Challenges in marine ecology: genomic investigations of dispersal patterns and phenotypic plasticity in Mediterranean Fishes

Desafíos en ecología marina: investigaciones genómicas de los patrones de dispersión y de la plasticidad fenotípica en peces del Mediterráneo

Celia Schunter

TESIS DOCTORAL
UNIVERSITAT DE BARCELONA

FACULTAT DE BIOLOGIA
DEPARTAMENT DE GENETICA
PROGRAMA DE DOCTORAT EN GENETICA

Challenges in marine ecology: genomic investigations of dispersal patterns and phenotypic plasticity in Mediterranean Fishes

Desafíos en ecología marina: investigaciones genómicas de los patrones de dispersión y de la plasticidad fenotípica en peces del Mediterráneo

Memoria presentada por

Celia Schunter

para acceder al grado de Doctor por la Universitat de Barcelona

Celia Schunter

Barcelona September 2013

PhD advisor
Dra. Marta Pascual Berniola
Universitat de Barcelona

PhD advisor
Dr. Enrique Macpherson Mayol
CEAB-CSIC

Acknowledgements

Ever since I was a little girl I have wanted to be a marine biologist and a researcher. Today, by handing in my doctoral thesis, I am moving a very important step towards this dream. However, this big step is made up of uncountable little steps that lead me to this achievement which would not have been possible without the help and support of so many people. I feel extremely fortunate to have only encountered encouragement and inspiration on my journey. Here, briefly, I will try to put my immense gratitude into words.

First and foremost, I would like to thank my two PhD advisors Marta and 'Mac'. They listened, advised, helped. Everything I believe a great teacher and mentor does. In these past 6 years that we have been working together in my Master's thesis as well as this PhD thesis, I have learnt a great deal and a broad range of skills from them both. What I am most grateful for though is the fact that they gave me liberty to explore and develop and I always felt respected and encouraged to try whatever crazy idea I might have come up with. Our discussions and arguments have always been, even though endless, inspiring and fruitful. So there is not much more I can say than a heartfelt thank you for being such great advisors as well as such genuinely great people.

Throughout my thesis, I have spent a wonderful time in various labs. In Catalunya I am so fortunate to already have two 'home' labs, one at the Centre d'Estudis Avancats de Blanes and the other one at the University of Barcelona. Cinta, Gemma, Victor, Victor Hugo, Ferran, Rosanna, Aymee and Pedro...you have not only made working so much more fun, but have become amazing and important friends and my Barcelona family to me. I also want to thank the CEAB-gang, as although I showed up sporadically I was always welcomed back with open arms and a smile. Throughout my stays abroad in the United States I have met some incredible people who helped me get through the hard parts and celebrated the good ones with me. I am very grateful to Steve Vollmer and Carlos Garza to invite me into their labs and help me in my fight with highthroughput sequencing data. Steve, Silvia, Liz, Kylla, Annick and the rest of the crazy people from the MSC, it has been a great pleasure to spend my 'island' life there with you. In Santa Cruz I worked at a great lab with even greater people; Devon, Alicia, Eric, Eric, Martha…and all the non-lab friends…I loved being 'Santa Cruzian' because of you all.

Thank you to all my supporting, loving and fun friends in Barcelona and all over the world, who in one way or the other contributed to me finishing this thesis and making me a happy person. A special thanks to Annelie, Laura and Irene who were there for me to lift me up when I needed them most and were and are truly great friends to me.

Last but definitely not least, I am so very grateful to my family. Marc, although 99% of the time on the other side of the world, I am so glad you are my big brother. As to my parents, when I thought I was ready to fly out of the nest, I probably wasn't, but they still let me go to explore everything there is. It would have been impossible for me to fly without knowing that they were always, even though far way, right by my side.

And this is why I dedicate this doctoral thesis to my parents; none of this would have been possible without your immense support. THANK YOU.

The sea, once it casts its spell, holds one in its net of wonder forever

Jacques Yves Cousteau

Für meine Eltern

1. INTRODUCTION

INTRODUCTION

The Mediterranean Sea

Approximately 70% of planet earth is covered by water. Less than 1% of more than 350 million square kilometers of ocean surface corresponds to the Mediterranean Sea. However, over 17.000 species of marine fishes, cetaceans, crabs, sea turtles, mollusks and plants are living within its waters and the Mediterranean basin displays one of the world richest animal and plant diversity (Coll *et al.* 2010). For this reason the Mediterranean Sea has been defined as a biodiversity hotspot for conservation priorities (Myers *et al.* 2000; Macpherson 2002).

The Mediterranean is a semi-enclosed sea confined by the different continental land masses with highly populated coastlines. For millennia has the human population inhabiting these coasts exploited and altered its marine resources and habitats. Therefore, anthropogenic impacts and alterations are proportionally larger in the Mediterranean Sea than in any other sea (Coll *et al.* 2010). Due to the risks to biodiversity the Mediterranean basin could be named one of the 'hottest' of all biodiversity hotspots. The threats to this rich environment are multiple, but at present the most impacting are: habitat loss, fishing impacts, pollution, euthrophication and invasive species. Infrastructural development, climate change and agricultural farming have indirect effects on the degradation of the Mediterranean Sea. Destruction of the marine habitat and overexploitation, both caused by recreational and commercial fishing, are some of the direct impacts on its marine life.

It is becoming common consensus that marine life in the Mediterranean Sea and other oceans around the world are increasingly altered and under considerable threat by human activity (Hutchings 2000). This strain on the marine ecosystem worldwide has led for calls of new management approaches especially for the coastal areas (Botsford 1997). Such measures, for instance, are the regulation of fisheries towards more sustainable fishing and the establishment of a network of Marine Protected Areas (MPA). It has been shown, that the establishment of marine reserves that in practice are too small and scattered can have a reduced or null effect on the protection of the Mediterranean ecosystems (Botsford *et al.* 2009). To avoid such waste of resources and establish efficient ecosystem protection approaches, MPAs have to be designed properly based on multi-disciplinary knowledge.

The design of MPAs often focus on simply protecting a target fraction of key habitats, and assume the ecosystems naturally occurring in those habitats will persist in the MPAs and contribute to unprotected habitats outside of the MPAs. However, it has been demonstrated that a network of Marine Protected Areas provides more protection than a set of individual unconnected protected areas (Halpern & Warner 2003). Furthermore, modeling studies have indicated the great importance of spatial configurations of MPAs to promote population persistence (Kaplan *et al.* 2009). Although these interactions are clear in modeling results, efforts to assess and design MPAs are hindered by the lack of knowledge about several important factors. One of the mayor issues is the uncertainty about propagule dispersal, which is one of the essential processes connecting areas and populations with each other (Almany *et al.* 2009). This dispersal is potentially influenced by currents or other oceanographic processes which are not being considered. Moreover, although one of the aims of MPAs is the protection of genetic diversity, such measures are often neglected especially at the inter-population level.

For the establishment of ecologically coherent MPA networks a broad array of scientific assessments, together with clear attainable and measurable objectives, are necessary. This objective often falls short due to the complexity of the marine environment and the challenges in assessing marine populations. Integrated and multidisciplinary studies combining oceanographic modeling, larval ecology and population genetics also including genetic diversity between and within populations are essential to face these challenges.

Population genetics

The study of the genetic composition of biological populations is called population genetics. The basic concept of population genetics is the study of genetic variation within and between populations (Hartl & Clark 2007). For a population of individuals to succeed over evolutionary time, it must contain genetic variability. Therefore, to get an overview of the population we analyze the variability of different phenotypes and genotypes. The fundamental tool of this approach is the analysis and quantification of allele frequencies. The quantification of the number of alleles and their commonness in a population can provide information on the genetic diversity of a species population as well as the richness of its gene pool.

Furthermore, allele frequency estimates taken from different locations can show the degree of genetic differentiation. If allele frequencies are not similar between two locations, we can consider these two locations two genetically distinct units suggesting that the exchange of genetic material is low (Hudson *et al.* 1992). This dissimilarity is measured as the genetic distance and is established commonly by using Wright's F-statistics (Wright 1969).

The distribution and change in genetic variation is under the influence of various evolutionary processes such as genetic drift, mutation, inbreeding, gene flow and also selection (Hedrick 2009). For marine species we would like to know how many individuals disperse from one population to another to understand how connected or how isolated certain populations are to evaluate the possible impact of a network of MPAs. However, this is not possible for thousands of minuscule larvae and therefore we can use the flow of genes as a proxy for the exchange of individuals and their genetic material.

Whereas the measurement of population structure is mainly based on the genetic distance and differences between populations, it is also possible to use allele frequencies to estimate similarities between individuals. The coefficient of relatedness is used in population genetics to estimate the amount of inbreeding within a population. However, the measurements of relationship can also be used to detect different degrees of kinship and family structure within populations and possibly exchange of offspring and genetic material between populations.

Molecular markers

Various types of genetic markers are used in population genetics such as mitochondrial genes, microsatellite loci or single nucleotide polymorphisms (SNPs). Some of the advantages of working with mitochondrial DNA is that mitochondrial genes are well-known and relatively conserved (Wan *et al.* 2004), allowing for an easier development of primers to be used in different species. However, the genetic variation in these genes may retain historical processes whether demographic or selective that may hinder current population structure assessment (Hurst & Jiggins 2005). Microsatellite loci, which have been the marker of choice in population genetics in the last decade, are nuclear loci with single sequences repeats. Most microsatellite loci are on non-coding regions, allowing mutation to occur on a more rapid pace and therefore being neutral they can represent somewhat more present day demographic processes. One great

advantage of microsatellite markers is that they are often highly polymorphic, meaning that there are many different alleles, which gives a marker an elevated power for analysis (Selkoe & Toonen 2006). Morin and coauthors (2004) comment that with the usage of such neutral markers the evaluation of genes that affect fitness are neglected. SNP markers, on the other hand, have the advantage of being both, neutral or allow for the discovery of genes under selection. Recent studies reveal that, when comparing neutral and non-neutral SNP markers, non-neutral markers reflected genetic differentiation in a much more refined manner than neutral markers (De Wit & Palumbi 2012; Lamichhaney *et al.* 2012). The application of single nucleotide polymorphisms (SNPs) as genetic markers has been common practice for more than a decade in model species, such as humans and mice. The discovery of SNP markers was mainly based on two methods, either by Sanger sequencing unigene-derived amplicons or by mining through existing EST databases and PCR-based validation (Picoult-Newberg *et al.* 1999). However, the development by these methods was time-consuming and expensive and therefore restricted to small-scales and model-species (Vignal *et al.* 2002). With the advances in high-throughput sequencing technologies in the past years, it seems possible and more feasible to discover SNP markers also for non-model species. This would more easily allow large-scale ecological studies on non-model species, as SNP genotyping can be high-throughput by analyzing hundreds or thousands of SNPs simultaneously. Hence, this type of marker could become the new marker of choice for a variety of studies in population genetics.

Population structure and Connectivity

Dispersal is a critical ecological process which can determine genetic population structure as well as the diversity within and among populations across evolutionary and ecological time scales (Waples & Gaggiotti 2006). On evolutionary time scales, by mitigating the effects of genetic drift and reducing the mutation load on small populations, dispersal can reduce the risk of extinction. Despite its importance, understanding dispersal patterns still remains a major challenge especially in marine ecology.

The extent to which populations are linked by the movement of individuals among populations is termed connectivity (Palumbi 2003; Cowen & Sponaugle 2009a). The movement from one population to another can be accomplished through dispersal by offspring or by juvenile or adult migration. However, in the marine environment it has been demonstrated that many species,

especially non-pelagic fish species, display homing behavior as adults and do not migrate. Therefore it is mostly the pelagic larval stage that allows for the dispersal and movement between populations and maintains connectivity between populations (Planes 2002).

Formerly, it was generally assumed that marine populations present high genetic fluxes, favored by the inexistence of physical barriers, so that connectivity would be mainly determined by the dispersive capacity of each species. Nevertheless, more recently it has been showed that a number of species present a spatial differentiation which is higher than expected if we only considered their dispersive abilities (Calderón *et al.* 2007; Palero *et al.* 2008). Connectivity patterns have been shown to be influenced by other factors such as oceanographic features (White *et al.* 2010), larval behavior (Gerlach *et al.* 2007) and the distribution of habitat (Pinsky *et al.* 2012). In the north-western Mediterranean Sea it has recently been demonstrated that oceanographic processes, such as current patterns, oceanographic discontinuities and fronts are crucial factors influencing population connectivity in the marine environment (Galarza *et al.* 2009). Hence, to understand connectivity patterns on a large-scale it is important to consider the physical processes possibly influencing the gene flow between populations.

The approach of inferring population connectivity by the genetic differentiation of locations is effective for the identification of major biogeographical or oceanographic barriers (Planes 2002) and allows for the measurement of gene flow across evolutionary timescales (Waples & Gaggiotti 2006). Therefore, large scale dispersal and exchange between populations in the marine environment is mostly measured by indirect approaches such as population genetics or larval behavior ecology or biophysical models (Cowen *et al.* 2007). This means that by measuring connectivity through population differentiation we are measuring processes from the past and present and the detected connectivity could be compromised by present-day rapid changes, such as anthropogenic disturbances (Cowen & Sponaugle 2009b). Hence, it is also essential to understand how and if connectivity is maintained in the present time. This can be achieved by direct measurement of larval dispersal such as the quantity of dispersing individuals and the direction of dispersal. However, obtaining direct estimates of larval movement is difficult due to the technical difficulties in tracking small larvae in the plankton. Therefore larvae trajectories remain largely unknown and understanding present-day dispersal patterns is still a great challenge. Parentage analysis permits for the direct estimation of connectivity, as the

detection of parent offspring pairs allows in many cases to reconstruct the movement of the offspring providing direct evidence of dispersal (Hedgecock *et al.* 2007). However, the successful application of paternity analysis requires sampling a large proportion of the parental population for the successful encounter of parent offspring pairs especially for species with a high reproductive output (Jones & Ardren 2003). For this reason, only few studies in the marine environment have applied this methodology (e.g. Planes *et al.* 2009; Saenz-Agudelo *et al.* 2011; Berumen *et al.* 2012), whereas owing to these investigations our understanding of dispersal patterns has greatly increased. Nonetheless, these studies were undertaken in tropical coral reef environments mostly on species with confined habitats and it would be important to expand such investigations to other habitat types such as the temperate Mediterranean Sea.

Connectivity, if from en evolutionary perspective or a more current ecological perspective, is highly important in determining the natural regulation of populations. By considering both time scales, it is possible to estimate the persistence of a species and its populations. Hence, correct management decisions on conservation issues should include genetic population data combined with oceanographic processes as well as direct measurements of larval dispersal for the full understanding of the population dynamics of species (Almany *et al.* 2009).

