STRUCTURE Simulation based on the use of all candidate outlier loci with the LocPrior function used to group sampling locations based on the bathymetry threshold of 40m. a) Plot of values for simulated mean Logarithmic Probabilities from each K. b) Plot of best K based on Evanno method. c) Barplots of best K outcomes. d) Barplot of K=2 outocome under the scenario of different bathymetry clustering.



c) K = 3 (above); K = 5 (middle); K = 7 (below)



d) K = 2

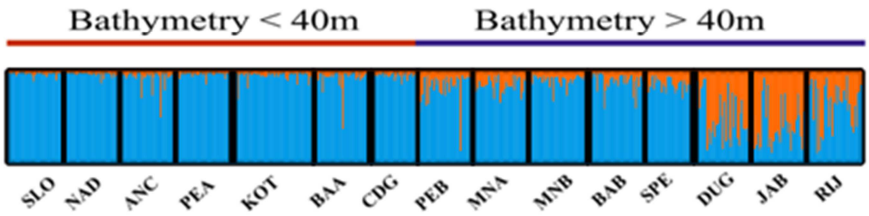


Fig. S2 Graphical plot of fdist/Lositan simulation. Blue dots represent interpolation between mean values of F_{ST} and H_E in each marker (microsatellite locus) analyzed. Blue dots falling within light grey area represent neutral markers whereas those within the yellow and the red areas represent outlier markers under balancing and directional selection. respectively.

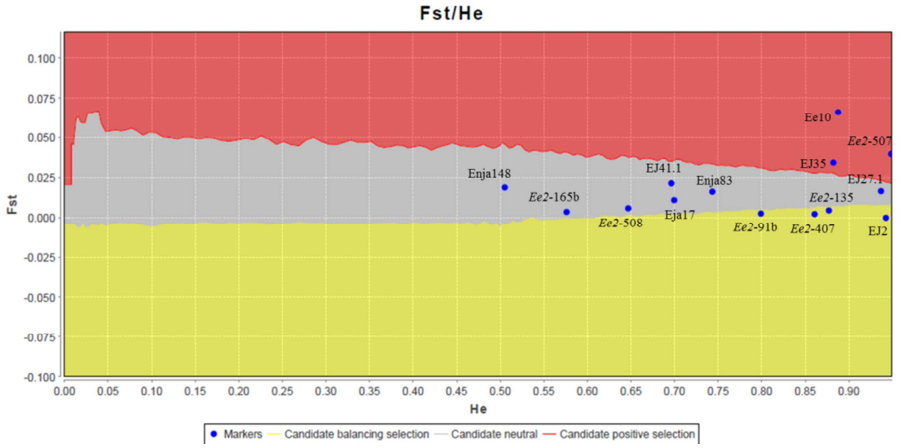
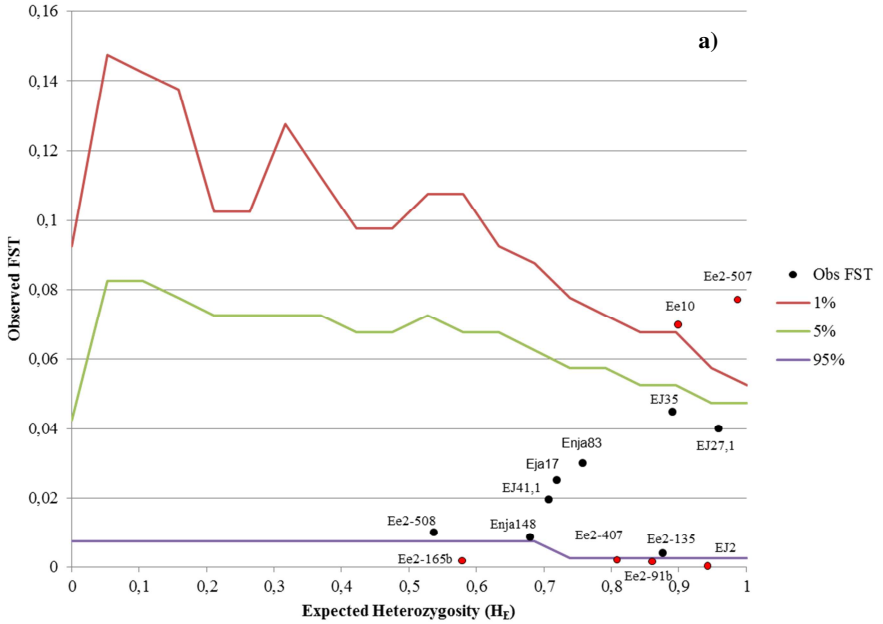


Fig. S3 Graphical plot of Hierarchical Island Model simulations under the assumption of population structured in two (K = 2; Fig. S3A) and three (K = 3; Fig. S3B) clusters. Dots represent the interpolation between mean values of F_{ST} and H_E in each marker (microsatellite locus) analyzed. Red, green and purple lines represent 1%, 5% and 95% percentiles of F_{ST} null distribution, respectively.



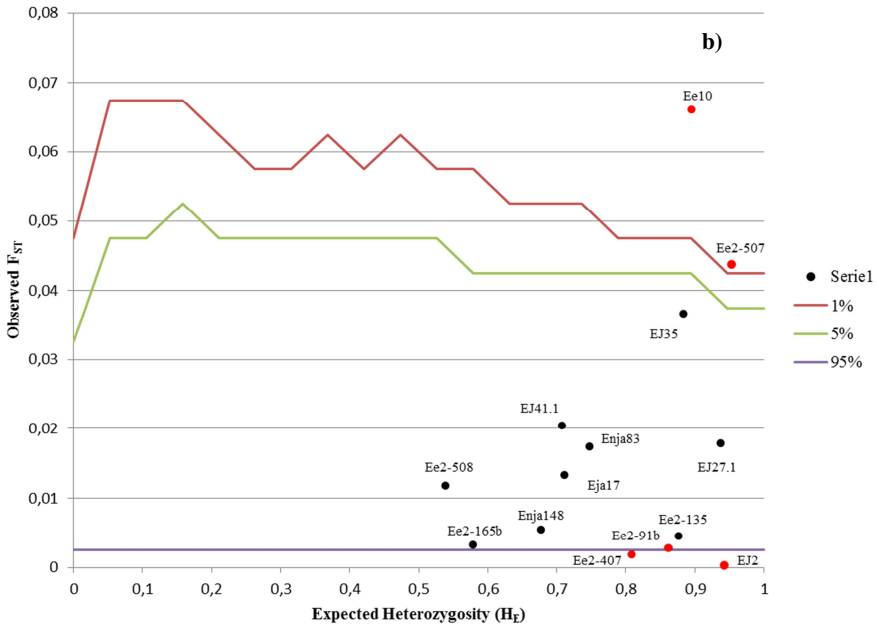

b)
0,08
0,07
0,06
0,05
0,04
0,03
0,02
0,01
0
Observed F_{ST}
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
Expected Heterozygosity (H_E)
Ee10
Ee2-507
EJ35
EJ41.1
Enja83
Ee2-508
EJ27.1
Eja17
Enja148
Ee2-135
Ee2-165b
Ee2-91b
Ee2-407
EJ2
Serie1
1%
5%
95%

Table S1 Summary of the microsatellite loci and primers for their PCR amplification used in the present study.

Locus Name	Primers Sequences	Repeat motifs	T_A	Fluor.	Multiplex	Species
Ee2-91b	F: GGTCTTGAGCTTGGCATAGG R: CCGGAAGACACTCTGCACAC	$(CCGCA)_7$	60	6-FAM	A	*E. encrasicolus*
Ee2-407	F: AGGAATCTCCTTCCCGTCTC R: GTGGGTCTGTGGGTGTTTTG	$(CA)_{13}$	60	VIC	A	*E. encrasicolus*
Ej41-1	F: TCTACCCCTGGAGGACACAC R: ACAGGGGGTTGAGAAAGAGG	$(CACAA)_8$	55	NED	A	*E. japonicus*
Ee10	F: GGTGGATGAAGTGGCAATCT R: CTGGGGTGGCATAACTGAAG	$[(GT)_9CT]_2$ $[(GT)_2CT]_3$	54	PET	A	*E. encrasicolus*
Ej27.1	F: GACTGTGAAGGAACGCTGGT R: AATAGGATTAGTCATCACAGGG	$(GA)_{36}$	58	6-FAM	B1	*E. japonicus*
Ej35	F: AGTGAGAGGACTCGCAAAGC R: CACACGAAGACAGACAAGCAA	$(TG)_{15}$	60	PET	B1	*E. japonicus*
Enja83	F: AAGGGACATCGGGTAGTGA R: AAGGCAAGTTCTCAGACGAG	$(AC)_6(TG)_7$	55	NED	B1	*E. japonicus*
Ee2-507	F: GGAAGGGACCTAGATGGAGTG R: ATCCCATTGATGTCCTGAGC	$(GAAA)_n$	60	VIC	B1	*E. encrasicolus*
Eja17	F: CCATTCAACTCCTCCCCAAGC R: GGCTCTTCAGCTCCCTGAGAC	$(CA)_7$	55	6-FAM	B2	*E. japonicus*
Ej2	F: AGCAAGGGAGCAAACAATC R: TGCAATTTGACAGAAACCACA	$(CT)_{43}$	58	NED	B2	*E. japonicus*
Ee2-135	F:AGGGCAGTGACAGGAGAGTC R: TCGTTACCCTGCGTTTATACTG	$(ATTAG)_{10}$	55	VIC	B2	*E. encrasicolus*
Ee2-165b	F: GGGTGGGTTAAAGATGAAGC R: AGGGATCTTCAGGGAACCAG	$(CCT)_7$	59			*E. encrasicolus*
Enja-148	F: CTCCATCTTCCGTATTCT R: GTAAATGATGTTGATGCTAA	$(TG)_6$	55			*E. japonicus*
Ee2-508	F: CACATGCTCGCTAAACATTG R: ACCTGATGCTGCTTGGTAGC	$(AGG)_8$	55			*E. encrasicolus*

Table S2 Values of the 13 environmental variables considered for each locality.

Localities	T0	T10	T20	T-samp	S0	S10	S20	S-samp	Oxyg0	Oxyg10	Oxyg20	Oxyg-samp	Chla
MNA	24.02	20.99	17.02	20.99	37.99	38.43	38.53	38.43	4.77	5.23	5.58	5.23	0.199
MNB	23.92	21.04	17.15	15.20	38.02	38.43	38.54	38.67	4.77	5.25	5.62	5.22	0.095
SLO	25.98	23.59	18.75	23.59	35.98	37.10	37.93	37.10	5.09	4.61	3.20	4.61	0.919
EMA	24.68	22.99	18.38	15.03	34.83	36.46	37.82	37.98	4.93	5.09	5.33	4.46	0.436
BAA	21.47	19.94	17.51	17.51	37.82	37.92	38.07	38.07	5.24	5.39	5.57	5.39	0.358
BAB	21.96	20.52	18.26	13.65	37.80	37.97	38.16	38.46	5.20	5.39	5.57	5.32	0.153
KOT	24.50	23.03	17.83	23.03	38.43	38.48	38.59	38.48	4.76	5.11	5.68	5.11	0.01
ANC	24.30	23.26	18.23	15.24	35.31	36.51	37.86	38.05	4.98	5.06	5.32	5.71	0.565
EMB	23.85	23.08	19.36	14.22	37.55	37.73	38.17	38.37	4.99	5.11	5.64	5.97	0.087
EMC	21.13	20.83	18.99	13.00	38.58	38.30	38.17	37.87	5.10	5.29	5.56	5.30	0.135
EMD	23.04	22.29	20.97	18.04	39.07	38.52	38.21	38.26	4.97	5.16	5.56	5.84	0.205
PEA	25.33	25.21	22.02	25.21	36.82	37.24	37.85	37.24	4.84	4.88	5.46	4.88	0.441
PEB	25.26	25.06	22.13	13.91	36.46	36.94	37.70	38.34	4.87	4.89	5.38	6.14	0.287
SPE	20.17	19.33	17.23	15.99	37.41	37.75	37.97	38.12	5.94	6.12	6.37	6.06	0.691
CDG	15.37	15.27	15.09	15.27	37.52	37.53	37.57	37.61	5.81	5.80	5.81	5.78	0.074

T0 = surface temperature (°C); T10 = 10 meters depth temperature (°C); T20 = 20 meters depth temperature (°C); T-samp = sampling depth temperature (°C); S0 = surface salinity (psu); S10 = 10 meters depth salinity (psu); S20 = 20 meters depth salinity (psu); S-samp = sampling depth salinity (psu); Oxyg0 = surface oxygen content (ml/l); Oxyg10 = 10 meters depth oxygen content (ml/l); Oxyg20 = 20 meters depth oxygen content (ml/l); Oxyg-samp = sampling depth oxygen content (ml/l); Chla = Chlorophyll-a surface concentration (mg/m^3).