Genetics of Individuals

Each individual organism has the ability to alter traits, such as morphological or behavioral traits, in response to external conditions. The 'activation' of certain coding genes is regulated and is different at each given moment. This activation of genes that leads to transcription into RNA sequences and possibly translation into proteins is termed gene expression. The expression of genes is what provides cellular control over structure and function and is the basis for biological processes such as morphogenesis, cell differentiation and the adaptability of an organism. Gene expression can help understand basic physiology as well as evolutionary processes such as adaptation. In particular the latter can give indications on the resilience of species by evaluating its potential to adapt. Furthermore, gene expression and adaptation analyses of the individuals can also help us to understand the population differences among areas, and how these differences are being produced.

Measuring Gene Expression

Not long ago, gene expression was studied by comparing one gene at a time with time-consuming techniques such as for example Northern Blots. This method was widely used despite many shortcomings and was based on the measurement of the level of messenger RNA (mRNA). In 1977, with the discovery of reverse transcriptase, which is an enzyme that can generate complementary DNA (cDNA) from RNA, new and more accurate techniques were introduced (Bustin 2000). Although the reverse transcription polymerase chain reaction (RT-PCR) allowed for more accurate measurements, it is still limited to the analysis of one gene at a time. With new scientific and technological advancements, scientists can now analyze many genes at one time by using for example serial analysis of gene expression or microarrays. Microarrays are powerful analytical tools that can measure the type and amount of mRNA and are widely used in the study of gene expression (Schulze & Downward 2001). Nonetheless, there are drawbacks to consider when using this type of methodology. Microarrays are designed for species-species hybridizations and using cross-species hybridizations, which is necessary for the study of non-model species, are non-standard applications and need to be interpreted with caution (Bar-Or *et al.* 2007). Therefore, most knowledge in gene expression has been won with the studies on model-species such as human or mice and great advances have been made in the field. Roberge and co-authors (2008) point out as a limiting factor to microarrays that some important genes (some candidate genes or novel candidate genes) might be missing on the microarray chip and therefore are not accounted for. With the progress and improvement in technology, high throughput sequencers or sometimes also called next generation sequencers (NGS) were developed in the 1990s. This advancement permitted the real beginning of the 'genome era' as mass sequencing produces sufficient sequence data to generate whole genome assemblies. One important part of the NGS is the 'de novo' sequencing, which refers to methods used to determine DNA sequences with no previous sequence knowledge such as for instance shotgun sequencing. This allows for the study of unknown parts of a model-species genome as well as for the study of non-model species with no prior known genetic sequences. One such method can be used for the evaluation of gene expression, as it uses mass sequencing to sequence cDNA in order to get information on RNA in a sample: RNAseq or also called whole transcriptome shotgun sequencing. RNAseq has an important advantage over microarray analysis because it is not limited to pre-known genes and commercially available arrays (Wang *et al.* 2009). RNAseq

analysis can give valid and relevant information on molecular ecological questions even for complete non-model species. Although this method has been announced as revolutionary in transcriptomics, is comes with large analytical challenges especially for the analysis of species with no reference genome sequence.

Gene expression plasticity

Gene expression is dynamic, and the same gene may act in different ways under different circumstances. This can be the case for the same gene in different parts of the organism or for a given individual in response to external stimuli (phenotypic plasticity).

Phenotypic plasticity is the ability of an organism to change its phenotype in response to changes in the environment. In other words, plasticity allows for the activation or deactivation of certain genes with external cues. Although its importance has long been underestimated it is now clear that it forms an integral part of how organisms develop and interact with their environment. A variety of studies clearly demonstrate the influence of plasticity on ecological and evolutionary processes in many species (Huey & Kingsolver 1989; Pigliucci 2005; Ghalambor *et al.* 2007). Moreover, a heightened awareness of the scope of phenotypic plasticity has shed new light and changed views on the basic paradigms in evolutionary biology (Pigliucci 2007). These new insights into the variability of genomic gene expression I believe follow the motto: 'Life is not just what you (genetically) get, it´s what you make out of it'.

The change in phenotype in fishes is often observed at the moment of reproduction, where two types of males, exhibiting similar genotypes, change behavior or their appearance. Therefore, temporally creating a social hierarchy with males displaying alternative mating strategies where for instance a secondary male or so-called sneaker male can enter the territorial males' nest and fertilize eggs without investing in the nest making or attracting the female (Immler *et al.* 2004). Here, it is an external social cue that provokes the change in phenotype for the dominant male, which often displays a different body coloring throughout the reproductive period. Social structuring in fish has mostly been studied in the brain in cichlid fish species, due to their extreme diversity and the facility to be handled and kept in the laboratory (Taborsky 2008). Also, the Atlantic salmon (*Salmo salar*) exhibits alternative reproductive tactics and the plasticity in the development of these reproductive phenotypes has been investigated (Aubin-Horth & Renn

2009; Guiry *et al.* 2010). Gene expression patterns related to phenotypic plasticity in different mating strategies have been analyzed either by targeting single genes or via microarray analyses (Aubin-Horth *et al.* 2005; Burmeister *et al.* 2005; Renn *et al.* 2008), but no attempt to characterize phenotypic plasticity in alternative mating tactics using a genome-wide approach has been carried out.

The Study Species

For a broader evaluation and understanding of population dynamics, I chose to work with three different types of fish species: the dusky grouper (*Epinephelus marginatus*), the comber (*Serranus cabrilla*) and the black-faced blenny (*Tripterygion delaisi*). These three species display very distinct life history traits from long lived to short lived (Table1). Whereas all three species form an important part of the Mediterranean marine ecosystem, each species has a different conservation status and is of different relevance to humans.

Table 1: Life history traits and Status of the three study species.

Species	**Max. body length (cm)**	**Max. age**	**Age of maturity**	**Length of Pelagic Larval Phase (days)**	**Status of Species**
Epinephelus marginatus	100	60	5-12 years	21- 30	endangered /commercially relevant
Serranus cabrilla	40	5	3(?)	21-28	Commercially relevant
Tripterygion delaisi	8	3	several months	14-24	no documented relevance

Epinephelus marginatus

Photo: Enrique Ballesteros

Epinephelus marginatus (Pisces: Serranidae, subfamily Epinephelinae), also called the dusky grouper is a wide-ranging species, occurring in the Mediterranean Sea and the Eastern Atlantic. It can also be found in the Western Atlantic in South American waters and in the Western Indian Ocean. Especially in the Mediterranean Sea but also in most other locations the dusky grouper is an important and valuable demersal species in coastal fisheries (Heemstra & Randall 1993). As a popular food fish and commercial species it is seen as a flagship species for conservation purposes and in the establishment of Marine Protected Areas (MAPs) in the Mediterranean region (Maggio *et al.* 2006).

The species is a large sized, long-lived (up to 60 years), slow-growing fish with a late onset of maturity and a complex reproductive biology as a protogynous hermaphrodite (Reñones *et al.* 2007). It inhabits rocky shore reefs up to a depth of about 50 meters on the Mediterranean coast (Heemstra & Randall 1993). The spawning behavior and ecology have been observed in the Medes Marine Reserve in Spain and other aggregation sites where the dusky grouper form small spawning aggregations and show strong site fidelity (Zabala *et al.* 1997; Hereu *et al.* 2006). Although migration pattern have been documented in many species of the family Serranidae *Epinephelus marginatus* demonstrated strong homing with mainly the pelagic larvae contributing to possible gene flow (Sadovy *et al.* 1994). The dusky grouper larva is seldom captured in the plankton but it is known to remain for 21-30 days in the pelagic stage (Macpherson & Raventós 2006). Alike numerous other commercially over-fished and threatened marine species, data on its dispersal potential is scarce (Jones *et al.* 2007).

Serranus cabrilla

Photo: Enrique Ballesteros

The comber *Serranus cabrilla* is a common demersal species inhabiting the Eastern Atlantic Ocean and Mediterranean Sea and it is found in sea grass beds and rocky bottoms, with a wide bathymetric range (5-500m). The comber is considered one of the most important predators of early stage of fish and vagile invertebrates (Guidetti & Cattaneo-Vietti 2002). Combers are economically relevant and included in FAO catalogues as species of interest to fisheries in the Central-eastern Atlantic, the Mediterranean and the Black Sea. Their larvae remain in the plankton for 21-28 days (Raventós & Macpherson 2001) and have been collected not only inshore but also over the continental shelf at a considerable distance from the habitats of the adults (Sabates *et al.* 2003). Much is known about its ecology and biology (Torcu-Koc *et al.* 2004), however there is no information about its population structure and degree of connectivity between populations.

Tripterygion delaisi

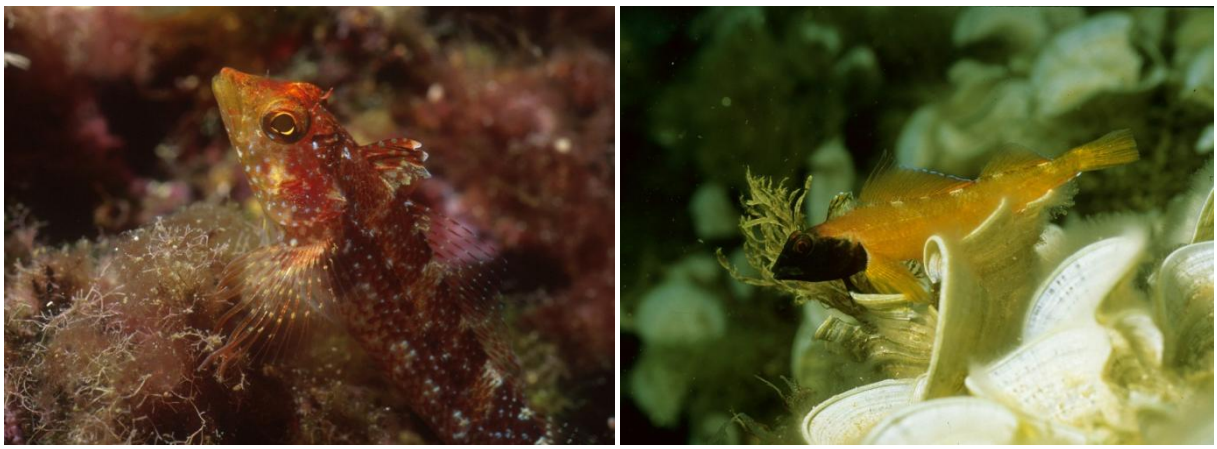

Photo: Enrique Ballesteros

Tripterygion delaisi also called the black-faced blenny, is a common small rocky shore fish from the Mediterranean Sea and the east Atlantic coast (Carreras-Carbonell *et al.* 2005; Domingues *et al.* 2007). The black-faced blennies live camouflaged with the rock or algae they inhabit for most time of the year. In spring, when the reproductive period starts, some males change their color to a black head and a bright yellow coloring for the rest of the body. These males will start protecting a small territory, which is referred to as their nest, against predators and other secondary males (Jonge & Videler 1989). The sneaker males as well as the females display the same camouflaged phenotype throughout the whole year. The dominant male then tries to attract the female by courtship, which could be referred to as dancing, above the nest. The female subsequently lays the eggs directly on the rock, sponge or algae where the nest is found and the dominant male fertilizes the eggs by ejecting the sperm directly on them (Wirtz 1978). However, the sneaker male can dart by and eject its sperm as well, only from a further distance and thus 'sneak' reproduction.

The population structure of the black faced blenny has been evaluated in previous studies (Carreras-Carbonell *et al.* 2006, 2007) providing valuable information for analysis of connectivity on a smaller scale by applying parentage analysis. Due to the particular mating process of this species, especially as the dominant males build nests and become territorial as well as change color and are easily seen, this species is an ideal candidate for such an investigation. Furthermore, as the species displays different phenotypes during the reproductive

period and individuals of either phenotype are easy to catch, *Tripterygion delaisi* is also optimal for the genomic investigation of phenotypic plasticity.

In this thesis, I approach a variety of ecological and evolutionary challenges from different angles. In **chapter 3.1** I analyze the genetic population structure of two different Mediterranean fish species. The first analysis focuses on the conservation of the endangered species *Epinephelus marginatus* and attempts to give indications on the status of the species as well as population genetic information for the correct design of management strategies. As it becomes clear that oceanographic processes, such as fronts and currents, are important factors in influencing genetic population structure, I proceed with a more multi-disciplinary analysis of the genetic structuring of *Serranus cabrilla*, a common Mediterranean fish species. The approach includes the comparison of genetic data with oceanographic particle simulations and can give indications on the degree of influence that the physical environment can have on a species genetic distribution.

In **chapter 3.2** I move from a population approach to the individual level approach. The development, physiology and behavior of an organism determine the life history traits as well as the adaptability to changing conditions. I investigate this by evaluating the differences in gene expression and function for males displaying alternative reproductive tactics as well as females. This is the first genome-wide study for a non-molecular model species in the context of alternative mating strategies and provides essential information on the molecular basis of social dominance.

With the production of a de novo transcriptome assembly in chapter 3.2, it can be possible to identify *in silico* single nucleotide polymorphism markers from this type of data. I start in **chapter 3.3** with the development of such markers in *Tripterygion delaisi* exploring for an optimal protocol allowing future SNP developments in non-model species. This type of genetic markers permits a more time and resource-efficient genotyping for a large amount of samples. Therefore, I used SNPs to explore connectivity patterns by using paternity analysis. The direct evidence of dispersal requires a large amount of individuals, but can provide very important insights and is essential to understanding current connectivity patterns in different marine

habitats. This study especially complements the investigations in chapter 3.1 granting a wholesome understanding of population connectivity on evolutionary as well as ecological time scales.

2. OBJECTIVES

OBJECTIVES

1. Evaluate the genetic population structure of an endangered species, *Epinephelus marginatus*, and provide input for the establishment of management units for conservation.

2. Analyze the genetic structuring of a commercially relevant fish species (*Serranus cabrilla*) within the Mediterranean Sea and evaluate connectivity between populations.

3. Identify to what extent the genetic structuring of *Serranus cabrilla* is related or even defined by oceanographic processes, such as current or oceanographic fronts and what effect oceanographic processes can have on gene flow and population connectivity.

4. Directly measure larval dispersal for the black faced blenny (*Tripterygion delaisi*) and, in turn, identify the connectivity pattern on an ecological time-scale.

6. Construct a *de novo* transcriptome assembly for brain tissue of *Tripterygion delaisi* and determine the main biological functions that are expressed on a genome-wide scale.

5. Profile the differential gene expression between the brains of territorial males, sneaker males, and females of *Tripterygion delaisi* to study the molecular signatures of male dimorphism.

6. Develop single nucleotide polymorphism markers (SNPs) from next generation sequencing techniques and evaluate the optimal number of individuals needed for an efficient development of such type of markers for non-model species.

3.1 Genetic population connectivity and oceanographic fronts in littoral fish species

3. PUBLICATIONS

3.1 Genetic population structure and oceanographic fronts in littoral fish species

Publication 1: Genetic connectivity pattern in an endangered species: The dusky grouper (*Epinephelus marginatus*)

Publicación 1: Patrones de conectividad genética para una especie amenazada: El mero (*Epinephelus marginatus*)

RESUMEN

El mero, *Epinephelus marginatus,* es un pez popular y comercial que se considera una especie bandera en la conservación y en el establecimiento de Reservas Marinas en el Mediterráneo. Fue declarada una especie amenazada e incluida en la Lista Roja de la IUCN debido a su sobreexplotación combinado con un comportamiento de fidelidad territorial y de cambio de sexo. En este estudio, analizamos la estructura genética del mero en el Mediterráneo y aguas adyacentes del Atlántico usando 11 marcadores microsatélites altamente polimórficos. *Epinephelus marginatus* presenta una estructura poblacional débil pero significativa y no es panmíctico dentro del área de estudio. La diferenciación genética entre el Atlántico centro-oriental y la cuenca Mediterránea puede estar basada en sucesos históricos y actuales. Se observó que las muestras de Argelia estaban significativamente diferenciadas de las otras localidades del Mediterráneo, probablemente debido a procesos oceanográficos, los cuales forman un patrón complejo de retención y advección en la costa mediterránea del oeste de África. Basados en estos resultados, proponemos tres diferentes unidades de gestión y conservación: Senegal (centro-este del Atlántico), la costa mediterránea del oeste de África (Argelia) y las demás localidades del Mediterráneo, las cuales no están diferenciadas significativamente, aunque se observa un agrupamiento de localidades con influencia de aguas atlánticas. Sugerimos una red de reservas marinas interconectadas por todo el Mediterráneo, incluyendo información genética como herramienta esencial tanto para la conservación y gestión del mero como para la supervivencia a largo plazo de esta especie amenazada.

Journal of Experimental Marine Biology and Ecology 401 (2011) 126–133

Contents lists available at ScienceDirect

Journal of Experimental Marine Biology and Ecology

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

Genetic connectivity patterns in an endangered species: The dusky grouper (*Epinephelus marginatus*)

Celia Schunter [a,b,*], Josep Carreras-Carbonell [a,b], Serge Planes [c], Enric Sala [a], Enric Ballesteros [a], Mikel Zabala [d], Jean-Georges Harmelin [e], Mireille Harmelin-Vivien [e], Enrique Macpherson [a], Marta Pascual [b]

[a] Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Acc. Cala St. Francesc 14, Blanes, 17300 Girona, Spain
[b] Dept. Genètica, Universitat de Barcelona, 08028 Barcelona, Spain
[c] USR 3278 CNRS-EPHE, Criobe, Universite de Perpignan, 66860 Perpignan cedex, France
[d] Dept. Ecología, Universitat de Barcelona, 08028 Barcelona, Spain
[e] Centre d'Océanologie de Marseille, UMR CNRS 6540, Université de la Méditerranée, Station Marine d'Endoume, 13007 Marseille, France

ARTICLE INFO

Article history:
Received 14 July 2010
Received in revised form 19 January 2011
Accepted 28 January 2011

Keywords:
Effective population size
Endangered species
Epinephelus marginatus
Marine Protected Area
Microsatellites
Population structure

ABSTRACT

The dusky grouper *Epinephelus marginatus* is a popular and commercial fish and it is seen as a flag species for conservation purposes and for the establishment of coastal Marine Protected Areas in the Mediterranean region. Due to many years of over-exploitation combined with its site fidelity and sex change behaviour it has been declared endangered by the IUCN red list. Herein we analysed the genetic structure of the dusky grouper within the Mediterranean Sea and adjacent Atlantic waters using 11 highly polymorphic microsatellite markers. *E. marginatus* showed a weak but significant genetic population structure and was not panmictic throughout the studied range. Genetic differentiation between the central-eastern Atlantic and the Mediterranean basin was putatively based on present and historical events. Samples from Algeria were significantly differentiated from other Mediterranean localities, probably due to historical and present oceanographic processes forming a complex pattern of retention and advection on the West African Mediterranean coast. Based on these results, we propose three different management or conservation units: Senegal (central-eastern Atlantic), the West African Mediterranean coast (Algeria) and the rest of the Mediterranean locations, which are not significantly differentiated, although, clustering of locations with Atlantic water influence is observed. We suggest a Mediterranean-wide network of interconnected Marine Protected Areas, taking genetic information into account, as an essential conservation and management tool for the protection and long-term survival of the endangered dusky grouper.

1. Introduction

Understanding population connectivity has become a major factor in determining and defining threats to marine biodiversity and has proven crucial for conservation (Jones et al., 2007; Sheaves, 2009). The detection of reduced connectivity is of particular importance to conservation because of its effect on the risk of species extinction (Frankham, 2006). It also provides the relevant spatial scale and limits of conservation units. Connectivity is the genetic exchange of individuals and their propagules by means of migration and dispersal. In marine ecosystems, genetic long distance exchange is principally driven by pelagic larval stages, and is directly influenced by oceanographic processes such as currents, winds or eddies (Palumbi, 2003). Oceanographic processes have recently shown to greatly influence population connectivity of species (Galarza et al., 2009a), although still relatively few studies directly couple genetic exchange with oceanographic conditions (Gonzalez-Wangüemert et al., 2010; White et al., 2010).

The extent to which genetic population units are connected can have crucial consequences for the state of marine ecosystems. The Mediterranean Sea is regarded as a marine biodiversity hot-spot (Macpherson, 2002), but many fish stocks are in decline, possibly the first signs of ecosystem collapse (Pauly, 2009). This emphasises the necessity of creating a network of Marine Protected Areas. However, the establishment of small and scattered marine reserves that do not consider population connectivity can have a limited effect on the protection of Mediterranean ecosystems (Botsford et al., 2009). Hence, management decisions regarding conservation issues should combine genetic data with oceanographic processes to fully understand the connectivity and population dynamics of a species (Almany et al., 2009).

Epinephelus marginatus (Lowe 1834) (Pisces: Serranidae, subfamily Epinephelinae), commonly called the dusky grouper, is a wide-

* Corresponding author at: Centre d'Estudis Avançats de Blanes (CSIC), Acc. Cala St. Francesc 14, Blanes, 17300 Girona, Spain. Tel.: +34 972 33 61 01; fax: +34 972 33 78 06.
E-mail address: cschunter@ceab.csic.es (C. Schunter).

0022-0981/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jembe.2011.01.021

ranging species occurring in the Mediterranean Sea and the eastern and western Atlantic (Heemstra and Randall, 1993). In the Mediterranean Sea, the dusky grouper is an important and valuable demersal species in coastal fisheries. For the reason that it is a popular food fish and commercial species, it is regarded a flagship species for conservation purposes and in the establishment of Marine Protected Areas (MPAs) in the Mediterranean region (Maggio et al., 2006).

The species grows large in size, is long-lived (up to 60 years), but slow-growing with a late onset of maturity (at the age of five) and has a complex reproductive biology as a protogynous hermaphrodite (Reñones et al., 2007). It inhabits rocky shore reefs down to a depth of 50 m along the Mediterranean coast (Heemstra and Randall, 1993). The dusky grouper forms small spawning aggregations and displays strong site fidelity (Zabala et al., 1997; Hereu et al., 2006). Based on its life history characteristics, it is presumably the pelagic larvae that contribute mainly to the gene flow (Sadovy et al., 1994). The dusky grouper larvae are rarely captured in plankton nets, but are known to remain for 21–30 days in the planktonic stage (Macpherson and Raventos, 2006). As for numerous other commercially over-fished and threatened marine species, data on its dispersal potential is scarce (Jones et al., 2007).

Owing to its popularity as a target species for commercial and recreational fishermen, the dusky grouper population has dramatically declined in numbers over the last decades with alarming declines reported across its range in the Mediterranean Sea (Cornish and Harmelin-Vivien, 2004). Consequently, the species was listed as endangered by the International Union for the Conservation of Nature (IUCN) in 1996 and by the European Centre for Nature Conservation (ECNC) (1998) (De Almeida Rodrigues Filho et al., 2009).

Genetic differentiation between the Atlantic and the Mediterranean as well as within the Mediterranean has been reported for *E. marginatus* (Gilles et al., 2000; De Innocentiis et al., 2001; Maggio et al., 2006). However, in a previous study only few individuals of each locality were analysed and several distant localities were pooled together in some analyses (De Innocentiis et al., 2001). Several of these pooled localities have been shown to be genetically differentiated due to oceanographic processes in other fish species (Galarza et al., 2009a).

The predominant circulation patterns and oceanographic patterns within the Mediterranean Sea are well described (Millot, 1999, 2005; Fernandez et al., 2005). The Mediterranean Sea consists of two partly isolated basins, the western and eastern basin, which are connected by the Strait of Sicily (SS) and the narrow Strait of Messina (Lermusiaux and Robinson, 2001) (Fig. 1). After passing through the Strait of Gibraltar, the Atlantic water encounters the higher density Mediterranean water. This generates different oceanographic discontinuities, such as the Almeria-Oran Front (AOF) and the Balearic Front (BF) which have been observed to create barriers to gene flow (Patarnello et al., 2007; Galarza et al., 2009a). Besides, there are also some well known oceanographic processes around the Ibiza channel (IC) that can cause water mass discontinuities (Fernandez et al., 2005). These have rarely been considered in population genetic studies of marine species. Consequently, for the conservation of *E. marginatus* it is vital to understand its population dynamics and the influencing oceanographic processes within its range. In this study, we focus on identifying the genetic population structure of the dusky grouper within the Mediterranean Sea and adjacent Atlantic waters. Eleven microsatellite markers and approximately 40 individuals per sample site from a total of 10 localities along the Mediterranean region and the eastern Central Atlantic Ocean were used. The objective of this study was to identify separate management units for this flagship species based on genetic differentiation providing a sound support for suitable conservation measures.

The specific aims of this study were to: (1) analyse the alternative hypotheses of panmixia versus genetic population subdivision, (2) identify to what extent the genetic population structure of *E. marginatus* is affected or even shaped by oceanographic processes, (3) give an indication of the genetic diversity and effective population size of the species, and (4) provide input in defining managements units for conservation of the genetic diversity.

2. Materials and methods

2.1. Sampling and DNA extraction

A total number of 362 samples of *E. marginatus* were collected by line and spear-fishing between 2 and 30 m depth between 2002 and 2006. Sampling was carried out in nine locations within the Mediterranean Sea (Cyclades Islands (Greece, n = 40), Annaba (Algeria, n = 37), Tunis (Tunisia, n = 40), Corsica (France, n = 32), Menorca (Spain, n = 40), Mallorca (Spain, n = 40), Barcelona (Spain, n = 40), the marine reserve of the Medes Islands (Spain, n = 20) and Cabo de Gata (Spain, n = 33)) and Dakar (Senegal, n = 40) as one location on the Atlantic coast of Africa (Fig. 1). Samples from Tunisia and Senegal were collected from fishermen at Mercabarna wholesale fish market in Barcelona. Tissue samples of caught specimens, as well as fin clips from live fish, were taken and stored in 100% EtOH at − 20 °C. DNA was extracted of a 3 mm large fin piece with 500 µl of 10% Chelex (Estoup et al., 1996) or QIAamp DNA mini Kit (50 µl) tissue protocol.

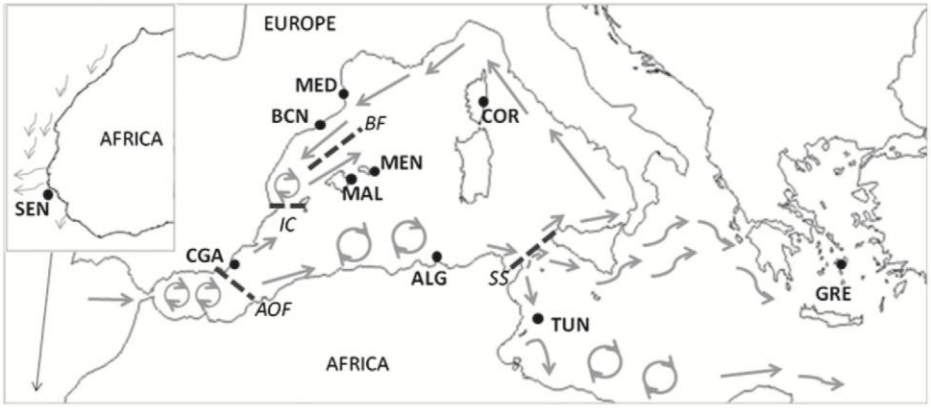

Fig. 1. Map of the ten locations where *Epinephelus marginatus* individuals were sampled. GRE = the Cyclades, ALG = Algeria, TUN = Tunisia, COR = Corsica, MEN = Menorca, MAL = Mallorca, BCN = Barcelona, MED = The Medes Islands, CGA = Cabo de Gata, SEN = Senegal.Basic current patterns (modified from Millot, 2005) and flow directions are indicated by arrows. The main oceanographic discontinuities are identified by dashed lines. AOF = Almeria-Oran front, BF = Balearic front, IC = Ibiza Channel, SS = Sicily Strait.

2.2. Microsatellite DNA analyses

Eleven microsatellites were amplified in three multiplex reactions (Supplementary material Table S1), of which four were isolated from *Mycteroperca microlepis* (GAG007, GAG038, GAG045, and GA049; Chapman et al., 1999), five from *Epinephelus guttatus* (RHCA7, RHCA1, RHCA3, RHCA4, and RHCA8; Ramirez et al., 2006), and one loci (D076) was cloned and sequenced from that same *E. guttatus* library. One primer pair, (SC06), was developed for *Serranus cabrilla* (Carreras-Carbonell et al., 2006b). Each primer pair was colour-labelled with either 6FAM, NED, VIC or PET and amplified using a PCR cycle of 94 °C for 15 min followed by 30 cycles of 30 s denaturation at 94 °C, 90 s of annealing at 52 °C and 60 s of extension at 72 °C, which was followed by a final extension of 60 °C for 30 min. Each amplification reaction of 11 μl consisted of 5 μl of master mix of the multiplex PCR kit (Qiagen), 2.5 μl of Primer mix (2 mM of each Primer), 2 μl of distilled water and 1.5 μl of DNA from the Chelex supernatant (approx. 60 μg of DNA). The amplified PCR products were screened using an ABI 3700 automatic sequencer. Alleles were sized and scored with the Genemapper™ software by comparison to the internal size standard genescan-500 LIZ (Applied Biosystems, Inc.).