Table S3 Summary of genetic variability observed at 14 microsatellite loci and a mtDNA genome portion in the sampled populations.

		MNA	*MNB*	*SLO*	*NAD*	*BAA*	*BAB*	*KOT*	*ANC*	*DUG*	*JAB*	*RIJ*	*PEA*	*PEB*	*SPE*	*CDG*	$N_{A\ TOT}$
Ee2-91b	N_A	8	9	9	10	7	10	11	10	8	9	7	8	10	8	6	15
	N	35	37	34	29	26	32	50	30	35	35	27	35	35	23	29	
	f_{null}	-0.029	0.041	-0.015	0.053	0.057	0.049	-0.064	0.037	-0.061	0.030	0.070	0.057	-0.015	0.068	-0.062	
	H_E	0.765	0.758	0.808	0.835	0.810	0.821	0.831	0.814	0.788	0.808	0.806	0.800	0.843	0.839	0.767	
	H_O	0.800	0.676	0.823	0.724	0.692	0.719	0.940	0.733	0.886	0.743	0.667	0.686	0.857	0.696	0.862	
	F_{IS}	-0.046	0.109	-0.019	0.135	0.148	0.126	-0.132	0.101	-0.126	0.081	0.176	0.144	-0.017	0.174	-0.126	
	R_S	6.84	6.56	7.58	9.06	6.71	8.99	8.72	8.59	7.03	7.67	6.89	7.06	8.89	7.63	5.77	7.60
Ee2-407	N_A	18	17	18	17	17	19	16	16	17	13	13	20	13	17	16	38
	N	35	37	35	34	35	34	50	35	35	35	35	35	35	30	28	
	f_{null}	0.053	-0.013	0.011	**0.062**	-0.020	**0.059**	0.013	0.049	0.006	0.048	-0.036	-0.001	-0.029	0.063	**0.075**	
	H_E	0.884	0.826	0.833	0.895	0.861	0.889	0.853	0.845	0.881	0.843	0.803	0.896	0.789	0.901	0.907	
	H_O	0.771	0.838	0.800	0.765	0.886	0.765	0.820	0.743	0.857	0.743	0.857	0.886	0.829	0.767	0.750	
	F_{IS}	0.129	-0.014	0.040	0.148	-0.030	0.141	0.039	0.122	0.028	0.120	-0.068	0.012	-0.051	0.151	0.176	
	R_S	13.57	11.88	13.61	13.66	12.65	14.81	11.54	12.10	12.56	10.81	9.64	14.80	9.74	13.66	13.45	12.57
EJ41.1	N_A	14	12	15	14	11	10	12	12	8	10	13	11	10	9	7	25
	N	35	31	34	35	35	34	50	35	30	28	31	35	33	30	29	
	f_{null}	-0.002	0.031	0.032	0.037	0.072	0.046	-0.019	0.079	0.099	**0.095**	0.061	-0.042	0.029	-0.026	-0.041	
	H_E	0.723	0.845	0.773	0.611	0.672	0.613	0.694	0.717	0.596	0.670	0.764	0.710	0.697	0.669	0.601	
	H_O	0.714	0.774	0.706	0.543	0.543	0.529	0.720	0.571	0.467	0.500	0.645	0.771	0.636	0.700	0.655	
	F_{IS}	0.013	0.085	0.088	0.113	0.194	0.138	-0.037	0.205	0.219	0.257	0.158	-0.088	0.088	-0.047	-0.092	
	R_S	9.71	10.16	11.48	10.14	8.24	7.33	8.21	9.05	6.62	8.61	10.28	8.59	8.22	7.38	5.51	8.64
Ee10	N_A	19	18	18	17	19	18	25	15	19	18	15	21	21	16	19	37
	N	35	35	35	32	35	33	50	35	34	31	28	35	35	30	27	
	f_{null}	0.042	0.060	-0.024	0.048	0.032	**0.063**	-0.040	0.051	-0.021	0.056	**0.114**	0.014	-0.064	-0.014	**0.065**	
	H_E	0.891	0.836	0.769	0.852	0.873	0.889	0.796	0.880	0.797	0.755	0.828	0.808	0.892	0.821	0.877	
	H_O	0.800	0.714	0.800	0.750	0.800	0.758	0.860	0.771	0.824	0.645	0.607	0.771	1.000	0.833	0.741	
	F_{IS}	0.104	0.147	-0.042	0.122	0.084	0.150	-0.081	0.124	-0.034	0.147	0.270	0.045	-0.123	-0.015	0.158	
	R_S	14.35	13.44	12.68	12.87	14.18	13.58	14.41	12.24	13.93	13.57	12.44	14.55	15.57	12.91	15.66	13.76

		MNA	MNB	SLO	NAD	BAA	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG	$N_{A\,TOT}$
EJ27.1	N_A	23	27	26	22	22	25	25	26	17	17	21	22	25	23	23	36
	N	35	37	35	34	35	35	49	35	35	34	34	35	35	29	28	
	f_{null}	0.099	0.032	0.002	0.010	-0.007	0.042	**0.058**	-0.022	-0.087	-0.054	-0.014	0.064	0.069	0.095	0.001	
	H_E	0.949	0.940	0.960	0.945	0.914	0.952	0.939	0.942	0.852	0.798	0.841	0.935	0.947	0.961	0.948	
	H_O	0.743	0.865	0.943	0.912	0.914	0.857	0.816	0.971	1.000	0.882	0.853	0.800	0.800	0.759	0.929	
	F_{IS}	**0.219**	0.081	0.018	0.036	0.000	0.101	0.132	-0.031	-0.177	-0.107	-0.015	0.147	0.157	**0.213**	0.021	
	R_S	18.97	19.53	20.55	17.81	16.63	19.77	18.66	19.07	13.39	12.89	15.86	17.51	19.05	19.90	19.27	17.92
EJ35	N_A	15	13	14	9	14	10	13	12	9	9	10	12	16	13	14	25
	N	35	37	35	35	35	34	50	35	35	35	35	35	35	30	29	
	f_{null}	0.013	-0.002	-0.005	-0.102	0.065	**0.064**	0.014	0.011	-0.104	-0.126	-0.005	0.034	-0.023	0.043	0.014	
	H_E	0.836	0.901	0.831	0.774	0.907	0.866	0.833	0.890	0.797	0.762	0.832	0.876	0.913	0.861	0.904	
	H_O	0.800	0.892	0.829	0.943	0.771	0.735	0.800	0.857	0.971	0.971	0.829	0.800	0.943	0.767	0.862	
	F_{IS}	0.044	0.010	0.004	-0.222	0.152	0.153	0.040	0.038	-0.223	-0.280	0.004	0.088	-0.033	0.111	0.048	
	R_S	11.42	11.60	10.43	7.99	12.38	8.84	9.75	10.37	7.64	7.72	8.74	10.09	13.44	10.89	12.79	10.27
Enja83	N_A	9	8	13	7	7	8	14	10	6	7	8	7	10	8	9	23
	N	31	36	35	35	35	35	48	31	35	34	35	35	35	30	27	
	f_{null}	**0.083**	-0.038	-0.064	-0.085	-0.021	0.058	-0.010	-0.006	-0.119	-0.047	-0.093	-0.014	0.020	-0.080	-0.015	
	H_E	0.771	0.722	0.784	0.616	0.746	0.739	0.803	0.776	0.695	0.639	0.683	0.757	0.761	0.741	0.766	
	H_O	0.613	0.778	0.886	0.743	0.771	0.629	0.813	0.774	0.886	0.706	0.829	0.771	0.714	0.867	0.778	
	F_{IS}	0.207	-0.078	-0.132	-0.210	-0.035	0.151	-0.012	0.002	-0.280	-0.106	-0.217	-0.019	0.062	-0.173	-0.016	
	R_S	7.93	6.60	9.74	5.57	6.44	6.87	9.55	8.54	5.43	5.73	6.57	6.32	8.28	7.08	7.73	7.23
Ee2-507	N_A	24	22	19	14	22	22	29	24	20	17	16	26	21	22	25	41
	N	35	37	35	35	35	35	50	34	35	35	34	35	35	30	29	
	f_{null}	0.001	0.027	0.031	-0.179	-0.015	-0.003	-0.009	-0.003	-0.051	-0.123	-0.089	0.003	0.012	0.053	0.041	
	H_E	0.958	0.957	0.930	0.707	0.957	0.951	0.953	0.949	0.862	0.793	0.821	0.963	0.952	0.952	0.959	
	H_O	0.943	0.892	0.857	1.000	0.971	0.943	0.960	0.941	0.943	1.000	0.971	0.943	0.914	0.833	0.862	
	F_{IS}	0.016	0.069	0.080	**-0.423**	-0.015	0.009	-0.008	0.008	-0.096	-0.266	-0.186	0.021	0.040	0.127	0.103	
	R_S	19.61	18.60	15.27	9.52	18.66	18.23	19.63	19.00	14.62	11.99	12.27	20.59	18.04	18.18	20.66	17.00
Eja17	N_A	8	8	7	7	6	6	9	7	7	7	6	9	8	5	6	15
	N	29	33	29	35	35	35	49	29	35	35	33	30	35	22	29	

	f_{null}	0.078	0.046	0.035	0.074	0.006	0.069	0.025	0.065	-0.027	0.066	0.035	0.041	0.055	0.040	0.047	
	H_E	0.733	0.790	0.765	0.765	0.706	0.698	0.702	0.747	0.595	0.600	0.610	0.750	0.764	0.672	0.676	
	H_O	0.586	0.697	0.690	0.543	0.686	0.571	0.653	0.621	0.629	0.486	0.546	0.667	0.657	0.591	0.586	
	F_{IS}	0.203	0.119	0.100	0.198	0.029	0.184	0.070	0.171	-0.057	0.192	0.108	0.112	0.141	0.124	0.135	
	R_S	6.85	6.97	6.49	5.68	5.67	5.05	6.76	6.73	5.42	6.00	4.97	7.39	7.19	4.73	5.53	6.10

		MNA	***MNB***	***SLO***	***NAD***	***BAA***	***BAB***	***KOT***	***ANC***	***DUG***	***JAB***	***RIJ***	***PEA***	***PEB***	***SPE***	***CDG***	$N_{A\,TOT}$
EJ2	N_A	21	20	15	22	21	21	25	22	22	18	19	21	24	20	23	32
	N	35	37	35	34	35	35	50	35	34	35	35	35	35	30	29	
	f_{null}	-0.007	0.018	0.068	0.012	0.025	-0.007	0.014	0.025	0.076	0.049	0.096	0.007	0.015	-0.020	0.038	
	H_E	0.944	0.939	0.913	0.950	0.947	0.942	0.936	0.947	0.955	0.936	0.940	0.941	0.957	0.944	0.951	
	H_O	0.943	0.892	0.771	0.912	0.886	0.943	0.900	0.886	0.794	0.829	0.743	0.914	0.914	0.967	0.862	
	F_{IS}	0.001	0.050	0.157	0.040	0.066	-0.001	0.038	0.065	0.170	0.116	**0.212**	0.029	0.045	-0.024	0.095	
	R_S	17.36	17.00	13.07	17.89	17.60	16.78	17.60	17.81	18.29	15.90	16.01	17.03	19.26	16.95	19.15	17.18
Ee2-135	N_A	12	14	13	10	10	12	13	11	11	10	10	12	10	12	12	16
	N	35	37	35	35	35	35	50	35	35	35	34	35	35	30	29	
	f_{null}	-0.046	0.036	-0.015	0.038	-0.001	-0.069	-0.029	-0.043	-0.002	-0.024	-0.014	-0.074	-0.006	-0.080	-0.063	
	H_E	0.897	0.891	0.899	0.854	0.869	0.884	0.874	0.875	0.865	0.854	0.870	0.875	0.858	0.866	0.865	
	H_O	0.971	0.811	0.914	0.771	0.857	1.000	0.920	0.943	0.857	0.886	0.882	1.000	0.857	1.000	0.966	
	F_{IS}	-0.084	0.091	-0.018	0.098	0.014	-0.133	-0.053	-0.079	0.009	-0.038	-0.014	-0.145	0.001	-0.158	-0.199	
	R_S	10.73	11.39	11.38	8.94	9.31	10.06	10.63	9.91	9.41	9.00	9.24	10.11	8.75	10.42	10.43	9.98
Ee2-508	N_A	6	7	10	6	6	5	8	6	6	8	7	8	6	4	7	11
	N	27	31	24	27	23	27	38	19	34	35	34	30	25	21	29	
	f_{null}	0.069	0.079	**0.127**	**0.095**	**0.118**	0.072	**0.077**	**0.109**	0.078	0.039	0.052	0.065	**0.102**	0.098	0.057	
	H_E	0.724	0.621	0.832	0.732	0.691	0.648	0.691	0.671	0.604	0.645	0.559	0.723	0.661	0.652	0.620	
	H_O	0.593	0.484	0.583	0.556	0.478	0.519	0.553	0.474	0.471	0.571	0.471	0.600	0.480	0.476	0.517	
	F_{IS}	0.184	0.223	0.303	0.245	0.313	0.203	0.202	0.300	0.223	0.116	0.161	0.173	0.278	0.274	0.168	
	R_S	5.62	6.12	9.53	5.97	5.80	4.89	6.83	6.00	5.25	6.21	5.45	7.11	5.69	4.00	5.62	6.00
Ee2-165b	N_A	5	5	7	4	4	6	5	5	8	5	4	4	6	6	5	12
	N	35	37	35	33	35	35	49	35	34	32	35	35	35	30	29	
	f_{null}	0.002	0.080	-0.032	-0.015	-0.027	0.042	-0.029	0.051	0.079	-0.042	-0.013	-0.026	0.028	-0.059	0.023	
	H_E	0.525	0.654	0.642	0.591	0.565	0.649	0.592	0.579	0.638	0.630	0.559	0.511	0.595	0.423	0.456	

	H_O	0.514	0.514	0.686	0.606	0.600	0.571	0.633	0.486	0.500	0.688	0.571	0.543	0.543	0.500	0.414	
	F_{IS}	0.020	0.217	-0.069	-0.026	-0.063	0.121	-0.069	0.163	0.219	-0.093	-0.022	-0.063	0.089	-0.185	0.093	
	R_S	4.35	4.96	5.57	3.97	3.90	5.44	4.31	4.53	6.29	4.58	3.89	3.45	5.04	5.24	4.54	4.67
Enja-148	N_A	6	5	6	5	4	6	7	4	7	5	5	4	4	6	5	16
	N	28	36	23	32	23	35	34	19	21	27	23	35	25	22	18	
	f_{null}	**0.122**	0.075	**0.151**	0.005	**0.173**	0.018	**0.140**	**0.168**	**0.176**	**0.104**	**0.134**	**0.097**	**0.127**	**0.122**	**0.176**	
	H_E	0.563	0.383	0.617	0.468	0.487	0.482	0.490	0.548	0.656	0.460	0.587	0.491	0.682	0.571	0.532	
	H_O	0.393	0.278	0.391	0.438	0.261	0.486	0.294	0.316	0.400	0.321	0.391	0.343	0.480	0.409	0.278	
	F_{IS}	0.306	0.278	0.371	0.067	0.470	-0.009	0.403	0.430	0.397	0.360	0.339	0.306	0.300	0.288	0.485	
	R_S	5.51	3.50	5.69	4.30	4.00	4.54	5.27	3.95	6.78	4.27	4.78	3.41	7.35	5.78	5.00	5.45

		MNA	***MNB***	***SLO***	***NAD***	***BAA***	***BAB***	***KOT***	***ANC***	***DUG***	***JAB***	***RIJ***	***PEA***	***PEB***	***SPE***	***CDG***	$N_{A\,TOT}$
Average*	$N_{A\,M}$	14.0	13.9	14.2	12.2	12.8	13.2	15.8	13.5	12.2	11.4	11.5	13.9	13.9	12.5	13.2	
	H_E	0.815	0.821	0.826	0.722	0.809	0.811	0.808	0.818	0.763	0.749	0.763	0.811	0.818	0.792	0.792	
	H_O	0.753	0.756	0.791	0.751	0.758	0.734	0.799	0.752	0.776	0.742	0.728	0.781	0.780	0.750	0.753	
	F_{IS}	**0.077**	**0.081**	0.043	0.028	0.064	**0.096**	0.011	**0.082**	-0.016	0.009	0.046	0.038	0.046	0.054	0.051	
	R_S	9.8	9.7	9.8	8.6	9.2	9.4	9.8	8.4	8.1	8.2	9.6	9.8	9.3	10.7	11.2	

N_{ATOT} = Number of alleles observed in a specific locus; N_A = number of alleles observed per location; N = number of individuals correctly genotyped; N_{AM} = mean number of alleles observed per location; f_{null} = null allele frequency;H_O = observed heterozygosity; H_E = expected heterozygosity; F_{IS} = inbreeding coefficient estimates; R_S = allelic richness estimates standardized at 19 individuals. Bold F_{IS}values are significant (<0.05) after a sequential Bonferroni correction (Weir & Cockerham 1984). Bold f_{null}values are those still showed null allele signals after applying Brookfield null allele correction method (Dempster *et al.* 1977); * The average was produced on the basis of all microsatellite loci excluding Enja-148.For the local sample codes, see Table 1-A.

Table S4A Hedrick`s $G`_{ST}$ (Jost 2008) values are shown below the diagonal, while Nei`sG_{ST} (Hedrick 2005) values above the diagonal. Significant values (P<0.0005) are shown in bold.

	MNA	MNB	SLO	NAD	BAA	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG
MNA	/	0.008	**0.015**	**0.031**	-0.001	0.003	0.009	0.004	**0.017**	**0.025**	**0.017**	0.008	0.002	0.003	-0.001
MNB	0.004	/	0.009	**0.036**	**0.013**	0.010	0.006	0.004	**0.035**	**0.040**	**0.015**	0.005	**0.015**	**0.021**	**0.017**
SLO	**0.008**	0.004	/	**0.037**	**0.016**	0.009	0.005	**0.014**	**0.042**	**0.048**	**0.020**	0.005	**0.020**	**0.029**	**0.030**
NAD	**0.016**	**0.019**	**0.019**	/	**0.025**	**0.020**	**0.036**	**0.030**	**0.048**	**0.057**	**0.038**	**0.030**	**0.029**	**0.035**	**0.027**
BAA	-0.001	**0.006**	**0.008**	**0.012**	/	0.003	**0.013**	0.005	0.011	**0.018**	**0.015**	**0.010**	-0.004	0.004	-0.001
BAB	0.001	0.005	0.005	**0.010**	0.001	/	0.006	0.002	**0.019**	**0.029**	**0.011**	0.004	0.004	0.008	0.007
KOT	0.005	0.003	0.003	**0.018**	**0.006**	0.003	/	0.008	**0.026**	**0.034**	0.010	0.002	**0.015**	**0.021**	**0.017**
ANC	0.002	0.002	**0.007**	**0.015**	0.002	0.001	0.004	/	**0.018**	**0.027**	**0.014**	0.007	0.004	0.011	0.005
DUG	**0.009**	**0.018**	**0.021**	**0.025**	0.006	**0.010**	**0.013**	**0.009**	/	-0.003	**0.014**	**0.032**	**0.015**	**0.021**	**0.019**
JAB	**0.013**	**0.021**	**0.025**	**0.030**	**0.009**	**0.015**	**0.017**	**0.014**	-0.002	/	**0.016**	**0.039**	**0.022**	**0.027**	**0.028**
RIJ	**0.008**	**0.007**	**0.010**	**0.019**	**0.007**	**0.006**	0.005	**0.007**	**0.007**	**0.008**	/	**0.013**	**0.021**	**0.025**	**0.022**
PEA	0.004	0.003	0.003	**0.015**	**0.005**	0.002	0.001	0.003	**0.016**	**0.020**	**0.007**	/	**0.017**	**0.016**	**0.013**
PEB	0.001	**0.007**	**0.010**	**0.014**	-0.002	0.002	**0.008**	0.002	**0.007**	**0.011**	**0.011**	**0.008**	/	0.007	0.006
SPE	0.002	**0.011**	**0.014**	**0.018**	0.002	0.004	**0.010**	0.005	**0.011**	**0.014**	**0.013**	**0.008**	0.004	/	-0.001
CDG	0.000	**0.008**	**0.015**	**0.013**	0.000	0.003	**0.008**	0.002	**0.010**	**0.014**	**0.011**	**0.006**	0.003	0.000	/

Table S4B Jost`s D (Peakall & Smouse 2012), D_{EST} values. Significant values (P<0.0005) are shown in bold.