2.3. Data analysis

Genetic diversity, allele frequencies, linkage disequilibrium and deviation from the Hardy–Weinberg equilibrium were estimated with the Genepop 3.4 software (Raymond and Rousset, 1995). The presence of null alleles was inferred using the programme Micro-checker (Van Oosterhout et al., 2004). Allelic richness values were assessed with FSTAT by accounting for the sample size (Goudet, 2002).

F-statistics (Wright, 1969) were calculated for all populations and all pairs of populations as for Weir and Cockerham (1984) with Genepop. To correct for multiple tests (Benjamini and Yekutieli, 2001), the modified False Discovery Rate (FDR) was applied (Narum, 2006). This method has been found to provide the most biologically important critical value when comparing it with the Bonferroni correction or the original FDR. Furthermore, the more recent heterozygosity-independent Jost D (Jost, 2008), which has been shown to reliably reflect genetic differentiation (Gerlach et al., 2010), was calculated with SMOGD (Crawford, 2010). To test the hypothesis of isolation by distance, a Mantel test was performed between the matrix of pairwise genetic ($F_{ST}/(1-F_{ST})$) and geographic (log km) distance with Genepop. Geographic distances were measured by following predominant current patterns between each location. In order to identify genetic barriers to gene flow, the software Barrier 2.2 (Manni et al., 2004) was employed with F_{ST} values. This programme applies the Delaunay triangulation producing a connectivity network of populations to which possible genetic barriers are superimposed. The number of genetic clusters was inferred by using the admixture analysis in the Structure 2.3.3 software (Pritchard et al., 2000) 100000 long burning periods, 500000 MCMC Reps and with K tested from 1 to 14. Each K was run for 20 iterations and their likelihood values tested following an ad hoc statistical test suggested by Evanno et al. (2005). This supplementary procedure can result in a clear and exact peak of K, thus allowing the identification of management units.

Principle coordinate analyses were performed using the F-statistics matrix of standardised population distances, plotting patterns within a multivariate data set. The procedure was computed and graphed with GenAlex (Peakall and Smouse, 2006) which is based on the algorithm of Orloci (1978). The effect of oceanographic processes on genetic population connectivity in the Mediterranean Sea was assessed by standardising the genetic distance by geographic distance. We used the Cavalli-Sforza chord distance (Cavalli-Sforza and Edwards, 1967), computed with the software MSA (Dieringer and Schlötterer, 2003), because it has only positive values. To evaluate the effects of barriers on genetic structuring of grouper populations, different oceanographic processes were tested by comparing population pairs affected by the discontinuities of interest (BF, IC and SS) with population pairs which are not affected (No front) by any processes.

Effective population sizes (Ne) were estimated using two different approaches. Direct estimates were calculated based on microsatellite heterozygosity using the stepwise mutation model (Ohta and Kimura, 1973) with the formula $Ne = [1/(1-He)^2 - 1]8\mu$. A full-likelihood Markov chain Monte Carlo (MCMC) based approach was used to estimate the likelihood surface for θ, using the programme MIGRATE-n (Beerli, 2008). Mean values out of five runs using default options with the Brownian microsatellite model were presented. The mutation rate (μ) of $5.56^{*}10^{-4}$ originally estimated for the common carp *Cyprinus carpio* (Yue et al., 2007), was applied for all approaches. Effective population size Ne was then calculated for each management unit previously established in Structure. In order to avoid biases due to sample size, an equal random subset of individuals (approximately 40 individuals) were chosen for each unit.

3. Results

3.1. Genetic variability and diversity

A total of 269 alleles were found within 11 loci. These were all highly polymorphic, except RHCa3 and GAG007, for which only two to four alleles could be observed. Following the False Discovery Rate correction, no linkage disequilibrium was found, which allows for the loci to be considered statistically independent. The mean observed heterozygosity (Ho) for the different samples ranged from 0.658 in Mallorca to 0.769 for the samples collected at the Medes Islands (Table 1; Supplementary material Table S2).

Two samples (Menorca and Senegal) demonstrated some deviations from Hardy–Weinberg equilibrium, but at different loci. Allelic richness ranged between 7.61 and 9.46, with the highest allelic richness observed in Senegal (9.46), Cabo de Gata (8.39) and the Medes Islands (8.27) (Table 1).

3.2. Genetic population structure

The global F_{ST} was highly significant ($F_{ST} = 0.0108$, $p<0.001$) revealing significant genetic differentiation between populations. Pairwise F_{ST} values as well as Jost D values were significant and highest mainly for comparisons involving Senegal and Algeria (Table 2). Only one significant F_{ST} estimate which did not include one of these two samples was observed, namely between Greece and Corsica. Algeria was significantly different from all other samples except the Medes Islands and Menorca. The highest F_{ST} values were observed when comparing each location with the samples from Senegal, with the highest F_{ST} estimate observed between Senegal and Greece ($F_{ST} = 0.0471$). The Mantel test revealed significant isolation

Table 1
Summary statistics of genetic diversity of *Epinephelus marginatus* for all localities. N: number of individuals, Ho: observed heterozygosity, He: expected heterozygosity, F_{IS}: fixation index.

Location	N	Allelic richness	Ho	He	F_{IS}
Greece	40	7.78	0.725	0.705	−0.028
Corsica	32	7.88	0.729	0.733	0.006
Mallorca	40	7.86	0.658	0.703	0.065
Menorca	40	7.89	0.706	0.722	0.022
Medes	20	8.20	0.768	0.743	−0.035
Barcelona	40	7.60	0.683	0.704	0.030
Cabo de Gata	33	8.39	0.735	0.733	−0.002
Algeria	37	7.73	0.753	0.712	−0.058
Tunisia	40	7.96	0.731	0.724	−0.010
Senegal	40	9.45	0.722	0.762	0.056

Table 2
Estimates of pairwise F_{ST} (below diagonal) and Jost D genetic distances (above diagonal) among population pairs of *Epinephelus marginatus*. Significant values of F_{ST} ($p<0.01$ after FDR correction) are highlighted in grey.

	Greece	Corsica	Mallorca	Menorca	Medas	Barcelona	Cabo Gata	Algeria	Tunisia	Senegal
Greece	---	0.0075	0.0122	0.0042	0.0009	0.0029	0.0000	0.0168	0.0019	0.1424
Corsica	0.0085	---	0.0017	0.0004	0.0000	0.0002	0.0007	0.0116	0.0002	0.0626
Mallorca	0.0099	0.0043	---	–0.0003	0.0045	0.0000	–0.0015	0.0057	0.0016	0.0859
Menorca	0.0082	0.0037	–0.0018	---	0.0009	0.0000	0.0000	0.0005	0.0009	0.0739
Medas	0.0051	–0.0006	0.0064	0.0005	---	0.0000	–0.0036	0.0002	0.0001	0.0473
Barcelona	0.0054	0.0016	–0.0025	0.0004	0.0035	---	0.0007	0.0050	0.0000	0.0672
Cabo Gata	0.0012	0.0023	–0.0009	0.0002	–0.0038	0.0021	---	0.0046	0.0000	0.0706
Algeria	0.0135	0.0189	0.0116	0.0033	0.0032	0.0165	0.0092	---	0.0043	0.1095
Tunisia	0.0035	0.001	0.004	0.0036	–0.0007	0.0006	0.0005	0.0126	---	0.0840
Senegal	0.0471	0.0271	0.0329	0.0234	0.0205	0.0306	0.0218	0.0454	0.0275	---

by distance including all 10 samples ($p=0.027$). When Senegal was excluded from the analysis, no significant isolation by distance ($p=0.167$) was found.

Patterns of genetic structure revealed by the Barrier analysis were consistent with analysis of F_{ST} values, which significantly separated Senegal from all other samples (Fig. 2). A second barrier was observed between some of the western (Algeria, Cabo de Gata, Mallorca, Menorca, Medes) and eastern (Barcelona, Corsica, Tunisia and Greece) Mediterranean sampling sites. The bi-plot of PCA with only Mediterranean localities (Fig. 3), including the first two PCs, explained in total 80% of the variance, with PC1 explaining 51% and clearly separating Algeria from the other localities. Three genetically differentiated populations were revealed by the ad hoc test after Evanno et al. (2005) (Supplementary material Fig.S1) although the estimated membership coefficient (Q) for each individual in each of the $K=3$ inferred clusters only separated in a different group the individuals from Senegal (data not shown). The three groups found with the ad hoc test could correspond, as indicated by the pairwise F_{ST} comparisons and the PCA analysis, to Senegal as one unit, Algeria as the second one and all other samples grouping together within a third genetic unit. Furthermore, the results obtained with Barrier (Fig. 2) suggest that Algeria is related to the Mediterranean populations with a higher influence of Atlantic water.

No effects of oceanographic processes on the genetic structuring of *E. marginatus* in the Mediterranean were revealed. Standardised Cavalli-Sforza chord distance pairwise comparisons of samples from either side of the Ibiza Channel ($0.25+/-0.050$), the Sicily Strait ($0.26+/-0.01$) and the Balearic Front ($0.23+/-0.06$) did not differ significantly from the population pairs without influence of oceanographic processes ($0.25+/-0.04$). The small standard deviations are indicative of similar genetic distances among population pairs independently of geographic distance.

Effective population sizes (Ne) were calculated for the three genetically differentiated management units Senegal, Algeria and the

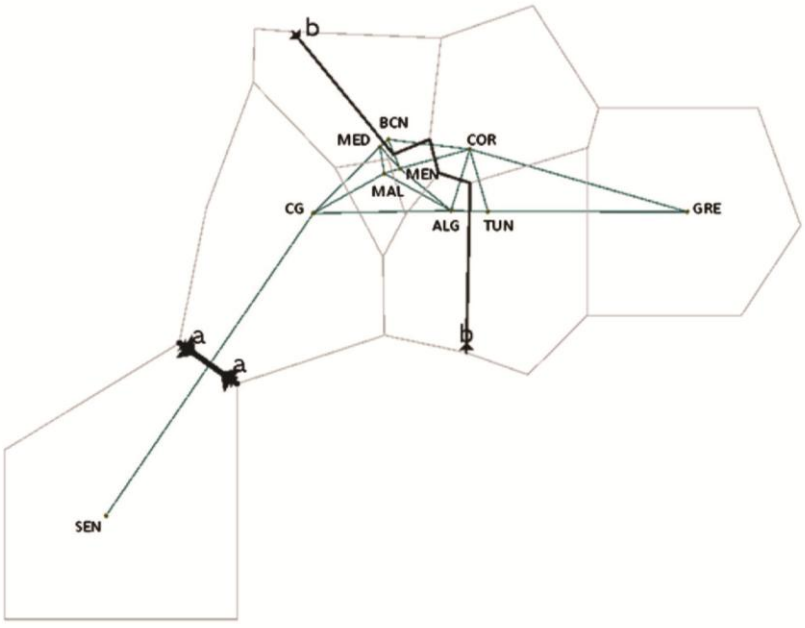

Fig. 2. Connectivity network of *Epinephelus marginatus* including geographical and genetic data by Barrier 2.2. Putative barriers to gene flow (a and b) are indicated with a grey line. Thickness is indicative of strength.

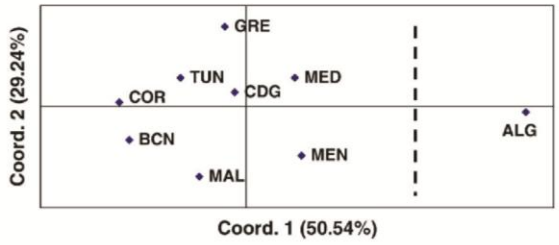

Fig. 3. Principal coordinates analysis based on F_{ST} pairwise distance matrix of the nine Mediterranean populations of *Epinephelus marginatus*.

rest of the Mediterranean Sea. Both methods used indicate that Senegal has the overall highest Ne and that Algeria has a lower effective population size than the rest of the Mediterranean (Table 3).

4. Discussion

4.1. Genetic variability and diversity

The large number of highly polymorphic microsatellites permitted a survey of neutral genetic variation over a part of *E. marginatus*' range. Compared to a previous study on *E. marginatus* in the Mediterranean that used fewer microsatellite loci, we obtained a similar level of genetic diversity (De Innocentiis et al., 2001). In this study, allelic richness was higher, but not significantly different, in three populations: Senegal, Cabo de Gata and the Medes Islands. Elevated diversity of alleles displayed by samples from Cabo de Gata and the Medes Islands could be a result of conservation measures: Cabo de Gata was declared a Marine Protected Area in 1987 and the Medes Island in 1983. In comparison, the high allelic richness in Senegal might suggest that the Atlantic is an important element of the present genetic diversity possibly reflecting also the general healthier state of the ecosystem on the Atlantic central eastern coast owing to less exploitation and environmental destruction in comparison to the Mediterranean Sea. Besides, the East Atlantic might be considered as the historical source of expansion of the species into the Mediterranean Sea with Senegal representing an elevated number of alleles. High allelic richness can be observed in populations from the ancestral range of the species and diversity patterns have been used to explain historical expansion events of marine species, e.g. *Raja clavata* (Chevolot et al., 2006) or *Diplodus sargus* (Bargelloni et al., 2005). Similarly, the higher allelic richness in Cabo de Gata could respond to either ancestral or current processes relative to the Atlantic influence.

4.2. Genetic population structure

Genetic structuring of *E. marginatus* populations within the studied area was revealed by significant global and pairwise F_{ST} values. Significant genetic differentiation of *E. marginatus* populations between areas in the Mediterranean Sea was also encountered by De Innocentiis et al. (2001). Neither De Innocentiis et al. (2001) nor this study, however, observed isolation by distance. Several other studies also failed to detect a significant correlation between genetic and geographic distance, such as for the goliath grouper *Epinephelus itajara* along the northern coast of Brazil (Silva-Oliveira et al., 2008). In contrast, isolation by distance over a large scale, including the Mediterranean Sea and the Atlantic Ocean, has been revealed for *D. sargus* (Domingues et al., 2007), *Tripterygion delaisi* (Carreras-Carbonell et al., 2006a) and *Mullus surmuletus* (Galarza et al., 2009b). The presence of isolation by distance does not seem to be related to the pelagic larval duration as it is similar (between 27 and 31 days) in these Mediterranean species including *E. marginatus* (Macpherson and Raventos, 2006). The higher population connectivity in the dusky grouper might be related to the decrease of the strength of oceanographic discontinuities in summer (Fernandez et al., 2005), which coincides with the reproductive season and the pelagic stage of the species (Zabala et al., 1997).

Table 3
Effective population sizes of *Epinephelus marginatus* for the three genetically differentiated management units calculated by the Ohta and Kimura method and with Migrate-n.