	MNA	MNB	SLO	NAD	BAA	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG
MNA	/														
MNB	0.038	/													
SLO	**0.070**	0.042	/												
NAD	**0.122**	**0.149**	**0.151**	/											
BAA	-0.006	**0.057**	**0.072**	**0.095**	/										
BAB	0.013	**0.046**	0.041	**0.078**	0.011	/									
KOT	0.039	0.027	0.024	**0.139**	**0.054**	0.024	/								
ANC	0.018	0.020	**0.065**	**0.120**	0.020	0.008	0.033	/							
DUG	**0.066**	**0.139**	**0.169**	**0.166**	0.042	**0.072**	**0.098**	**0.069**	/						
JAB	**0.093**	**0.154**	**0.187**	**0.194**	**0.066**	**0.107**	**0.123**	**0.099**	-0.009	/					
RIJ	**0.063**	**0.056**	**0.079**	**0.130**	**0.054**	**0.042**	0.035	**0.053**	**0.047**	**0.049**	/				
PEA	0.033	0.023	0.024	**0.118**	**0.043**	0.018	0.007	0.030	**0.123**	**0.145**	**0.049**	/			
PEB	0.011	**0.069**	**0.095**	**0.114**	-0.019	0.016	**0.068**	0.017	**0.057**	**0.080**	**0.081**	**0.074**	/		
SPE	0.013	**0.091**	**0.125**	**0.132**	0.016	0.032	**0.084**	0.045	**0.077**	**0.093**	**0.092**	**0.065**	0.031	/	
CDG	-0.004	**0.070**	**0.131**	**0.098**	-0.004	0.027	**0.069**	0.021	**0.069**	**0.097**	**0.078**	**0.052**	0.025	-0.003	/

Table S5A

RECEIVING SAMPLING LOCATIONS	NATIVE SAMPLING LOCATIONS														
	MNA	**MNB**	**SLO**	**NAD**	**BAA**	**BAB**	**KOT**	**ANC**	**DUG**	**JAB**	**RIJ**	**PEA**	**PEB**	**SPE**	**CDG**
MNA												1			
MNB								1							
SLO		1						1					1		
NAD	1	1						2		1		1			
BAA														1	
BAB	2						1		1					1	
KOT					1						1				
ANC											1	1			
DUG	1	1			1			1							
JAB	3														
RIJ	2						1		1						
PEA		1						1			1		1		
PEB	1														
SPE					1	1									
CDG	1				1										

Table S5B

THETA (ϴ)	MEAN VALUE	CI 2.5%	CI 97.5%
ANC	0.015	0.007	0.018
BAA	0.010	0.000	0.013
BAB	0.039	0.029	0.046
CDG	0.003	0.001	0.005
NAD	0.023	0.013	0.022
DUG	0.014	0.009	0.017
JAB	0.046	0.029	0.045
RIJ	0.026	0.019	0.030
KOT	0.009	0.000	0.012
MNA	0.005	0.003	0.007
MNB	0.028	0.036	0.052
PEA	0.008	0.004	0.012
PEB	0.031	0.000	0.005
SLO	0.012	0.006	0.014
SPE	0.023	0.002	0.012

Table S5C

M VALUES	ANC	BAA	BAB	CDG	NAD	DUG	JAB	RIJ	KOT	MNA	MNB	PEA	PEB	SLO	SPE
ANC	/	163.06	170.35	46.19	697.85	464.88	58.11	168.56	146.28	566.37	239.54	**667.99**	170.57	25.77	478.64
BAA	155.30	/	61.19	777.57	156.79	361.36	47.07	139.09	502.98	233.53	227.70	238.82	216.61	247.11	332.05
BAB	362.36	**646.64**	/	516.40	40.57	211.32	120.00	187.39	48.76	230.92	627.63	119.96	102.56	308.10	318.94
CDG	**735.09**	19.15	**567.70**	/	142.35	727.25	90.30	252.11	87.50	199.23	100.11	360.42	270.02	261.80	**578.03**
NAD	76.90	263.34	66.87	122.71	/	136.52	18.50	138.27	**736.04**	79.93	70.51	323.79	145.04	**538.47**	556.55
DUG	395.07	177.90	317.18	452.48	190.96	/	149.17	64.59	54.97	379.27	104.01	116.00	78.69	323.77	106.66
JAB	49.91	310.02	277.80	556.76	151.06	499.03	/	101.87	156.45	74.80	151.51	648.48	93.63	154.08	103.00
RIJ	686.49	88.44	128.86	44.03	231.50	488.34	298.34	/	209.86	**786.42**	541.27	341.93	184.06	316.06	200.39
KOT	461.10	255.46	295.46	116.44	797.84	122.65	229.07	60.62	/	106.99	489.36	99.91	686.56	207.59	63.73
MNA	519.89	116.15	174.76	**902.00**	76.19	232.29	299.95	317.28	141.25	/	366.87	450.95	85.21	114.55	91.20
MNB	459.43	381.29	101.35	177.92	614.20	89.03	535.04	**723.41**	329.71	142.10	/	187.26	730.60	152.29	304.12
PEA	175.20	63.13	144.25	711.34	691.86	**944.01**	486.92	103.28	61.16	49.16	485.79	/	548.78	120.47	442.16
PEB	440.48	282.40	37.96	209.35	**892.44**	86.93	**620.28**	64.73	89.10	419.68	259.70	230.42	/	215.24	40.16
SLO	220.45	159.60	188.16	695.52	31.48	486.81	258.91	138.39	166.79	44.22	239.29	205.58	240.68	/	144.59
SPE	381.44	85.30	322.63	113.65	72.79	663.58	705.12	422.21	477.59	192.24	**882.87**	363.45	**834.12**	401.60	/
	5119.11 source	3011.88	2854.50	**5442.34** source	4787.87	**5513.99** source	3916.77	2881.78	3208.45	3504.85	4786.15	4354.95	4387.14	3386.90	3760.22

Table S6

Locus	**MNA *vs*** MNB	SLO	NAD	BAA	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG
Ee91b	0.07	-0.42	-0.73	-0.44	-0.57	-0.69	-0.49	-0.21	-0.42	-0.40	-0.33	-0.83	-0.78	-0.02
Ee407	0.82	0.74	-0.21	0.37	-0.09	0.48	0.59	0.04	0.62	1.08	-0.23	1.22	-0.33	-0.46
EJ41,1	-1.21	-0.42	0.77	0.38	0.76	0.22	0.05	0.86	0.39	-0.34	0.10	0.20	0.40	0.83
Ee10	0.84	1.55	0.62	0.32	0.04	1.28	0.21	1.27	1.67	0.93	1.17	-0.02	1.01	0.25
EJ27,1	0.32	-0.51	0.13	1.03	-0.13	0.33	0.23	2.14	**2.78**	2.29	0.46	0.06	-0.53	0.02
EJ35	-1.02	0.06	0.66	-1.16	-0.41	0.04	-0.82	0.45	0.78	0.06	-0.57	-1.29	-0.34	-1.10
Enja83	0.41	-0.13	1.14	0.22	0.27	-0.32	-0.05	0.61	0.99	0.70	0.12	0.09	0.26	0.04
Ee507	0.05	1.00	**3.96**	0.04	0.29	0.23	0.37	2.39	**3.22**	**2.92**	-0.25	0.26	0.26	-0.05
Eja17	-0.51	-0.27	0.43	0.21	0.26	0.23	-0.11	0.94	0.91	0.84	-0.14	-0.26	0.45	0.42
EJ2	0.18	0.88	-0.22	-0.12	0.06	0.27	-0.11	-0.44	0.26	0.11	0.10	-0.52	-0.01	-0.28
Ee135	0.12	-0.02	0.71	0.49	0.25	0.41	0.40	0.55	0.72	0.48	0.39	0.65	0.55	0.56
Ee508m	0.71	-1.04	-0.07	0.25	0.54	0.25	0.38	0.82	0.56	1.07	0.01	0.45	0.52	0.72
Ee2-165b	-0.77	-0.69	-0.37	-0.23	-0.73	-0.38	-0.30	-0.66	-0.61	-0.19	0.07	-0.40	0.54	0.37

	MNB *vs*												
Locus	SLO	NAD	BAA	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG
Ee91b	-0.49	-0.80	-0.51	-0.63	-0.76	-0.56	-0.28	-0.48	-0.47	-0.40	-0.90	-0.85	-0.09
Ee407	-0.08	-1.03	-0.45	-0.90	-0.34	-0.23	-0.77	-0.20	0.26	-1.05	0.41	-1.15	-1.27
EJ41,1	0.79	1.98	1.59	1.97	1.43	1.27	2.07	1.60	0.87	1.32	1.42	1.61	2.04
Ee10	0.71	-0.22	-0.52	-0.80	0.44	-0.63	0.43	0.84	0.09	0.33	-0.86	0.17	-0.59
EJ27,1	-0.84	-0.19	0.71	-0.46	0.01	-0.10	1.82	2.45	1.96	0.14	-0.26	-0.86	-0.30
EJ35	1.08	1.69	-0.14	0.61	1.06	0.20	1.47	1.80	1.08	0.45	-0.27	0.68	-0.08
Enja83	-0.53	0.73	-0.19	-0.14	-0.73	-0.46	0.21	0.58	0.29	-0.29	-0.32	-0.15	-0.37
Ee507	0.95	**3.91**	-0.01	0.24	0.18	0.32	2.34	**3.17**	**2.87**	-0.30	0.21	0.21	-0.10
Eja17	0.24	0.94	0.72	0.77	0.74	0.40	1.44	1.42	1.35	0.37	0.25	0.96	0.93
EJ2	0.71	-0.39	-0.30	-0.12	0.10	-0.28	-0.62	0.09	-0.06	-0.07	-0.69	-0.19	-0.45
Ee135	-0.14	0.59	0.37	0.13	0.29	0.28	0.44	0.60	0.36	0.27	0.53	0.43	0.44
Ee508m	-1.75	-0.78	-0.46	-0.17	-0.46	-0.33	0.10	-0.15	0.36	-0.70	-0.26	-0.19	0.00
Ee2-165b	0.08	0.39	0.54	0.03	0.39	0.46	0.11	0.16	0.57	0.84	0.37	1.30	1.13

Locus	**SLO vs** NAD	BAA	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG
Ee91b	-0.31	-0.02	-0.14	-0.27	-0.07	0.21	0.01	0.02	0.09	-0.41	-0.36	0.41
Ee407	-0.95	-0.37	-0.83	-0.26	-0.15	-0.70	-0.13	0.33	-0.97	0.48	-1.07	-1.20
EJ41,1	1.19	0.80	1.18	0.64	0.47	1.28	0.81	0.08	0.53	0.63	0.82	1.25
Ee10	-0.93	-1.23	-1.51	-0.27	-1.35	-0.28	0.12	-0.62	-0.39	-1.57	-0.54	-1.30
EJ27,1	0.65	1.54	0.38	0.85	0.74	**2.65**	**3.29**	**2.80**	0.97	0.58	-0.02	0.54
EJ35	0.61	-1.21	-0.47	-0.02	-0.87	0.39	0.72	0.00	-0.63	-1.34	-0.39	-1.15
Enja83	1.26	0.34	0.40	-0.19	0.08	0.74	1.12	0.82	0.24	0.21	0.38	0.17
Ee507	**2.96**	-0.96	-0.71	-0.77	-0.63	1.39	2.22	1.92	-1.25	-0.74	-0.74	-1.05
Eja17	0.70	0.48	0.54	0.51	0.16	1.21	1.18	1.12	0.13	0.01	0.72	0.69
EJ2	-1.10	-1.00	-0.82	-0.61	-0.99	-1.32	-0.62	-0.77	-0.78	-1.40	-0.90	-1.16
Ee135	0.74	0.52	0.27	0.44	0.42	0.58	0.74	0.50	0.42	0.68	0.57	0.59
Ee508m	0.97	1.29	1.58	1.29	1.42	1.85	1.60	2.11	1.04	1.49	1.55	1.75
Ee2-165b	0.31	0.46	-0.05	0.31	0.38	0.03	0.08	0.50	0.76	0.29	1.22	1.05