Management units	Method	
	Ohta and Kimura	Migrate-n
Mediterranean Sea	2659	3498
Algeria	2496	2851
Senegal	3773	3830

When including samples from Senegal population in the analysis, the significant relationship between geographic and genetic distance is most likely erroneous, as it resulted from a putative barrier between the Atlantic and Mediterranean basins. Multiple studies on marine species have focused their research efforts on genetic population divergence between the Atlantic Ocean and the Mediterranean Sea. Some found a clear division between the two basins explained by current processes of oceanographic barriers to gene flow, such as the Strait of Gibraltar or the Almeria-Oran front (Patarnello et al., 2007; Galarza et al., 2009a,b). Other oceanographic discontinuities, such as the Sahara upwelling, along NW Africa, can also influence the gene flow in some species (Gonzalez and Zardoya, 2007). A previous genetic study of *E. marginatus* using ND2 RFLP between the Mediterranean and the Azores Islands in the Atlantic Ocean also found significant differences between these two regions (Maggio et al., 2006). Nonetheless some marine taxa of the Macaronesian islands appear as genetically divergent when compared with most Atlantic "continental" populations due to historical processes (Carreras-Carbonell et al., 2005; Chevolot et al., 2006). Thus, more refine sampling within the Atlantic is needed in order to identify the oceanographic discontinuities causing the genetic differentiation between the Atlantic and Mediterranean basins in *E. marginatus*.

Within the Mediterranean Sea, we found genetic differentiation between the Algerian population and the rest of the Mediterranean localities. A previous study using mtDNA of 29 individuals of *E. marginatus* from several Mediterranean areas reported extensive genetic differentiation among Algerian individuals, and claimed the presence of a cryptic subspecies in that area (Gilles et al., 2000). In a preliminary study with mitochondrial genes Cytochrome b and Cytochrome Oxidase I we observed no genetic differentiation between the samples from Algeria and those from the other localities. Furthermore, based on microsatellite loci, our samples grouped together and were clearly differentiated from other species of the same genus (data not shown). Thus, we can rule out the presence of a cryptic subspecies in our data to explain the genetic differentiation in Algeria. The divergence of the Algerian population could be a result of historical isolation in combination with the mesoscale eddy system produced by the entering of the Atlantic water through the Strait of Gibraltar thus creating a complex pattern of retention and advection. Millot (2005) clearly documented the presence of a mesoscale phenomenon of coastal eddies, that were already established during the Pleistocene and have been stable until the present (Cacho et al., 1999). These eddies, while slowing down the eastward moving current, are connected to open-sea eddies in the middle basin towards the Balearic Islands. Retention of larvae by coastal eddies could explain why the Algerian population was genetically differentiated, but also to some extent connected by off-shore eddies to the Balearic Island of Menorca and the marine reserve of the Medes Islands. Such a south–north recruitment of dusky grouper has been previously suggested (Bodilis et al., 2003). Additionally, coastal eddies have

been documented to cause retention of larvae, for example in the rock lobster *Jasus edwardsii* on the coast of New Zealand (Chiswell and Booth, 1999), or the reef-building polychaete *Sabellaria alveolata* in the English channel (Ayata et al., 2009). Effects of retention nevertheless depend on the type of larvae and are most likely coupled with several other processes, such as local topography, vertical larval movement and tidal regimes (Cowen, 2002).

Additionally, a putative barrier indicated a slight influence of the Atlantic waters on the gene flow among localities. The Atlantic water entering through the Strait of Gibraltar flows north-eastward reaching the Balearic Islands as well as moves along the coast of Northern Africa, creating an area with greater Atlantic water influence which includes the Southeast Spanish coast, the Balearic Islands and the coast of Morocco and Algeria (Fernandez et al., 2005; Millot, 2005). This pattern is reflected by a weak genetic barrier between the Atlantic-influenced localities and the more Mediterranean sites. However, this separation is only significant for the Algerian population.

The modified Atlantic water then crosses through the Strait of Sicily. There is a gradient of increasing salinity from the Atlantic along the coast of Northern Africa and across the SS. This salinity gradient causes an important disjunction at the SS, separating water masses from the eastern Mediterranean basin from the water masses of the western Mediterranean basin (Lermusiaux and Robinson, 2001). Significant genetic differentiation between fish populations from the eastern and the western basin of the Mediterranean Sea was revealed, for example, in the Mediterranean Sea bass *Dicentrarchus labrax* (Bahri-Sfar et al., 2000) and the red mullet *Mullus barbatus* (Galarza et al., 2009b). Similarly, an effect of the SS was discovered on mackerel species (*Scomber* spp., Zardoya et al., 2004) as well as the European sprat (*Sprattus sprattus*, Debes et al., 2008). In this study, the genetic differences between Algeria and Tunisia may suggest an effect of the SS on the dusky grouper populations as our Tunisian locality was located on the eastern side of the SS. However, the strong connectivity between the eastern and western Mediterranean populations may indicate that the Strait of Sicily was not a barrier to dispersal of *E. marginatus*, but that the retention processes around Algeria were more likely responsible for isolation and differentiation of the Algerian population.

The low genetic population differentiation in *E. marginatus* within the Mediterranean as well as the lack of effect of predominant oceanographic processes may be explained by the species larval behaviour and reproductive and recruitment strategies. *E. marginatus* displays high site fidelity where adults do not migrate (Hereu et al., 2006), thus resulting in adult genetic isolation. The existing intra-Mediterranean panmixia (excluding West African Mediterranean coast) could therefore be explained by the movement of larvae or juvenile fish. The dusky grouper has a moderately long pelagic larval phase of 21–30 days (Spedicato et al., 1998; Macpherson and Raventos, 2006) which could account for the observed gene flow, spreading genetic variation throughout most of the Mediterranean Sea. The development of *Epinephelus* larvae has recently been described owing to the successful rearing of larvae, indicating that the larvae are fast growing and large in size (Cuhna et al., 2009; Kawabe and Kohno, 2009). However, larval behaviour in the plankton is still not well understood and thus warrants further research. Nevertheless, the lack of effect of predominant oceanographic processes on the gene flow in most Mediterranean populations can also be related with the decrease of the water transport during late summer and early autumn, affecting the strength of some frontal systems (Fernandez et al., 2005; Millot, 2005). This decrease occurs during the larval period of *E. marginatus* (Zabala et al., 1997) facilitating the gene flow among localities. Other species with a planktonic development during late summer and early autumn, e.g. *Diplodus puntazzo*, also show an intra-Mediterranean panmixia (Bargelloni et al., 2005). The hydrography of the Mediterranean Sea is strongly influenced by seasonality and temperature can oscillate up to 15 °C. The presence of frontal structures and their dynamism has a complex interaction with the spatial variability and the stratification development (Fernandez et al., 2005). Therefore, the variations in the strength of the frontal systems and their effect on the population genetic structure of Mediterranean organisms might be species specific. The differences in population structure of closely related species, e.g. *Mullus surmuletus–M. barbatus* (Galarza et al., 2009b), *Scomber japonicus–S. scombrus* (Zardoya et al., 2004) demonstrate the complexity of the relationships between life history, oceanography and population structure.

4.3. Effective population size

Effective population size (Ne) measures inbreeding and the loss of genetic variation (Frankham, 2006) and is thus an assessment of the genetic vulnerability of an endangered species (Mace and Lande, 1991). The reduction of genetic variability is particularly detrimental to the recovery of an overexploited species (Hutchingson and Reynolds, 2004). However, Ne calculations not based on temporal data represent the historical Ne and thus can only provide a rough indication of the present status of the species. *E. marginatus* revealed an effective population size (Mediterranean population: 2659, West African Mediterranean population: 2496, and Senegal population: 3773) with values similar to other endangered species, such as the copper redhorse *Moxostoma hubbsi* (Lippé et al., 2006), the flatfish *Pleuronectes platessa* (Hoarau et al., 2005), and the Atlantic cod *Gadus morhua* (Poulsen et al., 2006). With an effective population size of at least 500 to 5000 individuals, a species is considered to retain its evolutionary potential, at which it is able to adapt and evolve in response to environmental changes (Frankham et al., 2002). According to effective population sizes documented in this study, *E. marginatus* may cope with environmental changes. Yet, as Ne values are historical, they do not indicate accurate changes in recent decades of overexploitation. Consequently, present day changes in genetic variability of the species should be monitored to estimate current Ne values, and in so doing warn of dramatic declines which could lead to a high risk of extinction.

5. Conclusions

The dusky grouper has a weak genetic population structure and is not panmictic throughout the studied range. Three significantly differentiated management or conservation units were revealed: Senegal (central-eastern Atlantic), Algeria (West African Mediterranean coast) and the rest of the Mediterranean populations. Genetic differentiation between the Atlantic and the Mediterranean basin is putatively based on present and historical events; nonetheless, further research is required to define the expanse of the Central Eastern Atlantic unit. Algeria is significantly differentiated from other Mediterranean localities probably due to historical and present oceanographic processes forming a complex pattern of retention and advection on the West African Mediterranean coast. The rest of the Mediterranean sampling sites are not significantly differentiated, although, clustering of locations with Atlantic water influence is observed. The Mediterranean Sea has long suffered from high exploitation and effective implementation and execution of conservation measures is long overdue (Lleonart and Maynou, 2003). We suggest a Mediterranean-wide network of interconnected Marine Protected Areas as an essential conservation and management tool for the protection and long-term survival of the endangered dusky grouper.

Supplementary materials related to this article can be found online at doi:10.1016/j.jembe.2011.01.021.

Acknowledgements

We thank all the people who helped us to collect samples, with special mention to Laura Caldentey and the Gremi de Consignataris Majoristes de Peix i Marisc of Barcelona. We are also grateful to the Direcció General de Pesca i Cultius Marins del Govern de les Illes Balears for the sampling permits in the Balearic Islands. We are also thankful to Mara Mackle with her help on the manuscript. This work was partially funded by the Spanish Ministry from Science and Innovation through projects GRACCIE (C5D2007-0067), MARMOL (CICYT-CTM2007-66635-C02-01) and BIOCON08-187 (FBBVA). The authors are part of the research group 2009SGR-636, 2009SGR-665 of the Generalitat de Catalunya. **[RH]**

References

Almany, G.R., Connolly, C.S., Heath, D.D., Hogan, J.D., Jones, G.P., McCook, L.J., Mills, M., Pressey, R.L., Williamson, D.H., 2009. Connectivity, biodiversity conservation and the design of marine reserve networks for coral reefs. Coral Reefs 28, 339–351.

Ayata, S., Ellien, C., Dumas, F., Dubois, S., Thiébaut, E., 2009. Modelling larval dispersal and settlement of the reef-building polychaete *Sabellaria alveolata*: role of hydroclimatic processes on the sustainability of biogenic reefs. Cont. Shelf Res. 29, 1605–1623.

Bahri-Sfar, L., Lemaire, C., Ben Hassine, O.K., Bonhomme, F., 2000. Fragmentation of the seabass populations in the western and eastern Mediterranean as revealed by microsatellite polymorphism. Proc. R. Soc. Lond. B 267, 929–935.

Bargelloni, L., Alarcon, J.A., Alvarez, M.C., Penzo, E., Magoulas, A., Palma, J., Patarnello, T., 2005. The Atlantic–Mediterranean transition: discordant genetic patterns in two seabream species, *Diplodus puntazzo* (Cetti) and *Diplodus sargus* (L.). Mol. Phylogenet. Evol. 36, 523–535.

Beerli, P., 2008. Migrate version 3.0 — a maximum likelihood and Bayesian estimator of gene flow using the coalescentDistributed over the internet at http://popgen.scs.edu/migrate.html2008.

Benjamini, Y., Yekutieli, D., 2001. The control of false discovery rate under dependency. Ann. Stat. 29, 1165–1188.

Bodilis, P., Ganteaume, A., Francour, P., 2003. Recruitment of the dusky grouper (*Epinephelus marginatus*) in the Northwestern Mediterranean Sea. Cybium 27 (2), 123–129.

Botsford, L.W., Coffroth, M.-A., Paris, C.B., Planes, S., Shearer, T.L., Thorrold, S.R., Jones, G.P., 2009. Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. Coral Reefs 28, 327–337.

Cacho, I., Grimalt, J.O., Pelejero, C., Canals, M., Sierro, F.J., Flores, J.A., Shackleton, N.J., 1999. Dansgaard-Oeschger and Heinrich event imprint in the Alboran Sea temperatures. Paleoceanography 14, 698–705.

Carreras-Carbonell, J., Macpherson, E., Pascual, M., 2005. Rapid radiation and cryptic speciation in mediterranean triplefin blennies (Pisces: Tripterygiidae) combining multiple genes. Mol. Phylo. Evol. 37 (3), 751–761.

Carreras-Carbonell, J., Macpherson, E., Pascual, M., 2006a. Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. Mol. Ecol. 15, 3527–3539.

Carreras-Carbonell, J., Macpherson, E., Pascual, M., 2006b. Primer note: characterization of 12 microsatellite markers in *Serranus cabrilla* (Pisces: Serranidae). Mol. Ecol. Notes 6, 204–206.

Cavalli-Sforza, L.L., Edwards, A.W.F., 1967. Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Genet. 19, 233–257.

Chapman, R.W., Sedberry, G.R., Koenig, C.C., Eleby, B.M., 1999. Stock identification of gag, *Mycteroperca microlepis*, along the southeast coast of the United States. Mar. Biotechnol. 1 (2), 137–146.

Chevolot, M., Hoarau, G., Rijnsdorp, A.D., Stam, W.T., Olsen, J.L., 2006. Phylogeography and population structure of thornback rays (*Raja clavata* L., Rajidae). Mol. Ecol. 15, 3693–3705.

Chiswell, S.M., Booth, J.D., 1999. Rock lobster *Jasus edwardsii* larval retention by the Wairarapa Eddy off New Zealand. Mar. Ecol. Prog. Ser. 183, 227–240.

Cornish, A., Harmelin-Vivien, M., 2004. *Epinephelus marginatus*. IUCN 2010. IUCN Red List of Threatened Species. Version 2010.3www.iucnredlist.org.

Cowen, R.K., 2002. Larval dispersal and retention and consequences for population connectivity. In: Sale, P.F. (Ed.), Ecology of Coral Reef Fishes: Recent Advances. Academic Press, pp. 149–170.

Crawford, N.G., 2010. SMOGD: software for the measurement of genetic diversity. Mol. Ecol. Res. 50, 556–557.

Cuhna, M.E., Quental, H., Barradas, A., Pousao-Ferreira, P., Cabrita, E., Engrola, S., 2009. Rearing larvae of dusky grouper, *Epinephelus marginatus* (Lowe, 1834), (Pisces: Serranidae) in a semi-extensive mesocosm. Sci. Mar. 73 (S1), 201–212.