Locus	**NAD vs** BAA	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG
Ee91b	0.29	0.17	0.04	0.24	0.52	0.32	0.33	0.40	-0.10	-0.05	0.72
Ee407	0.58	0.12	0.69	0.80	0.25	0.82	1.29	-0.02	1.43	-0.12	-0.25
EJ41,1	-0.39	-0.01	-0.55	-0.71	0.09	-0.38	-1.11	-0.66	-0.56	-0.37	0.06
Ee10	-0.30	-0.58	0.66	-0.42	0.65	1.05	0.31	0.54	-0.64	0.39	-0.37
EJ27,1	0.90	-0.27	0.20	0.10	2.01	**2.64**	2.15	0.33	-0.07	-0.67	-0.11
EJ35	-1.82	-1.08	-0.63	-1.48	-0.22	0.11	-0.61	-1.24	-1.95	-1.00	-1.76
Enja83	-0.92	-0.86	-1.46	-1.19	-0.52	-0.15	-0.44	-1.02	-1.05	-0.88	-1.09
Ee507	**-3.92**	**-3.67**	**-3.73**	**-3.59**	-1.57	-0.74	-1.04	**-4.21**	**-3.70**	**-3.70**	**-4.01**
Eja17	-0.22	-0.16	-0.19	-0.54	0.51	0.48	0.42	-0.57	-0.69	0.02	-0.01
EJ2	0.10	0.28	0.49	0.11	-0.22	0.48	0.33	0.32	-0.30	0.21	-0.06
Ee135	-0.22	-0.47	-0.30	-0.31	-0.16	0.01	-0.23	-0.32	-0.06	-0.16	-0.15
Ee508m	0.31	0.61	0.31	0.45	0.88	0.62	1.14	0.07	0.52	0.58	0.78
Ee2-165b	0.15	-0.36	-0.01	0.07	-0.28	-0.24	0.18	0.45	-0.03	0.91	0.74

	BAA vs										**BAB vs**			
Locus	BAB	KOT	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG	KOT	ANC	DUG	JAB
Ee91b	-0.12	-0.25	-0.05	0.23	0.03	0.04	0.11	-0.39	-0.34	0.43	-0.12	0.08	0.35	0.15
Ee407	-0.46	0.11	0.22	-0.33	0.25	0.71	-0.60	0.85	-0.70	-0.83	0.57	0.67	0.13	0.70
EJ41,1	0.38	-0.16	-0.33	0.48	0.01	-0.72	-0.27	-0.17	0.02	0.45	-0.54	-0.70	0.10	-0.37
Ee10	-0.28	0.96	-0.12	0.95	1.35	0.61	0.85	-0.34	0.69	-0.07	1.24	0.17	1.24	1.64
EJ27,1	-1.16	-0.70	-0.80	1.11	1.75	1.26	-0.57	-0.97	-1.56	-1.01	0.47	0.36	2.27	**2.91**
EJ35	0.74	1.20	0.34	1.60	1.94	1.21	0.58	-0.13	0.82	0.06	0.45	-0.40	0.86	1.19
Enja83	0.06	-0.54	-0.27	0.40	0.77	0.48	-0.10	-0.13	0.04	-0.18	-0.59	-0.32	0.34	0.72
Ee507	0.25	0.20	0.34	2.35	**3.18**	**2.88**	-0.29	0.22	0.22	-0.09	-0.06	0.08	2.10	**2.93**
Eja17	0.06	0.03	-0.32	0.73	0.70	0.64	-0.35	-0.47	0.24	0.21	-0.03	-0.38	0.67	0.64
EJ2	0.18	0.39	0.02	-0.32	0.38	0.24	0.22	-0.40	0.11	-0.16	0.21	-0.16	-0.50	0.20
Ee135	-0.25	-0.08	-0.09	0.06	0.23	-0.01	-0.10	0.16	0.06	0.07	0.17	0.15	0.31	0.47
Ee508m	0.29	0.00	0.14	0.57	0.31	0.82	-0.24	0.20	0.27	0.47	-0.29	-0.16	0.27	0.02
Ee2-165b	-0.51	-0.15	-0.08	-0.43	-0.38	0.03	0.30	-0.17	0.76	0.59	0.35	0.43	0.07	0.12

	KOT vs												
Locus	RIJ	PEA	PEB	SPE	CDG	ANC	DUG	JAB	RIJ	PEA	PEB	SPE	CDG
Ee91b	0.16	0.23	-0.27	-0.21	0.55	0.20	0.47	0.27	0.28	0.36	-0.14	-0.09	0.67
Ee407	1.16	-0.14	1.31	-0.24	-0.37	0.11	-0.44	0.13	0.59	-0.71	0.74	-0.81	-0.94
EJ41,1	-1.10	-0.65	-0.55	-0.36	0.07	-0.17	0.64	0.17	-0.56	-0.11	-0.02	0.18	0.61
Ee10	0.90	1.13	-0.06	0.97	0.21	-1.08	-0.01	0.39	-0.35	-0.12	-1.31	-0.27	-1.04
EJ27,1	2.42	0.59	0.20	-0.40	0.16	-0.10	1.81	2.45	1.96	0.13	-0.27	-0.87	-0.31
EJ35	0.47	-0.16	-0.87	0.08	-0.68	-0.86	0.41	0.74	0.02	-0.61	-1.33	-0.38	-1.14
Enja83	0.42	-0.15	-0.18	-0.02	-0.23	0.27	0.93	1.31	1.02	0.44	0.41	0.58	0.36
Ee507	**2.63**	-0.54	-0.03	-0.03	-0.34	0.14	2.16	**2.99**	**2.69**	-0.49	0.02	0.02	-0.29
Eja17	0.58	-0.41	-0.53	0.18	0.15	-0.35	0.70	0.67	0.61	-0.38	-0.50	0.21	0.18
EJ2	0.06	0.04	-0.58	-0.07	-0.34	-0.38	-0.71	-0.01	-0.16	-0.17	-0.79	-0.28	-0.55
Ee135	0.23	0.15	0.41	0.30	0.32	-0.01	0.14	0.30	0.06	-0.02	0.24	0.13	0.15
Ee508m	0.53	-0.53	-0.09	-0.02	0.17	0.14	0.57	0.31	0.82	-0.24	0.20	0.27	0.47
Ee2-165b	0.54	0.80	0.33	1.27	1.10	0.08	-0.28	-0.23	0.19	0.45	-0.02	0.92	0.75

	ANC vs							DUG vs					
Locus	DUG	JAB	RIJ	PEA	PEB	SPE	CDG	JAB	RIJ	PEA	PEB	SPE	CDG
Ee91b	0.27	0.07	0.08	0.16	-0.34	-0.29	0.47	-0.20	-0.19	-0.12	-0.62	-0.57	0.20
Ee407	-0.55	0.03	0.49	-0.82	0.64	-0.92	-1.04	0.57	1.03	-0.27	1.18	-0.37	-0.50
EJ41,1	0.81	0.34	-0.39	0.05	0.15	0.34	0.78	-0.47	-1.20	-0.75	-0.66	-0.46	-0.03
Ee10	1.07	1.47	0.73	0.96	-0.23	0.80	0.04	0.40	-0.34	-0.11	-1.30	-0.26	-1.03
EJ27,1	1.91	2.55	2.06	0.23	-0.17	-0.76	-0.21	0.64	0.15	-1.68	-2.08	**-2.67**	-2.12
EJ35	1.26	1.60	0.87	0.24	-0.47	0.48	-0.28	0.33	-0.39	-1.02	-1.73	-0.78	-1.54
Enja83	0.66	1.04	0.75	0.17	0.14	0.30	0.09	0.38	0.08	-0.50	-0.52	-0.36	-0.57
Ee507	2.02	**2.85**	2.55	-0.63	-0.12	-0.12	-0.43	0.83	0.53	**-2.64**	-2.13	-2.13	-2.44
Eja17	1.05	1.02	0.96	-0.03	-0.15	0.56	0.53	-0.03	-0.09	-1.08	-1.20	-0.49	-0.52
EJ2	-0.33	0.37	0.22	0.21	-0.41	0.09	-0.17	0.70	0.55	0.54	-0.08	0.43	0.16
Ee135	0.16	0.32	0.08	-0.01	0.25	0.15	0.16	0.16	-0.08	-0.16	0.10	-0.01	0.01
Ee508m	0.43	0.17	0.69	-0.38	0.07	0.13	0.33	-0.26	0.26	-0.81	-0.36	-0.30	-0.10
Ee2-165b	-0.36	-0.31	0.11	0.37	-0.10	0.84	0.67	0.05	0.47	0.73	0.26	1.19	1.03

	JAB vs					**RIJ vs**				**PEA vs**			**PEB vs**		**SPE vs**
Locus	RIJ	PEA	PEB	SPE	CDG	PEA	PEB	SPE	CDG	PEB	SPE	CDG	SPE	CDG	CDG
Ee91b	0.01	0.08	-0.42	-0.36	0.40	0.07	-0.43	-0.38	0.39	-0.50	-0.45	0.32	0.05	0.82	0.76
Ee407	0.46	-0.84	0.61	-0.94	-1.07	-1.30	0.15	-1.41	-1.53	1.45	-0.10	-0.23	-1.55	-1.68	-0.13
EJ41,1	-0.73	-0.28	-0.19	0.01	0.44	0.45	0.54	0.74	1.17	0.10	0.29	0.72	0.19	0.63	0.43
Ee10	-0.74	-0.51	-1.70	-0.67	-1.43	0.23	-0.96	0.08	-0.69	-1.19	-0.16	-0.92	1.03	0.27	-0.76
EJ27,1	-0.49	-2.32	**-2.71**	**-3.31**	**-2.75**	-1.83	-2.22	**-2.82**	-2.26	-0.40	-1.00	-0.44	-0.60	-0.04	0.56
EJ35	-0.72	-1.35	-2.07	-1.12	-1.87	-0.63	-1.34	-0.39	-1.15	-0.71	0.24	-0.52	0.95	0.19	-0.76
Enja83	-0.29	-0.87	-0.90	-0.74	-0.95	-0.58	-0.61	-0.44	-0.65	-0.03	0.14	-0.08	0.17	-0.05	-0.21
Ee507	-0.30	**-3.47**	**-2.96**	**-2.96**	**-3.27**	**-3.17**	**-2.66**	**-2.66**	**-2.97**	0.51	0.51	0.20	0.00	-0.31	-0.31
Eja17	-0.06	-1.05	-1.17	-0.46	-0.49	-0.99	-1.11	-0.40	-0.43	-0.12	0.59	0.56	0.71	0.68	-0.03
EJ2	-0.15	-0.16	-0.78	-0.28	-0.54	-0.01	-0.63	-0.13	-0.39	-0.62	-0.11	-0.38	0.50	0.24	-0.27
Ee135	-0.24	-0.32	-0.06	-0.17	-0.15	-0.08	0.18	0.07	0.09	0.26	0.15	0.17	-0.11	-0.09	0.02
Ee508m	0.51	-0.55	-0.11	-0.04	0.16	-1.07	-0.62	-0.56	-0.36	0.45	0.51	0.71	0.06	0.26	0.20
Ee2-165b	0.42	0.68	0.21	1.15	0.98	0.26	-0.21	0.73	0.56	-0.47	0.46	0.30	0.94	0.77	-0.17

Table S7 AMOVA on microsatellite DNA. Samples were divided into groups according to the bottom depth.

Microsatellite DNA AMOVA					
Groups	**Source of variation**	**df**	**Sum of squares**	**Variancecomponents**	**% of variation**
Bottom depth	Among groups	1	14.777	0.0131	0.35
	Between populations within groups	11	93.750	**0.0670**	1.79
	Within populations	931	3413.868	**3.6668**	97.86

df = freedom degree. Bold values are significant (P<0.05).

Supplementary materials CHAPTER III

(see CHAPTER III for References)

File S1

Methods to candidate outliers detection

Detection of non-neutral distribution of allele frequencies was evaluated using two separate approaches: the *fdist* method (Beaumont & Nichols 1996) implemented in Lositan (Antao *et al.* 2008) and a microsatellite designed LnRH method (Kauer *et al.* 2003). The *fdist* method detects outliers loci using joint distributions of F_{ST} and expected heterozygosity H_e, under an island migration model (Beaumont & Nichols 1996). The *fdist* tests were conducted on the basis of 50,000 simulations under both IAM (Infinite Allele Model) and SMM (Stepwise Mutation Model) settings. Both tests were conducted on the whole dataset and subsequently, separating data, in two distinct temporal frames: before the demographic collapse (between 1978 and 1987) and after this event (between 1989 and 2010). These tests were conducted to define whether selection occurred before or after the demographic collapse in 1987 and to verify their neutrality.