De Almeida Rodrigues Filho, J., Gomes Sanches, E., De Oliveira Garcia, C.E., Vianna Pannutti, C., Figueiredo Sebastiani, E., Guimaraes Moreira, R., 2009. Threatened fishes of the world: *Epinephelus marginatus* (Lowe:1834) (Serranidea: Epinephelus). Environ. Biol. Fish. 85 (4), 301–302.

De Innocentiis, S., Sola, L., Cataudella, S., Bentzens, P., 2001. Allozyme and microsatellite loci provide discordant estimates of population differentiation in the endangered dusky grouper (*Epinephelus marginatus*) within the Mediterranean Sea. Mol. Ecol. 10, 2163–2175.

Debes, P.V., Zachos, F.E., Hanel, R., 2008. Mitochondrial phylogeography of the European sprat (*Sprattus sprattus* L., Clupeidae) reveals isolated climatically vulnerable populations in the Mediterranean Sea and range expansion in the northeast Atlantic. Mol. Ecol. 17, 3873–3888.

Dieringer, D., Schlötterer, S., 2003. Microsatellite analyzer (MSA) — a platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes 3, 167–169.

Domingues, V.S., Santos, R.S., Brito, A., Alexandrou, M., Almada, V.C., 2007. Mitochondrial and nuclear markers reveal isolation by distance and effects of Pleistocene glaciations in the northeastern Atlantic and Mediterranean populations of the white seabream (*Diplodus sargus*, L.). J. Exp. Mar. Biol. Ecol. 346, 102–113.

Estoup, A., Largiadèr, C.R., Perrot, E., Chourrout, D., 1996. Rapid one - tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. Mol. Mar. Biol. Biotechnol. 5, 295–298.

European Centre of Nature Conservation, 1998. Convention on the conservation of European wildlife and natural habitats. European Treaty Series, p. 104.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620.

Fernandez, V., Dietrich, D.E., Haney, R.L., Tintore, J., 2005. Mesoscale, seasonal and interannual variability in the Mediterranean Sea using a numerical ocean model. Prog. Oceanogr. 66, 321–340.

Frankham, R., 2006. Genetics and landscape connectivity. In: Crooks, K., Sanjayan, M. (Eds.), Connectivity Conservation. Cambridge University Press, New York, USA, pp. 72–96.

Frankham, R., Ballou, J.D., Briscoe, D.A., 2002. Introduction to Conservation Genetics. Cambridge University Press, Cambridge.

Galarza, J., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G., Rico, C., 2009a. The influence of oceanographic fronts and early-life history traits on connectivity among fish populations: a multi-species approach. PNAS 106, 1473–1478.

Galarza, J., Macpherson, E., Turner, G., Rico, C., 2009b. Patterns of genetic differentiation between two co-occurring demersal species: the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*). Can. J. Fish. Aquat. Sci. 66, 1478–1490.

Gerlach, G., Jueterbock, A., Kraemer, P., Deppermann, J., Harmand, P., 2010. Calculations of population differentiation based on GST and D: forget GST but not all of statistics. Mol. Ecol. 19 (18), 3845–3852.

Gilles, A., Miquelis, A., Quignard, J.-P., Faure, E., 2000. Molecular phylogeography of western Mediterranean dusky grouper *Epinephelus marginatus*. C. R. Acad. Sci. Paris 323 (2), 195–205.

Gonzalez-Wangüemert, M., Cánovas, F., Pérez-Ruzafa, A., Marcos, C., Alexandrino, P., 2010. Connectivity patterns inferred from the genetic structure of white seabream (*Diplodus sargus* L.). J. Exp. Mar. Biol. Ecol. 383, 23–31.

Goudet, J., 2002. Fstat version 2.9.3.2. A Program to Estimate and Test Gene Diversities and Fixation Indices. Institute of Ecology, Lausanne, Switzerlandhttp://www.unil.ch/izea/softwares/fstat.

Heemstra, P.C., Randall, J.E., 1993. Groupers of the world (Family Serranidae, subfamily Epinephelinae). An annotated and illustrated catalogue of the grouper, rockcod, hind, coral grouper and lyretail species known to date. FAO Fish. Synop. 125, 1–382.

Hereu, B., Diaz, D., Pasqual, J., Zabala, M., Sala, E., 2006. Temporal patterns of spawning of the dusky grouper *Epinephelus marginatus* in relation to environmental factors. Mar. Ecol. Prog. Ser. 325, 187–194.

Hoarau, G., Boon, E., Jongma, D.N., Ferber, S., Palsson, J., Van der Veer, H.W., Rijnsdorp, A.D., Stam, W.T., Olsen, J.L., 2005. Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (*Pleuronectes platessa* L.). Proc. R. Soc. Lond. B 272, 497–503.

Hutchingson, J.A., Reynolds, J.D., 2004. Marine fish population collapses: consequences for recovery and extinction risk. Bioscience 54 (4), 297–309.

Jones, G.P., Srinivasan, M., Almany, G.R., 2007. Population connectivity and conservation of marine biodiversity. Oceanography (Wash DC) 20, 100–111.

Jost, L., 2008. GST and its relatives do not measure differentiation. Mol. Ecol. 17 (18), 4015–4026.

Kawabe, K., Kohno, H., 2009. Morphological development of larval and juvenile blacktip grouper, *Epinephelus fasciatus*. Fish. Sci. 75, 1239–1251.

Lermusiaux, P.F.J., Robinson, A.R., 2001. Features of dominant mesoscale variability, circulation patterns and dynamics in the Strait of Sicily. Deep Sea Res. 9, 1953–1997.

Lippé, C., Dumont, P., Bernatchez, L., 2006. High genetic diversity and no inbreeding in the endangered copper redhorse, *Moxostoma hubbsi* (Catostomidae, Pisces): the positive sides of a long generation time. Mol. Ecol. 15, 1769–1780.

Lleonart, J., Maynou, F., 2003. Fish stock assessment in the Mediterranean: state of art. Sci. Mar. 67, 37–49.

Mace, G.M., Lande, R., 1991. Assessing extinction threats: toward a reevaluation of IUCN threatened species categories. Conserv. Biol. 5, 148–157.

Macpherson, E., 2002. Large scale gradients of species richness in the Atlantic Ocean. Proc. R. Soc. Lond. B 269, 1715–1720.

Macpherson, E., Raventos, N., 2006. Relationship between pelagic larval duration and geographic distribution of Mediterranean littoral fishes. Mar. Ecol. Prog. Ser. 327, 257–265.

Maggio, T., Andaloro, F., Arculeo, M., 2006. Genetic population structure of *Epinephelus marginatus* (Pisces, Serranidae) revealed by two molecular markers. Ital. J. Zool. 73, 275–283.

Manni, F., Guérard, E., Heyer, E., 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". Hum. Biol. 76 (2), 173–190.

Millot, C., 1999. Circulation in the Western Mediterranean Sea. J. Mar. Syst. 20, 423–442.
Millot, C., 2005. Circulation in the Mediterranean Sea: evidences, debates and unanswered questions. Sci. Mar. 69, 5–21.
Narum, S.R., 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. Conserv. Genet. 7, 783–787.
Ohta, T., Kimura, M., 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. 22, 201–204.
Orloci, L., 1978. Multivariate Analysis in Vegetation Research. Dr W. Junk B. V, The Hague.
Palumbi, S.R., 2003. Population genetics, demographic connectivity, and the design of marine reserves. Ecol. Applic. 13, 146–158.
Patarnello, T., Volckaert, F.A.M.J., Castilho, R., 2007. Pillars of Hercules: is the Atlantic Mediterranean transition a phylogeographical break? Mol. Ecol. 16, 4426–4444.
Pauly, D., 2009. Beyond duplicity and ignorance in global fisheries. Sci. Mar. 73 (2), 215–224.
Peakall, R., Smouse, P.E., 2006. Genealex 6.1: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295.
Poulsen, N.A., Nielsen, E.E., Schierup, M.H., Loeschcke, V., Gronkjaer, P., 2006. Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). Mol. Ecol. 15, 321–331.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure from multilocus genotype data. Genetics 155, 945–959.
Ramirez, M.A., Patricia-Acevedo, J., Planes, S., 2006. New microsatellite resources for groupers (Serranidae). Mol. Ecol. Notes 6, 813–817.
Raymond, M., Rousset, F., 1995. Genepop (Version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.
Reñones, O.P.C., Mas, X., Goñi, R., 2007. Age and growth of the dusky grouper *Epinephelus marginatus* (Lowe 1834) in an exploited population of the western Mediterranean Sea. J. Fish Biol. 71, 346–362.
Sadovy, Y., Colin, P.L., Domeier, M.L., 1994. Aggregation and spawning in the tiger grouper, *Mycteroperca tigris* (Pisces: Serranidae). Copeia 2, 511–516.
Sheaves, M., 2009. Consequences of ecological connectivity: the coastal ecosystem mosaic. Mar. Ecol. Prog. Ser. 391, 107–115.
Silva-Oliveira, G.C., Do Rego, P.S., Schneider, H., Sampaio, I., Vallanoto, M., 2008. Genetic characterisation of populations of the critically endangered Goliath grouper (*Epinephelus itajara*, Serranidae) from the Northern Brazilian coast through analyses of mtDNA. Genet. Mol. Biol. 31 (4), 988–994.
Spedicato, M.T., Contegiacomo, M., Carbonara, P., Lembo, G., Boglione, C., 1998. Riproduzione artificiale di *Epinephelus marginatus* orientata allo sviluppo delle tecniche di restocking. Biol. Mar. Mediterr. 5, 1248–1257.
Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER, Version 2.2.3. Department of Biological Sciences. University of Hull, Hull, U.K.
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 current help explain population genetic structure. Proc. R. Soc. Lond. B 277 (1688), 1685–1694.
Wright, S., 1969. Evolution and the Genetics of Populations. University of Chicago Press, Chicago, IL.
Yue, G.H., David, L., Orban, L., 2007. Mutation rate and pattern of microsatellites in common carp (*Cyprinus carpio* L.). Genetica 129, 329–331.
Zabala, M., Garcia-Rubies, A., Louisy, P., Sala, E., 1997. Spawning behaviour of the Mediterranean dusky grouper *Epinephelus marginatus* (Lowe, 1834) (Pisces, Serranidae) in the Medes Islands Marine Reserve (NW Mediterranean, Spain). Sci. Mar. 61, 65–77.
Zardoya, R., Castilho, R., Grande, C., Favre-Krey, L., Caetano, S., Marcato, S., Krey, G., Patarnello, T., 2004. Differential population structuring of two closely related fish species, the mackerel (*Scomber scombrus*) and the chub mackerel (*Scomber japonicus*), in the Mediterranean Sea. Mol. Ecol. 13, 1785–1798.

Supplementary material Table S1: Microsatellite loci used in *Epinephelus marginatus*, the repeat type and the exact forward and reverse sequence of the primers, the allele size range in base pairs. The loci amplified simultaneously are identified with the same multiplex number.

locus	repeat	primer sequence (5'-3')	size(bp)	Multiplex
D076	$(TCTA)_{12}$	**F** : 6FAM-ACCCCGTCCTCCATTAAGTC	350-424	1
		R : CCGAGCCATGGAAGAATTTA		
GAG007	$(TG)_8$	**F** : 6FAM-CTGTAATAGACAACCCACTG AC	141-151	1
		R : CCTGTAGCATCTTCACTAGCTG		
GAG038	$(CA)_{28}$	**F** : VIC-CCCCACCTC CCT TAA CA	71-131	1
		R : GCT GAA TTG AGG AAA TGA G		
GAG045	$(GT)_{14}$	**F** : 6FAM-GTG TGC ATG TGA GAG AAA GT	72-130	1
		R : GCC TTA ACG GAT GTC TTT CT		
GAG049	$(GT)_{24}$	**F** : 6FAM-ACT CTA ATC TAC AGC ATA TTC T	88-130	2
		R : CAG CTC GCC TGA AAG ACT		
RH-CA-7	$(GAT)_{12}$	**F** : VIC-CAG AAA CAT CTC CCC CAA AA	306-348	2
		R : CTG GCA GAG CAA TTA GAG GC		
RH-CA-1	$(CA)_{19}$	**F** : VIC-CGA GAT AAG CCC TGG TGA AA	371-413	2
		R : AGT CCC GAT GTG GTA ACG AG		
RH-CA3	$(CA)_{16}$	**F** : PET-ATA CTG CAC ACA ACC CAC CA	349-389	2
		R : ACA CGC GGT GTT TTA GAG GT		
RH-CA-4	$(CA)_5N(CA)_8$	**F** : PET-GAG AAC GAC ATT CCA GCA CA	205-237	3
		R : TGT GTG ACC AGA AAC CAG GA		
RH-CA-8	$(CA)_{21}$	**F** : 6FAM-AGT TGC CCA GGT TAC ACG AG	201-247	3
		R : TTG GGT CCT GGC ATT TAG AG		
SC06	(CA)12	**F** : NED-AAA AGA GGC AGT GAA GAA TTG G	190-226	3
		R : TCA TCC ATT TCC CTG TTT CA		

Supplementary material Table S2: Summary statistics of the 11 microsatellites in each of the 10 analysed populations of *Epinephelus marginatus*

Pop		**D076**	**GAG007**	**GAG038**	**GAG045**	**GAG049**	**RHCA7**	**RHCA1**	**RHCA3**	**RHCA4**	**RHCA8**	**SC06**
Gre	**N**	38	40	37	40	40	38	40	40	40	40	40
	Na	13	3	20	20	12	10	11	2	10	18	15
	Ho	0,842	0,725	0,946	0,800	0,750	0,763	0,700	0,025	0,775	0,850	0,800
	He	0,842	0,490	0,883	0,852	0,640	0,729	0,755	0,025	0,804	0,836	0,810
	F_{IS}	0,000	-0,481	-0,071	0,061	-0,172	-0,047	0,073	-0,013	0,036	-0,017	0,012
Cor	**N**	32	32	32	32	32	32	31	31	32	32	31
	Na	13	3	16	17	10	6	12	4	10	15	15
	Ho	0,906	0,438	0,875	0,844	0,688	0,781	0,871	0,065	0,906	0,750	0,903
	He	0,854	0,514	0,840	0,866	0,699	0,747	0,848	0,123	0,811	0,768	0,874
	F_{IS}	-0,061	0,149	-0,041	0,026	0,016	-0,046	-0,028	0,477	-0,118	0,024	-0,033
Mal	**N**	40	40	40	40	40	40	40	40	40	38	40
	Na	12	3	21	19	7	6	13	4	10	17	17

	Ho	0,725	0,375	0,775	0,800	0,550	0,600	0,800	0,100	0,775	0,789	0,950
	He	0,802	0,463	0,862	0,869	0,653	0,703	0,787	0,097	0,776	0,726	0,900
	F_{IS}	0,096	0,191	0,101	0,080	0,157	0,147	-0,017	-0,036	0,001	-0,088	-0,056
Men	**N**	40	40	40	40	40	40	40	39	40	40	39
	Na	12	2	17	24	9	8	13	2	9	17	16
	Ho	0,675	0,525	0,875	0,800	0,725	0,850	0,700	0,051	0,875	0,900	0,795
	He	0,829	0,475	0,823	0,889	0,712	0,728	0,809	0,050	0,793	0,837	0,904
	F_{IS}	**0,186**	-0,106	-0,064	0,100	-0,019	-0,168	0,135	-0,026	-0,103	-0,076	**0,120**
Med	**N**	19	19	19	14	19	19	19	20	19	20	20
	Na	11	3	15	12	7	7	11	2	7	14	12
	Ho	0,737	0,737	0,842	0,929	0,684	0,789	0,947	0,150	0,789	0,900	0,950
	He	0,817	0,548	0,866	0,855	0,647	0,748	0,855	0,139	0,785	0,839	0,861
	F_{IS}	0,098	-0,343	0,027	-0,087	-0,058	-0,056	-0,109	-0,081	-0,005	-0,073	-0,103
Bcn	**N**	40	40	40	40	40	40	39	40	40	40	40
	Na	11	2	18	22	9	8	12	3	12	14	17
	Ho	0,750	0,400	0,775	0,875	0,775	0,625	0,897	0,100	0,800	0,775	0,750
	He	0,803	0,480	0,802	0,883	0,658	0,738	0,789	0,096	0,819	0,709	0,876
	F_{IS}	0,065	0,167	0,033	0,010	-0,177	0,153	-0,137	-0,042	0,023	-0,093	0,144
Cga	**N**	33	33	32	33	33	33	33	33	33	33	33
	Na	13	3	18	19	10	9	13	3	10	16	16
	Ho	0,879	0,606	0,969	0,818	0,727	0,667	0,879	0,030	0,667	0,970	0,879
	He	0,854	0,495	0,885	0,859	0,659	0,756	0,813	0,088	0,805	0,867	0,869
	F_{IS}	-0,028	-0,223	-0,095	0,048	-0,103	0,118	-0,081	0,654	0,172	-0,119	-0,011
Alg	**N**	33	37	37	36	37	37	35	37	35	36	36
	Na	11	3	18	21	9	7	11	2	9	16	15
	Ho	0,848	0,730	0,784	0,944	0,757	0,757	0,943	0,027	0,857	0,806	0,833
	He	0,811	0,511	0,763	0,909	0,694	0,749	0,805	0,027	0,782	0,804	0,874
	F_{IS}	-0,046	-0,428	-0,028	-0,039	-0,090	-0,010	-0,171	-0,014	-0,097	-0,002	0,047
Tun	**N**	40	40	40	35	40	40	39	39	40	40	38
	Na	14	4	15	21	10	7	10	3	11	17	14
	Ho	0,900	0,650	0,875	0,857	0,675	0,700	0,821	0,128	0,700	0,850	0,895
	He	0,844	0,552	0,879	0,890	0,622	0,739	0,753	0,122	0,805	0,790	0,872
	F_{IS}	-0,066	-0,178	0,005	0,037	-0,085	0,052	-0,089	-0,051	0,131	-0,076	-0,027
Sen	**N**	39	40	40	40	40	40	40	40	40	40	40
	Na	14	3	20	20	10	8	16	4	15	22	18
	Ho	0,923	0,575	0,825	0,925	0,675	0,475	0,975	0,225	0,725	0,850	0,750
	He	0,867	0,455	0,816	0,883	0,699	0,722	0,899	0,250	0,868	0,918	0,908
	F_{IS}	-0,064	-0,263	-0,011	-0,047	0,035	**0,342**	-0,084	0,099	0,165	0,075	**0,174**

N: Number of individuals, Na: Number of alleles, Ho: Observed Heterozygozity, He: Expected Heterozygozity, F_{IS}: Fixation Index with indication of significance after False discovery rate correction in bold ($p<0.009$).

Supplementary material Fig.S1:

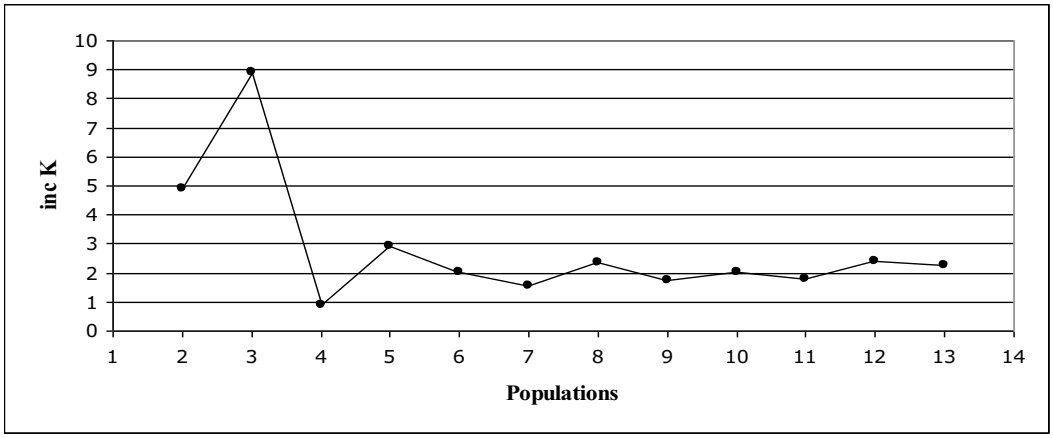

Fig. S1: The increment of K for the possible numbers of genetically different populations after applying the ad hoc by Evanno et al. (2005).

Supplementary material Fig.S2:

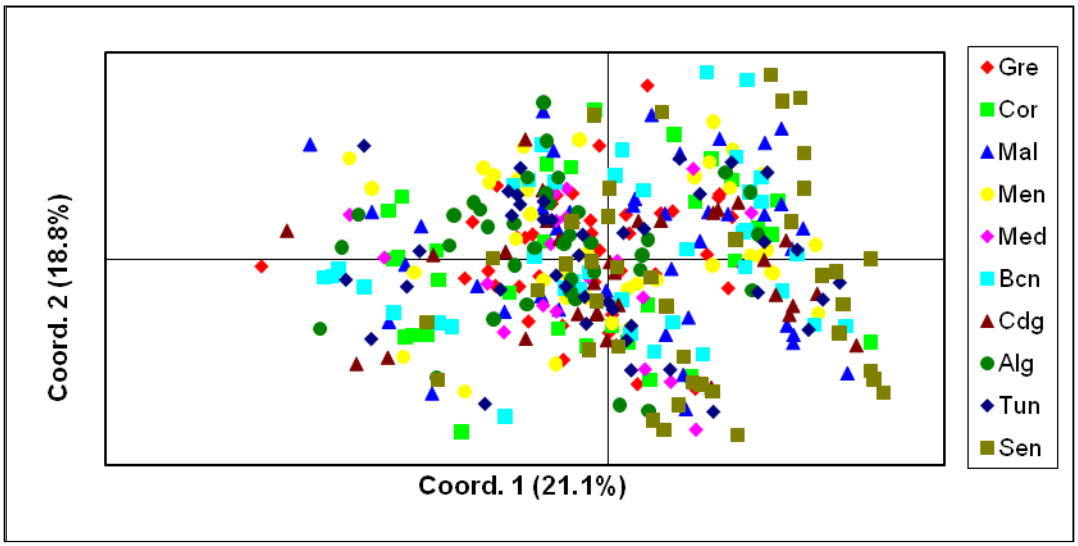

Fig. 2: Principal coordinates analysis of all individuals calculated with GenAlEx.

3.1 Genetic population connectivity and oceanographic fronts in littoral fish species

3.1 Genetic population connectivity and oceanographic fronts in littoral fish species

Publication 2: Matching genetics with oceanography: directional gene flow in a Mediterranean fish species

Publicación 2: Coincidencia entre la genética y la oceanografía: direccionalidad en el flujo de genes en un pez Mediterráneo

RESUMEN

La conectividad genética y la fragmentación geográfica son dos mecanismos opuestos que determinan la estructura poblacional de las especies. Mientras el primero homogeniza las poblaciones, el segundo permite su diferenciación. Por lo tanto, el conocimiento sobre los procesos que afectan a la dispersión de organismos marinos es crucial para entender sus patrones de distribución genética y para la gestión efectiva de sus poblaciones. En este estudio utilizamos análisis genéticos de once microsatélites en combinación con datos oceanográficos obtenidos por satélite y simulaciones de dispersión para determinar los patrones de distribución de *Serranus cabrilla*, un pez demersal que se reproduce ampliamente en el Mediterráneo. Los valores de FST por pares de poblaciones oscilan entre 0,003 y 0,135. Se identificaron dos grupos genéticamente distintos con una división clara localizada entre las discontinuidades oceanográficas del canal de Ibiza (IC) y el frente Almería-Oran (AOF), revelando una población genéticamente mezclada entre ambos. El frente balear (BF) también parece influir en la estructura poblacional. En la costa española se observó un flujo génico direccional en *S. cabrilla* con una dispersión del oeste al este a través del AOF, del norte hacia el sur en el IC y desde el sur del IC hacia las Islas Baleares. Las correlaciones entre los datos genéticos y oceanográficos fueron altamente significativas. Los cambios estacionales en los patrones de corrientes y la relación entre los patrones de circulación del océano y la época de reproducción pueden jugar un papel importante en la estructura poblacional alrededor de los frentes oceanográficos.

MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 5167–5181 doi: 10.1111/j.1365-294X.2011.05355.x

Matching genetics with oceanography: directional gene flow in a Mediterranean fish species

C. SCHUNTER,*† J. CARRERAS-CARBONELL,*† E. MACPHERSON,* J. TINTORÉ,‡ E. VIDAL-VIJANDE,‡ A. PASCUAL,‡ P. GUIDETTI§ and M. PASCUAL†

*Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Car. Acc. Cala St. Francesc 14, Blanes, 17300 Girona, Spain, †Department Genètica, Univ. Barcelona, 08028 Barcelona, Spain, ‡IMEDEA (CSIC-UIB), C/Miquel Marques 21, 07190 Esporles, Mallorca, Spain, §University of Salento, Department of Biological and Environmental Sciences, via Prov. Monteroni, 73100 Lecce, Italy

Abstract

Genetic connectivity and geographic fragmentation are two opposing mechanisms determining the population structure of species. While the first homogenizes the genetic background across populations the second one allows their differentiation. Therefore, knowledge of processes affecting dispersal of marine organisms is crucial to understand their genetic distribution patterns and for the effective management of their populations. In this study, we use genetic analyses of eleven microsatellites in combination with oceanographic satellite and dispersal simulation data to determine distribution patterns for *Serranus cabrilla*, a ubiquitous demersal broadcast spawner, in the Mediterranean Sea. Pairwise population F_{ST} values ranged between −0.003 and 0.135. Two genetically distinct clusters were identified, with a clear division located between the oceanographic discontinuities at the Ibiza Channel (IC) and the Almeria-Oran Front (AOF), revealing an admixed population in between. The Balearic Front (BF) also appeared to dictate population structure. Directional gene flow on the Spanish coast was observed as *S. cabrilla* dispersed from west to east over the AOF, from north to south on the IC and from south of the IC towards the Balearic Islands. Correlations between genetic and oceanographic data were highly significant. Seasonal changes in current patterns and the relationship between ocean circulation patterns and spawning season may also play an important role in population structure around oceanographic fronts.

Keywords: connectivity, microsatellites, population structure, seascape genetics, *Serranus*, surface currents

Received 17 December 2010; revised received 8 September 2011; accepted 30 September 2011

Introduction

Population connectivity is driven by the dispersal of individuals (Palumbi 2003), and it is generally assumed that for marine organisms pelagic stages represent the main vector of genetic exchange because of their high potential for dispersal (Galarza *et al.* 2009a). Seventy per cent of all marine life has a pelagic larval phase in their life cycle (Pinet 2009), but difficulties in studying early life stages still limit the understanding of how connectivity works within the marine environment. The high effort of measuring larval movement in the field, owing to the r-selection strategy of most marine organisms by producing hundreds and thousands of offspring in one brood or spawning event, is one of the first hurdles when analysing larval dispersal. Furthermore, larvae are extremely small in size (200 μm–20 mm, Levin 2006), which makes tagging studies problematic or highly elaborate (Thorrold *et al.* 2006). Moreover, especially for the pelagic larval phase biological features are closely connected with physical processes (Levinton 2001) such as winds and currents, which affect the passive transport of early-life stages (Leis 2007).

In marine waters, there is an apparent continuity potentially allowing larvae of marine organisms to be

Correspondence: C. Schunter, Fax: +34 972 33 78 06;
E-mail: cschunter@ceab.csic.es

dispersed over large spatial scales. In such cases, we would expect genetic isolation only by distance (Kimura 1953; Wright 1969), meaning that genetic differentiation increases because of increasing geographic distance. However, numerous marine species do not follow this stepping stone model of isolation by distance (Bradbury & Bentzen 2007), and recent genetic studies suggest that the extent of dispersal and gene flow between populations is more complex than previously assumed (Calderon *et al.* 2007; Carreras-Carbonell *et al.* 2007; Uriz *et al.* 2008; Planes *et al.* 2009). Moreover, small-scale genetic structuring is a common phenomenon even for highly mobile species with long larval phases such as cod (*Gadus morhua*) and spiny lobster (*Palinurus elephas*) (Knutsen *et al.* 2003; Palero *et al.* 2008).

Recent studies indicate that oceanographic processes and barriers to dispersal may greatly influence or even determine the connectivity of marine populations (Johansson *et al.* 2008; Galarza *et al.* 2009a; Selkoe *et al.* 2010; White *et al.* 2010). Phylogeographic or speciation events can be measured by studying ancient breaks or past climate patterns (Palero *et al.* 2008); however, the study of current genetic exchange patterns is essential especially in the presence of ever-increasing anthropogenic pressure (Saenz-Agudelo *et al.* 2009) and requires the analysis of present-day oceanographic processes (Gilg & Hilbish 2003). Fortunately, studies coupling oceanographic and genetic data are becoming increasingly popular (Galindo *et al.* 2006; Selkoe *et al.* 2006; Banks *et al.* 2007; Galarza *et al.* 2009a). Coastal and ocean circulations, as well as eddies and current discontinuities, have been demonstrated to strongly affect population connectivity (White *et al.* 2010; Schunter *et al.* 2011). Nevertheless, there is still limited knowledge on how oceanographic conditions may pose barriers for the population structure of a species and especially why different species with similar life-history traits show distinct responses to oceanographic discontinuities (Bargelloni *et al.* 2008; Galarza *et al.* 2009a).