The LnRH method compares genetic diversity estimates found in each locus between pairwise sampling locations applying the following formula:

$$LnRH_{popA\text{-}popB} = Ln\,[(1/(1-He_{popA}))^2 - 1/(1/(1-He_{popB}))^2 - 1]$$

The LnRH method assumes that microsatellite loci under selective constraints will display lowered levels of diversity (Kauer *et al.* 2003). Under neutral expectations LnRH values show a normal distribution (Antao *et al.* 2008), therefore, after data standardization (mean = 0, s. d. = 1), the 99% of neutral loci are expected to have

LnRH values comprised within ±2.58. A Bonferroni correction [42] was applied to LnRH values (LnRH< -3.21 and LnRH> 3.21) with the aim to maintain a more conservative approach and avoid the identification of false positives.

Candidate outliers detection results

The *fdist* method under both IAM and SMM models revealed no outlier loci both when the whole dataset and the temporal frames before (between 1978 and 1987) and after the demographic collapse (between 1989 and 2010) were considered (Table S5). The results from LnRH method showed a lack of significance in all the 36 LnRH pairwise comparisons between samples, suggesting even in this case no candidate outliers among loci used (Supplementary Table S6).

Figures and Tables

Fig. S1 Graphical plot of STRUCTURE (Pritchard *et al.* 2000; Falush *et al.* 2003) simulations results. Blue circles are mean Ln probability from each simulates K (1-6). Bars crossing blue circles explain the extent of standard deviations (SD) among replicates from each K simulated.

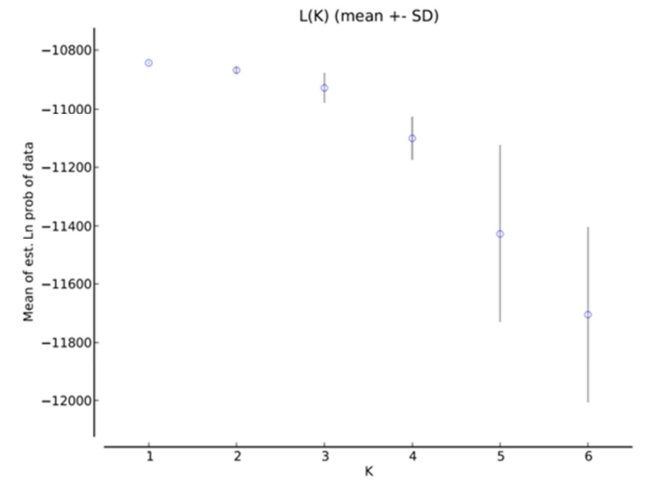


Table S1 Table illustrating the historical collection to which belongs each sample analyzed,the sample identification code (Sample), the sampling year (year) and months (Sampling months) and the number of analyzed individuals per sample (N° of total analysed individuals; the number includes individuals discarded from subsequent analysis due to contaminations or excess of missing data) and the number of genotyped individuals per sample (N° of genotyped individuals).

Historical collection	Sample	year	Sampling months	N° of total analysed individuals	N° of genotyped individuals
CHIOGGIA	CH78	1978	July-August	59	48
	CH87	1987	May-June-July	74	60
	CH94	1994	May-June-July-September	72	60
	CH00	2000	May-September	68	60
	CH10	2010	July-August	25	25
VIESTE	VI85	1985	May-June	69	60
	VI87	1987	May-June	72	60
	VI89	1989	May-June-July-September	70	60
	VI10	2010	July-August	48	48

Table S2 Information concerning the molecular markers employed in this study. Ta (°C) refers to annealing temperature expressed in Celsius degrees. * denotes newly designed primer sequences if compared to the original primer sequences described in the reference papers.

Locus name	Primers sequence (5'-3')	Repeated motif	Ta (°C)	Genome species
Ee2-91b	F: GGCTTGAGCTTGGCATAGG	$(CCGCA)_7$	60	*E. encrasicolus*
	R: CCGGAAGACACTCTGCACAC			
Ee2-135	F: AGGGCAGTGACAGGAGAGTC	$(ATTAG)_{10}$	60	*E. encrasicolus*
	R: TCGTTACCCTGCGTTTATACTG			
Ee2-508m*	F:CCTGACAATGTTTCAAAGTTGG	$(AGG)_8$	58	*E. encrasicolus*
	R:CACCTGATGCTGCTTGGTAG			
Ee2-407m*	F:CCGGAGTTGGTAGCATCTGT	$(CA)_{13}$	59	*E. encrasicolus*
	R:GTGTGGGTCTGTGGGTGTTT			
Ee2-165m*	F:CTGGAACCTTCCTCGTTTTG	$(CCT)_7$	58	*E. encrasicolus*
	R:GGGATCTTCAGGGAACCAGT			
Ee-10m*	F:TCTGATTCTTGCCTTTGGCTA	$[(GT)_9CT]_2[(GT)_2CT]_3$	57	*E. encrasicolus*
	R:ATGTTCTGGGGTGGCATAACT			
Eja-183m*	F:TTTGAATGGACACGATCATCA	$(TCA)_{13}$	54	*E. japonicus*
	R:TAAGGCCCCCTATCCAATGT			

Table S3 Table reporting Probability of Identity (P_{ID}) and Probability of Identity of Siblings ($P_{ID(sib)}$) estimations (Peakall & Smouse 2012). The estimates are produced both independently for every single locus and as complete set of loci analyzed from each sampling site/year.

P_{ID} by Locus

Pop	N	Ee91b	Ee165m	Ee135	Ee508m	Ee407m	Ee10m	Eja183	P_{ID}
CH78	48	0.091	0.252	0.037	0.222	0.025	0.126	0.046	2.820E-08
CH87	60	0.085	0.190	0.042	0.146	0.026	0.053	0.057	7.731E-09
CH94	60	0.105	0.273	0.047	0.274	0.097	0.146	0.055	2.854E-07
CH00	60	0.117	0.235	0.037	0.223	0.155	0.123	0.062	2.662E-07
CH10	25	0.084	0.154	0.032	0.168	0.104	0.035	0.085	2.130E-08
VI85	60	0.078	0.238	0.041	0.267	0.043	0.161	0.042	6.065E-08
VI87	60	0.085	0.203	0.036	0.276	0.010	0.142	0.052	1.329E-08
VI89	60	0.061	0.348	0.044	0.276	0.034	0.165	0.034	4.860E-08
VI10	48	0.073	0.253	0.036	0.186	0.043	0.053	0.054	1.535E-08

$P_{ID(sib)}$ by Locus

Pop	N	Ee91b	Ee165m	Ee135	Ee508m	Ee407m	Ee10m	Eja183	$P_{ID(sib)}$
CH78	48	0.391	0.519	0.331	0.526	0.322	0.431	0.344	1.682E-03
CH87	60	0.384	0.490	0.337	0.437	0.324	0.356	0.355	1.133E-03
CH94	60	0.402	0.542	0.344	0.558	0.412	0.449	0.354	2.743E-03
CH00	60	0.417	0.513	0.330	0.507	0.475	0.435	0.362	2.674E-03
CH10	25	0.382	0.448	0.324	0.475	0.417	0.337	0.384	1.419E-03
VI85	60	0.378	0.527	0.337	0.559	0.353	0.467	0.339	2.095E-03
VI87	60	0.385	0.484	0.331	0.558	0.291	0.459	0.351	1.609E-03
VI89	60	0.361	0.611	0.340	0.560	0.339	0.475	0.328	2.222E-03
VI10	48	0.372	0.543	0.329	0.488	0.344	0.361	0.354	1.425E-03

Table S4 Summary of genetic variability observed at 7 microsatellite loci from the sampled archived materials.

		CH78	CH87	CH94	CH00	CH10	VI85	VI87	VI89	VI10	TOT
Ee2-91b	N_A	7	7	6	8	6	8	8	9	9	11
	N	33	51	57	55	24	58	56	59	48	
	f_{null}	0.067	0.046	-0.034	-0.033	0.063	-0.006	0.043	-0.003	0.035	
	H_e	0.778	0.784	0.755	0.732	0.800	0.790	0.780	0.815	0.800	
	H_o	0.643	0.692	0.807	0.782	0.667	0.793	0.696	0.814	0.729	
	F_{IS}	0.176	0.118	-0.070	-0.069	0.171	-0.004	0.109	0.001	0.090	
	R_S	6.699	6.288	5.364	6.311	6.000	6.348	6.991	7.619	7.067	6.728
Ee2-165m	N_A	5	6	4	4	4	4	4	5	5	7
	N	48	58	60	59	25	59	60	60	46	
	f_{null}	-0.049	-0.079	-0.180	-0.118	0.072	-0.062	-0.081	-0.068	0.040	
	H_e	0.595	0.620	0.558	0.597	0.695	0.569	0.639	0.455	0.546	
	H_o	0.667	0.741	0.833	0.780	0.560	0.661	0.767	0.550	0.478	
	F_{IS}	-0.123	-0.198	**-0.500**	**-0.308**	0.197	-0.162	-0.202	-0.210	0.125	
	R_S	4.081	4.813	3.626	3.872	4.000	3.953	3.949	3.849	4.309	4.098
Ee2-135	N_A	9	9	13	11	10	11	11	11	11	14
	N	46	52	56	58	25	56	51	58	47	
	f_{null}	0.058	0.042	0.007	-0.012	0.090	0.033	0.007	0.017	0.040	
	H_e	0.869	0.856	0.842	0.865	0.890	0.854	0.865	0.848	0.870	
	H_o	0.750	0.769	0.821	0.879	0.700	0.786	0.843	0.810	0.786	
	F_{IS}	0.139	0.102	0.025	-0.016	0.218	0.081	0.026	0.045	0.098	
	R_S	8.229	8.177	9.515	8.682	9.791	9.064	9.163	9.138	8.851	9.273
Ee2-508m	N_A	8	6	7	9	7	8	7	6	9	12
	N	41	57	59	60	25	58	57	60	48	
	f_{null}	0.068	0.076	-0.014	0.020	-0.003	-0.013	-0.003	-0.011	**0.135**	
	H_e	0.569	0.709	0.526	0.603	0.647	0.519	0.527	0.522	0.628	
	H_o	0.455	0.568	0.542	0.567	0.640	0.534	0.526	0.533	0.400	
	F_{IS}	0.203	0.201	-0.031	0.061	0.012	-0.030	0.001	-0.023	0.367	
	R_S	6.338	5.096	5.108	5.577	6.497	5.431	4.463	4.781	7.273	5.562
Ee2-407m	N_A	22	21	11	14	11	24	21	25	17	38
	N	33	48	36	49	25	47	23	58	48	
	f_{null}	**0.059**	**0.125**	0.065	0.047	-0.048	**0.078**	**0.126**	0.053	0.012	
	H_e	0.883	0.877	0.734	0.634	0.733	0.826	0.945	0.847	0.843	
	H_o	0.759	0.632	0.611	0.551	0.800	0.674	0.682	0.741	0.813	

	F_{IS}	0.143	**0.282**	0.169	0.132	-0.093	0.185	**0.283**	0.125	0.037	
	R_S	16.722	15.105	8.979	9.025	9.432	14.676	18.995	15.063	11.161	13.784

		CH78	**CH87**	**CH94**	**CH00**	**CH10**	**VI85**	**VI87**	**VI89**	**VI10**	**TOT**
Ee2-10m	N_A	12	11	13	17	18	13	13	12	17	26
	N	46	57	60	60	25	60	59	60	48	
	f_{null}	0.055	**0.074**	-0.055	-0.065	0.024	0.039	0.026	-0.020	-0.004	
	H_e	0.709	0.824	0.680	0.696	0.861	0.652	0.659	0.639	0.813	
	H_o	0.609	0.681	0.767	0.800	0.800	0.583	0.610	0.667	0.813	
	F_{IS}	0.143	0.176	-0.129	-0.150	0.072	0.106	0.075	-0.044	0.001	
	R_S	8.043	8.999	7.679	8.945	14.939	7.131	8.388	7.096	11.818	9.786
Eja183m	N_A	15	10	11	13	8	11	13	16	16	21
	N	39	53	59	58	25	59	49	58	46	
	f_{null}	0.008	0.041	0.004	-0.030	-0.037	0.035	0.005	-0.001	0.056	
	H_e	0.847	0.827	0.827	0.815	0.791	0.851	0.834	0.868	0.828	
	H_o	0.821	0.745	0.814	0.862	0.840	0.780	0.816	0.862	0.717	
	F_{IS}	0.031	0.101	0.017	-0.058	-0.063	0.085	0.021	0.007	0.135	
	R_S	10.213	8.183	8.603	9.283	7.329	9.381	9.432	11.019	10.437	9.808
Average	N_{AM}	11.143	10.000	9.286	10.857	9.143	11.286	11.000	12.000	12.000	
	H_e	0.737	0.767	0.703	0.706	0.767	0.717	0.746	0.713	0.753	
	H_o	0.622	0.590	0.742	0.746	0.671	0.679	0.684	0.711	0.643	
	F_{IS}	**0.106**	**0.123**	-0.056	-0.057	0.078	0.050	0.060	0.003	**0.113**	
	R_S	8.618	8.094	6.982	7.385	8.284	7.998	8.769	8.366	8.702	8.434

N_A, number of alleles observed per location; N, number of individuals correctly genotyped; N_{AM}, mean number of alleles observed per location;f_{null}, null allele frequency calculated with (Brookfield 1996) formula; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, inbreeding coefficient estimates. R_S, allelic richness estimates standardized at 41 individuals. Bold F_{IS}values are significant (<0.05) after a sequential Bonferroni correction (Rice 1989). Bold f_{null}values are those exceed the 5% null allele tolerance threshold. Asterisk (*) denotes estimates still showed null allele signals after (Brookfield 1996) method.

Table S5 Results from the *fdist* (Beaumont & Nichols 1996) outlier loci detection tests.

				IAM			
Locus	$F_{ST\text{-}TOT}$	$F_{ST\text{-}CH\text{-}TOT}$	$F_{ST\text{-}VI\text{-}TOT}$	$F_{ST\text{-}CH78\text{-}87}$	$F_{ST\text{-}CH94\text{-}10}$	$F_{ST\text{-}VI85\text{-}87}$	$F_{ST\text{-}VI89\text{-}10}$
Ee2-91b	0.0055	0.0034	0.0059	0.0170	-0.0003	0.0104	0.0004
Ee2-165m	0.0273	0.0083	0.0618	0.0312	0.0023	0.0837	0.0023
Ee2-135	0.0029	0.0019	0.0033	-0.0013	0.0053	0.0011	0.0109
Ee2-508m	0.0162	0.0259	-0.0017	0.0733	0.0056	-0.0035	-0.0011
Ee2-407m	0.0197	0.0233	0.0093	-0.0016	0.0044	0.0232	0.0006
Ee-10m	0.0264	0.0218	0.0274	0.0189	0.0292	0.0332	0.0223
Eja183m	0.0004	-0.0021	0.0008	0.0071	-0.0064	-0.0064	0.0015

				SMM			
Locus	$F_{ST\text{-}TOT}$	$F_{ST\text{-}CH\text{-}TOT}$	$F_{ST\text{-}VI\text{-}TOT}$	$F_{ST\text{-}CH78\text{-}87}$	$F_{ST\text{-}CH94\text{-}10}$	$F_{ST\text{-}VI85\text{-}87}$	$F_{ST\text{-}VI89\text{-}10}$
Ee2-91b	0.0055	0.0034	0.0059	0.017	-0.0003	0.0104	0.0004
Ee2-165m	0.0273	0.0083	0.0618	0.0312	0.003	0.0837	0.0023
Ee2-135	0.0029	0.0019	0.0033	-0.0013	0.0053	0.0011	0.0109
Ee2-508m	0.0162	0.0259	-0.0017	0.0733	0.0056	-0.0035	-0.0011
Ee2-407m	0.0197	0.0233	0.0093	-0.0016	0.0044	0.0233	0.0006
Ee-10m	0.0264	0.0218	0.0274	0.0189	0.0293	0.0332	0.0223
Eja183m	0.0004	-0.0021	0.0008	0.0071	-0.0064	-0.0064	0.0015

IAM = Infinite Allele Mutation model; SMM = Stepwise Mutation Model. $F_{ST\text{-}TOT}$= mean F_{ST} value among all samples analyzed; $F_{ST\text{-}CH\text{-}TOT}$= mean F_{ST} value among all Chioggia samples analyzed; $F_{ST\text{-}VI\text{-}TOT}$ = mean F_{ST} value among all Vieste samples analyzed; $F_{ST\text{-}CH78\text{-}87}$ =mean F_{ST} value among all Chioggia samples before 1987; $F_{ST\text{-}CH94\text{-}10}$=mean F_{ST} value among all Chioggia samples after 1987; $F_{ST\text{-}VI85\text{-}87}$ =mean F_{ST} value among all Vieste samples before 1987; $F_{ST\text{-}VI89\text{-}10}$=mean F_{ST} value among all Vieste samples after 1987. Significant values for selection were denoted by: * < 0.05; ** < 0.01; *** < 0.001.

Table S6 Results from the LnRH method (Kauer *et al.* 2003) to detect outlier loci. Significant values (LnRH< -3.21 and LnRH> 3.21) were highlighted in bold.

Locus	CH78-CH87	CH78-CH94	CH78-CH00	CH78-CH10	CH78-VI85	CH78-VI87	CH78-VI89	CH78-VI10
Ee2-91b	-0.0027	0.0114	0.0242	-0.0098	-0.0054	-0.0010	-0.0156	-0.0098
Ee2-165b	-0.0236	0.0379	-0.0028	-0.0818	0.0254	-0.0401	0.1721	0.0516
Ee2-135m	0.0037	0.0079	0.0010	-0.0051	0.0041	0.0010	0.0059	-0.0003
Ee2-508m	-0.1170	0.0485	-0.0343	-0.0729	0.0575	0.0476	0.0542	-0.0570
Ee2-407m	0.0016	0.0599	0.1302	0.0600	0.0171	-0.0107	0.0101	0.0112
Ee2-10m	-0.0570	0.0200	0.0084	-0.0689	0.0407	0.0352	0.0516	-0.0528
Eja-183m	0.0065	0.0065	0.0111	0.0209	-0.0014	0.0043	-0.0062	0.0061
Locus	**CH87-CH94**	**CH87-CH00**	**CH87-CH10**	**CH87-VI85**	**CH87-VI87**	**CH87-VI89**	**CH87-VI10**	**CH94-CH00**
Ee2-91b	0.0141	0.0269	-0.0070	-0.0026	0.0018	-0.0129	-0.0071	0.0128
Ee2-165b	0.0615	0.0208	-0.0582	0.0490	-0.0164	0.1957	0.0752	-0.0407
Ee2-135m	0.0042	-0.0027	-0.0088	0.0004	-0.0027	0.0022	-0.0040	-0.0069
Ee2-508m	0.1656	0.0827	0.0441	0.1746	0.1646	0.1712	0.0600	-0.0829
Ee2-407m	0.0583	0.1286	0.0586	0.0156	-0.0123	0.0085	0.0096	0.0703
Ee2-10m	0.0770	0.0654	-0.0119	0.0977	0.0922	0.1086	0.0042	-0.0116
Eja-183m	-0.0001	0.0046	0.0144	-0.0080	-0.0022	-0.0128	-0.0004	0.0047
Locus	**CH94-CH10**	**CH94-VI85**	**CH94-VI87**	**CH94-VI89**	**CH94-VI10**	**CH00-CH10**	**CH00-VI85**	**CH00-VI87**
Ee2-91b	-0.0212	-0.0168	-0.0124	-0.0270	-0.0212	-0.0339	-0.0296	-0.0251
Ee2-165b	-0.1197	-0.0125	-0.0780	0.1342	0.0137	-0.0790	0.0282	-0.0373
Ee2-135m	-0.0130	-0.0038	-0.0069	-0.0020	-0.0082	-0.0061	0.0031	0.0000
Ee2-508m	-0.1215	0.0090	-0.0010	0.0056	-0.1056	-0.0386	0.0918	0.0819
Ee2-407m	0.0003	-0.0428	-0.0706	-0.0498	-0.0487	-0.0700	-0.1130	-0.1409
Ee2-10m	-0.0889	0.0207	0.0152	0.0316	-0.0728	-0.0773	0.0323	0.0268
Eja-183m	0.0144	-0.0079	-0.0021	-0.0127	-0.0004	0.0098	-0.0125	-0.0068
Locus	**CH00-VI89**	**CH00-VI10**	**CH10-VI85**	**CH10-VI87**	**CH10-VI89**	**CH10-VI10**	**VI85-VI87**	**VI85-VI89**
Ee2-91b	-0.0398	-0.0340	0.0044	0.0088	-0.0059	-0.0001	0.0044	-0.0103
Ee2-165b	0.1749	0.0544	0.1072	0.0417	0.2539	0.1334	-0.0654	0.1467
Ee2-135m	0.0049	-0.0013	0.0092	0.0061	0.0110	0.0048	-0.0031	0.0018
Ee2-508m	0.0885	-0.0227	0.1305	0.1205	0.1271	0.0159	-0.0009	-0.0033
Ee2-407m	-0.1201	-0.1190	-0.0431	-0.0709	-0.0501	-0.0490	-0.0278	-0.0070
Ee2-10m	0.0432	-0.0612	0.1096	0.1041	0.1205	0.0161	-0.0055	0.0109
Eja-183m	-0.0174	-0.0050	-0.0223	-0.0166	-0.0271	-0.0148	0.0057	-0.0048
Locus	**VI85-VI10**	**VI87-VI89**	**VI87-VI10**	**VI89-VI10**				
Ee2-91b	-0.0045	-0.0147	-0.0089	0.0058				
Ee2-165b	0.0262	0.2121	0.0917	-0.1205				
Ee2-135m	-0.0044	0.0049	-0.0013	-0.0062				
Ee2-508m	-0.1145	0.0066	-0.1046	-0.1112				
Ee2-407m	-0.0060	0.0208	0.0219	0.0011				
Ee2-10m	-0.0935	0.0164	-0.0880	-0.1044				
Eja-183m	0.0075	-0.0106	0.0018	0.0123				

*Splendiani A., ***Fioravanti T.**, Giovannotti M., Negri A., Ruggeri P., Olivieri L., Nisi Cerioni P., Lorenzoni M., Caputo Barucchi V. (2016) The Effects of Paleoclimatic Events on Mediterranean Trout: Preliminary Evidences from Ancient DNA. *PLoS ONE*, **11** (6), e0157975. doi:10.1371/journal.pone.0157975

** These authors contributed equally to this work*

Supplementary materials CHAPTER IV

Figures and Tables

Table S1 Details regarding the mtDNA D-loop sequences using to obtain the statistical parsimony network.

Haplotype	GenBank Accession no.	Reference
ADcs1	AY836330	Cortey *et al.* 2004
ADcs2	AY836331	Cortey *et al.* 2004
ADcs3	AY836332	Cortey *et al.* 2004
ADcs4	AY836333	Cortey *et al.* 2004
ADcs5	AY836334	Cortey *et al.* 2004
ADcs6	AY836335	Cortey *et al.* 2004
ADcs7	AY836336	Cortey *et al.* 2004
ADcs8	AY836337	Cortey *et al.* 2004
ADcs9	AY836338	Cortey *et al.* 2004
ADcs10	AY836339	Cortey *et al.* 2004
ADcs11	AY836340	Cortey *et al.* 2004
ADcs12	AY836341	Cortey *et al.* 2004
ADcs13	AY836342	Cortey *et al.* 2004
ADcs14	AY836343	Cortey *et al.* 2004
ADcs15	AY836344	Cortey *et al.* 2004
ADcs16	AY836345	Cortey *et al.* 2004
ADcs17	AY836346	Cortey *et al.* 2004
ADcs18	AY836347	Cortey *et al.* 2004
ADcs19	AY836348	Cortey *et al.* 2004
ADcs20	AY836349	Cortey *et al.* 2004
AD-Z1	DQ381565	Susnik *et al.* 2007
Haplo12	AY926570	Susnik *et al.* 2006
Ma2a	DQ841189	Meraner *et al.* 2007
Ma2b	DQ841190	Meraner *et al.* 2007
MAcs1	AY836365	Cortey *et al.* 2004
MEcs1	AY836350	Cortey *et al.* 2004
MEcs2	AY836351	Cortey *et al.* 2004
MEcs3	AY836352	Cortey *et al.* 2004
MEcs4	AY836353	Cortey *et al.* 2004
MEcs5	AY836354	Cortey *et al.* 2004
MEcs6	AY836355	Cortey *et al.* 2004
MEcs7	AY836356	Cortey *et al.* 2004
MEcs8	AY836357	Cortey *et al.* 2004
MEcs9	AY836358	Cortey *et al.* 2004
MEcs10	AY836359	Cortey *et al.* 2004
MEcs11	AY836360	Cortey *et al.* 2004
MEcs12	AY836361	Cortey *et al.* 2004
MEcs13	AY836362	Cortey *et al.* 2004
MEcs14	AY836363	Cortey *et al.* 2004

MEcs15	AY836364	Cortey *et al.* 2004
ATcs1	AF273086	Cortey & Garcia-Marin 2000
ATcs2	AF273087	Cortey & Garcia-Marin 2000
ATcs3	AF274574	Cortey & Garcia-Marin 2000
ATcs4	AF274575	Cortey & Garcia-Marin 2000
ATcs5	AF274576	Cortey & Garcia-Marin 2000
ATcs6	AF274577	Cortey & Garcia-Marin 2000
ATcs11	AY836327	Cortey *et al.* 2004
ATcs12	AY836328	Cortey *et al.* 2004
ATcs13	AY836329	Cortey *et al.* 2004
ATcs14	EF530476	Cortey *et al.* 2009
ATcs15	EF530477	Cortey *et al.* 2009
ATcs16	EF530478	Cortey *et al.* 2009
ATcs17	EF530479	Cortey *et al.* 2009
ATcs18	EF530480	Cortey *et al.* 2009
ATcs19	EF530481	Cortey *et al.* 2009
ATcs20	EF530482	Cortey *et al.* 2009
ATcs21	EF530483	Cortey *et al.* 2009
ATcs22	EF530484	Cortey *et al.* 2009
ATcs23	EF530485	Cortey *et al.* 2009
ATcs24	EF530486	Cortey *et al.* 2009
ATcs25	EF530487	Cortey *et al.* 2009
ATcs26	EF530488	Cortey *et al.* 2009
ATcs27	EF530489	Cortey *et al.* 2009
ATcs28	EF530490	Cortey *et al.* 2009
ATcs29	EF530491	Cortey *et al.* 2009
ATcs30	EF530492	Cortey *et al.* 2009
ATcs31	EF530493	Cortey *et al.* 2009
ATcs32	EF530494	Cortey *et al.* 2009
ATcs33	EF530495	Cortey *et al.* 2009
ATcs34	EF530496	Cortey *et al.*, 2009
ATcs35	EF530497	Cortey *et al.* 2009
ATcs36	EF530498	Cortey *et al.* 2009
ATcs37	EF530499	Cortey *et al.* 2009
ATcs38	EF530500	Cortey *et al.* 2009
ATcs39	EF530501	Cortey *et al.* 2009
ATcs41	EF530502	Cortey *et al.* 2009
ATcs42	EF530503	Cortey *et al.* 2009
ATcs43	EF530504	Cortey *et al.* 2009
ATcs45	EF530505	Cortey *et al.* 2009
ATcs46	EF530506	Cortey *et al.* 2009
ATcs47	EF530507	Cortey *et al.* 2009
ATcs48	EF530508	Cortey *et al.* 2009
ATcs49	EF530509	Cortey *et al.* 2009
ATcs50	EF530510	Cortey *et al.* 2009
ATcs51	EF530511	Cortey *et al.* 2009
ATcs52	EF530512	Cortey *et al.* 2009
AT11a	AY185578	Duftner *et al.* 2003

AT11b	AY185579	Duftner *et al.* 2003
At1e	DQ841192	Meraner *et al.* 2007
ATM1	JF297978	Snoj *et al.* 2011
ATM2	JF297979	Snoj *et al.* 2011
ATM3	JF297980	Snoj *et al.* 2011
ATM4	JF297975	Snoj *et al.* 2011
ATM5	JF297977	Snoj *et al.* 2011
ATM6	JF297976	Snoj *et al.* 2011
ATM7	JF297982	Snoj *et al.* 2011
Dades	JF297981	Snoj *et al.* 2011
Da1a	AY185568	Duftner *et al.* 2003
Da3	AY185571	Duftner *et al.* 2003
Da9	AY185572	Duftner *et al.* 2003
Da22	AY185573	Duftner *et al.* 2003
Da24	AY185576	Duftner *et al.* 2003
DUcs1	EF530513	Snoj *et al.* 2011
DUcs2	EF530514	Snoj *et al.* 2011
DUcs3	EF530515	Snoj *et al.* 2011
DUcs4	EF530516	Snoj *et al.* 2011
DUcs5	EF530517	Snoj *et al.* 2011
DUcs6	EF530518	Snoj *et al.* 2011
DUcs7	EF530519	Snoj *et al.* 2011
DUcs8	EF530520	Snoj *et al.* 2011
DUcs9	EF530521	Snoj *et al.* 2011
DUcs10	EF530522	Snoj *et al.* 2011
DUcs11	EF530523	Snoj *et al.* 2011
DUcs12	EF530524	Snoj *et al.* 2011
DUcs13	EF530525	Snoj *et al.* 2011
DUcs14	EF530526	Snoj *et al.* 2011
DUcs15	EF530527	Snoj *et al.* 2011
DUcs16	EF530528	Snoj *et al.* 2011
DUcs17	EF530529	Snoj *et al.* 2011
DUcs18	EF530530	Snoj *et al.* 2011
DUcs19	EF530531	Snoj *et al.* 2011
DUcs20	EF530532	Snoj *et al.* 2011
DUcs21	EF530533	Snoj *et al.* 2011
DUcs22	EF530534	Snoj *et al.* 2011
DUcs23	EF530535	Snoj *et al.* 2011

Table S2 Amplification success.

	Primer pairs							
Sample ID	**StCR-1**	**StCR-2a**	**StCR-2c***	**StCR-2b**	**StCR-3**	**StCR-4**	**StCR-5**	**StCR-7**
First set of samples analysed								
V44	NA	NA	NA	.	NA	.	.	NA
V46	.	.		.	.	.	.	.
M50	.	NA	.	.	.	.	.	.
PT57	.	NA	NA	.	.	NA	.	.
M62	.	.		.	.	.	.	.
TP65	.	NA	.	.	NA	NA	.	.
Second set of samples analysed								
M47	NA	NA	.	.	.	.	.	.
M48	.	.		.	.	.	.	.
Pr50	.	.		.	.	.	.	.
M57	.	.		.	.	.	.	.
V62	.	.		.	.	.	.	.
Pr65	NA	NA	NA	NA	NA	NA	NA	NA

Sequences obtained for each primer pair used. The dots indicate all fragments successfully amplified and sequenced. The acronym "NA" indicates not amplified.
*The StCR-2c primer pair was used only in case of non-amplification with StCR-2a.

LIST OF PAPERS PUBLISHED

Ruggeri P., Splendiani A., Giovannotti M., **Fioravanti T.**, Occhipinti G., Nisi Cerioni P., Caputo Barucchi V. (2015) The role of life-history traits, selective pressure and hydrographic boundaries in shaping the genetic structure of the transparent goby, *Aphia minuta*. *Marine Ecology*, **37**, 518–531. doi:10.1111/maec.12266

Ruggeri P., Splendiani A., Occhipinti G., **Fioravanti T.**, Santojanni A., Leonori I., De Felice A., Arneri E., Procaccini G., Catanese G., Tičina V., Bonanno A., Cerioni Nisi P., Giovannotti M., Grant W.S., Caputo Barucchi V. (2016) Biocomplexity in Populations of European Anchovy in the Adriatic Sea. *PLoS ONE*, **11** (4), e0153061. doi:10.1371/journal.pone.0153061

Ruggeri P., Splendiani A., Di Muri C., **Fioravanti T.**, Santojanni A., Leonori I., De Felice A., Biagiotti I., Carpi P., Arneri E., Nisi Cerioni P., Giovannotti M., Caputo Barucchi V. (2016) Coupling Demographic and Genetic Variability from Archived Collections of European Anchovy (*Engraulis encrasicolus*). *PLoS ONE*, **11** (3), e0151507. doi:10.1371/journal.pone.0151507

*Splendiani A., ***Fioravanti T.**, Giovannotti M., Negri A., Ruggeri P., Olivieri L., Nisi Cerioni P., Lorenzoni M., Caputo Barucchi V. (2016) The Effects of Paleoclimatic Events on Mediterranean Trout: Preliminary Evidences from Ancient DNA. *PLoS ONE*, **11** (6), e0157975. doi:10.1371/journal.pone.0157975

** These authors contributed equally to this work*

Other papers at which I have collaborated during my PhD:

Splendiani A., Ruggeri P., Giovannotti M., Pesaresi S., Occhipinti G., **Fioravanti T.**, Lorenzoni M., NisiCerioni P., Caputo Barucchi V. (2016) Alien brown trout invasion of the Italian peninsula: the role of geological, climate and anthropogenic factors. *Biological Invasions*, **18**, 2029-2044. doi:10.1007/s10530-016-1149-7

Splendiani A., **Fioravanti T.**, Giovannotti M., Olivieri L., Ruggeri P., Nisi Cerioni P., Vanni S., Enrichetti F., Caputo Barucchi V. (submitted) Museum samples could help to reconstruct the original distribution of genus Salmo in Italy. *Journal of Fish Biology*.

Splendiani A., **Fioravanti T.**, Giovannotti M., D'Amore A., Furii G., Nisi Cerioni P., Lucchetti A., Caputo Barucchi V. (in preparation) Mitochondrial DNA reveals the natal origin of loggerhead turtles (*Caretta caretta*) stranded or by-caught along the southwestern Adriatic coasts.