Plasticity of oceanographic processes presents a key challenge in understanding and predicting larval distribution patterns. Numerous fronts can be seasonal in intensity and even change direction in which the driving current flows (Astraldi *et al.* 1995). Neumann (1968) described how ocean currents may vary in terms of speed and direction, which highlights the difficulty in determining a general gene flow direction. Such current fluctuations can greatly influence recruitment and the genetic exchange between populations (Shanks & Eckert 2005; Stenseth *et al.* 2006). In particular, seasonal shifts in current direction can result in seasonal variation in larval transport. Thus, two groups of offspring from one same location may follow highly varied distribution patterns depending on the time of spawning. Hence, combining information of temporal and directional variation at different spatial scales in oceanographic patterns is of great interest and importance in the investigation of connectivity patterns.

The Mediterranean Sea is an ideal study area for a survey incorporating oceanographic features and gene flow. The circulation patterns within the Mediterranean Sea have been subject to studies for many decades and are well described (Fig. 1, e.g. Millot 1999; Fernandez *et al.* 2005; Rio *et al.* 2007). Moreover, several oceanographic discontinuities, mostly on the Spanish coast, originated by the entry of less saline Atlantic waters in the Mediterranean Sea, have been studied at the population genetic level in different marine organisms. The oceanographic processes occurring off the Gibraltar Strait has been demonstrated to act as barriers to gene flow for two fish species (e.g. Galarza *et al.* 2009b; Sala-Bozano *et al.* 2009). The most well-studied and quasi-permanent Almeria-Oran Front (AOF) has been proposed as the point of genetic break between the Atlantic

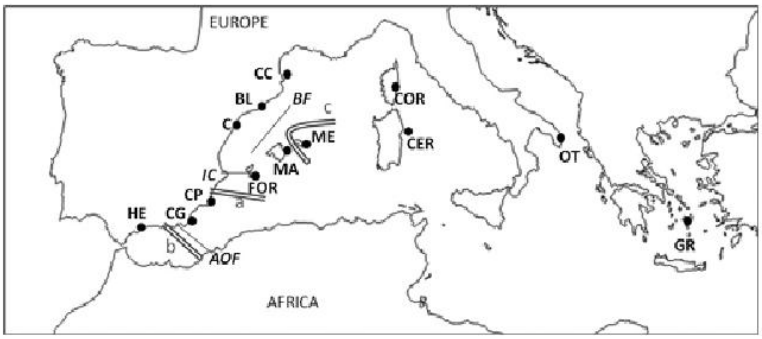

Fig. 1 Map of sample locations across the Mediterranean Sea showing putative barriers to gene flow (grey lines): Almeria-Oran Front (AOF), Balearic Front (BF), Ibiza Channel (IC). Inferred genetic barriers identified in the BARRIER analysis are plotted with a double line and are labelled a, b, c with a being the strongest. CC, Cap de Creus; BL, Blanes; C, Columbretes; ME, Menorca; MA, Mallorca; FOR, Formentera; CP, Cabo de Palos; CG, Cabo de Gata; CER, Sardinia; COR, Corsica; OT, Otranto; GR, Greece.

Ocean and the Mediterranean Sea in many species of different taxa including seaweeds, sponges, molluscs, crustaceans, fish and mammals (see the review in Patarnello *et al.* 2007). However, only few genetic studies sampled populations at either side of the front (Galarza *et al.* 2009a), and thus, the genetic break inferred can include several oceanographic discontinuities as well as genetic differentiation because of isolation by distance. The Balearic Front (BF), a temporal front, has only recently been studied, but has shown to be a strong barrier for genetic exchange in littoral fish species (Galarza *et al.* 2009a). Besides, the Ibiza Channel (IC) has rarely been considered in genetic population studies of marine species, but see the genetic study in red gorgonian by Mokhtar-Jamaï *et al.* (2011), although the oceanographic patterns and temporal processes are well known (Fernandez *et al.* 2005; Monserrat *et al.* 2008). Other oceanographic processes occurring within the Mediterranean Sea, for example along the Sicily Channel and the Sardinian Channel, can also act as barriers to gene flow such as unveiled in prawns and seaweeds (Zitari-Chatti *et al.* 2009; Serra *et al.* 2010).

Our study species *Serranus cabrilla*, also called the comber, is a common demersal fish inhabiting the Eastern Atlantic Ocean, and the Mediterranean and Black Seas. It inhabits sea-grass beds, and is found on rocky, sandy and muddy bottoms between 5 and 100 m deep. The comber is considered one of the most important predators of early fish stages and vagile invertebrates (Guidetti & Cattaneo-Vietti 2002). Combers are economically relevant and included in the Food and Agricultural Organization of the United Nations (FAO) catalogues as species of interest to fisheries in the Eastern Atlantic, the Mediterranean and the Black Sea. Much is known about its ecology and biology (Torcu-Koc *et al.* 2004); however, there is no information about its population structure and degree of connectivity between populations. Larvae of comber remain in the plankton stage for 21–28 days (Raventós & Macpherson 2001; Macpherson & Raventós 2006) and have been collected inshore and over the continental shelf at a considerable distance (50 km offshore) from the habitats of the adults (Sabatés 1990). This implies that comber larva may have a wide dispersal potential to maintain high connectivity between populations (Planes 2002).

The main aim of this study is to analyse the effects of oceanographic processes on the population structure of fish species. More specifically, we evaluate the connectivity pattern of the comber *S. cabrilla* with genetic markers and identify different genetic units as well as direction of gene flow. With the help of oceanographic satellite and dispersal simulation data, we analyse the flow of particles to assess the influence of predominant current patterns on the genetic structure of *S. cabrilla*. Finally, we discuss the seasonal and directional effects of oceanographic processes on the genetic connectivity of the species.

Materials and Methods

Sampling and DNA extraction

Thirteen locations were sampled to investigate the genetic structure of *S. cabrilla* within the Mediterranean Sea. Two locations from the East-Mediterranean basin [Greece (GR), $n = 22$ and Otranto (OT), $n = 30$] and two island localities from the central part of the Mediterranean basin [Sardinia (CER), $n = 30$ and Corsica (COR), $n = 30$] were chosen. More intensive sampling was carried out along the Spanish coast, because there is more information on oceanographic processes in this area. To examine the effects of oceanographic discontinuities on the gene-flow patterns, nine coastal and island locations were selected along the Spanish coast: Cap de Creus (CC) ($n = 30$), Blanes (BL) (30), Columbretes (C) (25), Mallorca (MA) (30), Menorca (ME) (30), Formentera (FOR) (30), Cabo de Palos (CP) (29), north of Cabo de Gata (CG) (31) and Herradura (HE) (35) (Fig. 1, Table 1). A total of 382 specimens were collected in the field by hook and line or spear gun, and CG samples were bought from local fishermen at the sampling location. All specimens were collected in 2003 with the

Table 1 Summary statistics of all localities

Population	N	Allelic richness	H_e	H_o	F_{IS}	Cluster 1	Cluster 2
GR	22	8.10	0.65	0.64	**0.015**	21	1
OT	30	8.43	0.69	0.62	**0.104**	30	0
CER	30	8.35	0.69	0.65	**0.047**	28	2
COR	30	8.27	0.69	0.66	0.041	30	0
CC	30	8.10	0.70	0.67	**0.045**	30	0
BL	30	7.49	0.67	0.60	**0.112**	28	2
C	25	8.03	0.69	0.64	**0.075**	24	1
ME	30	8.37	0.73	0.73	**−0.002**	27	3
MA	30	8.79	0.73	0.72	**0.016**	25	5
FOR	30	8.75	0.72	0.71	0.015	27	3
CP	29	9.03	0.77	0.74	0.039	8	21
CG	31	8.62	0.79	0.70	**0.110**	5	26
HE	35	8.04	0.76	0.73	**0.048**	0	35

CC, Cap de Creus; BL, Blanes; C, Columbretes; ME, Menorca; MA, Mallorca; FOR, Formentera; CP, Cabo de Palos; CG, Cabo de Gata; CER, Sardinia; COR, Corsica; OT, Otranto; GR, Greece, HE, Herradura; N, Number of individuals; H_o, Observed Heterozygozity; H_e, Expected Heterozygozity; F_{IS}, Fixation Index (significant values in bold after FDR ($P < 0.0157$)); Cluster 1 and Cluster 2, number of individuals assigned by STRUCTURE to each cluster with over 75%. Population abbreviations are as in Fig. 1.

exception of HE collected in 2005. The sampled fish were young adults (total length 15–20 cm), sexually active and therefore represent the young spawning population. Pectoral fin clips were removed and preserved in 100% ethanol at room temperature. Total genomic DNA was extracted from fin tissue using the Chelex 10% protocol (Estoup *et al.* 1996).

PCR amplification and screening

Eleven microsatellite loci previously isolated (Carreras-Carbonell *et al.* 2006a) were used in the analysis of all 13 populations (Supporting Information). Polymerase chain reactions were carried out under conditions described in Carreras-Carbonell *et al.* (2006a). Amplified products were scored using an ABI 3700 automatic sequencer from the Scientific and Technical Services of the University of Barcelona. Alleles were sized by GENEMAPPER™ software, with an internal size marker CST Rox 70-500 (BioVentures Inc.).

Statistical analyses

Allele frequencies, expected (H_E) and observed (H_O) heterozygosity per locus and population, were calculated using GeneAlEx 6.1 (Peakall & Smouse 2006), and standardized allelic richness was determined by FSTAT (Goudet 2002) (see Table S1, Supporting Information). Departures from the Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium were tested for each locus-population combination using GENEPOP version 4 (Rousset 2008) which employs a Markov chain method, with 5000 iterations, following the algorithm of Guo & Thompson (1992). The program MICRO-CHECKER (Van Oosterhout *et al.* 2004) was used to infer the most probable cause of HWE departures.

Genetic divergence between populations was estimated by computing the classical F_{ST} approach and the log likelihood ratio (G) tests for population differentiation was assessed by Markov chain algorithms as implemented in GENEPOP (Wright 1969; Weir & Cockerham 1984; Goudet *et al.* 1996). Also, the more recent heterozygosity-independent Jost *D* (Jost 2008), which has been shown to reliably reflect genetic differentiation, was calculated with DEMEtics (Gerlach *et al.* 2010), which uses a bootstrap method (1000 bootstrap repeats) after Manly (1997) to estimate *P*-values. To correct for possible type I errors when performing multiple tests, we employed a false discovery rate (FDR) approach described in Benjamini & Yekutieli (2001). With this approach, the cut-off of the significance value is fixed by the level at which to control FDR. The FDR can be viewed as the fraction of false positives among all tests that are declared significant. This approach has been stated to provide the critical value most relevant for biological questions (Narum 2006).

The correlation between pairwise multilocus distances ($F_{ST}/(1 - F_{ST})$) and geographical distance (Ln distance) of all locations and of only the Spanish localities was assessed using the Mantel permutation test (10 000 permutations; Mantel 1967) implemented in GENEPOP. The geographical distance in kilometres was computed as the coastline distance between continental sample locations and as the straight geographical distance for island populations. Putative barriers to gene flow were detected with BARRIER by using F_{ST} (Manni *et al.* 2004). Three barriers were chosen as there are three well-known oceanographic barriers in the sampling area on the Spanish coast. Separate tests were performed with each loci and all loci together to assure that the patterns were not driven by only one or few loci.

The program STRUCTURE 2.3.3. (Pritchard *et al.* 2000) was used to detect the number of genetically differentiated populations (*K*). The population structure was considered without prior information of the number of locations at which the individuals were sampled and to which location each individual belongs. Following recommendations from Evanno *et al.* (2005), we calculated an *ad hoc* statistic ΔK based on the rate of change in the log probability of data between successive *K* values. Twenty runs and 200 000 Markov Chain Monte Carlo (MCMC) were carried out in order to quantify the standard deviation (SD) of the likelihood of each *K* with a range of K_s between 1 and 14. Furthermore, we plotted the log probability of the data ($\mathrm{Ln}P(D)$) as a function of *K* across the 20 runs and looked for the value that captured the major structure in the data (Pritchard *et al.* 2000). Structure was run with different data subsets to identify a hierarchical structure: all populations, all populations except HE, only CP, CG and HE and finally all populations except CP, CG and HE. Also, a box plot of *Q* (estimated membership of each individual to a cluster) was produced with STRUCTURE (Pritchard *et al.* 2000) indicating the possible origin for each individual. Single individuals showing a *Q* of more than 0.75 to a different cluster were identified as putative migrants.

Furthermore, a pairwise relatedness coefficient r_{xy} was computed with all the samples from all localities to receive a mean relatedness value per location. This was carried out by using the Queller & Goodnight (1989) approach which is implemented in the GenAlex software (Peakall & Smouse 2006). Population means and 95% Confidence Intervals were determined by 10 000 bootstrap replicates. The 95% CI of the null hypothesis of 'no difference' across populations was established by 9999 permutations in GenAlex. To identify population relationship in a two-dimensional space, Principle Coordinate Analyses (PCoA) of the localities were computed

and graphed with GenAlEx (Peakall & Smouse 2006). PCoA was performed using the F-statistics and the Jost *D* values.

Recent migration patterns along the Spanish coast were established by assignment tests run in GENECLASS 2.0 (Piry *et al.* 2004) and migration rate estimates in BAYESASS (Wilson & Rannala 2003). In GENECLASS 2.0, the probability of an individual belonging to a certain locality is calculated using a Bayesian method (Rannala & Mountain 1997). All assignment probabilities above 0.8 were accepted as correctly assigned and used in the migration analysis. BAYESASS uses a non-equilibrium Bayesian method to estimate recent migration rates. Default settings were used in the program and the average across three runs was used. Migration rate analyses were carried out for all Spanish localities, but joining CC and BL together as the North Spanish Coast (NSC) because locations are very close and particle simulations could only be carried out for the area including both sites. Furthermore, these two locations were not genetically differentiated (see below in Results section) as reported in other fish species (Carreras-Carbonell *et al.* 2006b).

Oceanographic General Circulation Data

Numerical simulations of particle dispersion were carried out using the Mediterranean Forecasting System (MFS) model (Tonani *et al.* 2008), characterized by the highest horizontal and vertical resolution presently available for the Mediterranean Sea: $1/16° \times 1/16°$ in the horizontal (~6.5 km) and 72 vertical levels. The model can be therefore defined as a mesoscale-resolving model for the Mediterranean Sea, as the first internal Rossby radius is ~15 km (Robinson *et al.* 2001). In order to improve the simulations, the MFS system assimilates temperature and salinity vertical profiles from eXpandableBathyThermograph (XBT) and Argo, and Sea Level Anomalies from satellite altimetry (Tonani *et al.* 2009). Daily forecast of a wide range of physical variables such as temperature, salinity, density and currents are provided (http://gnoo.bo.ingv.it/mfs/). In this study, independent satellite altimeter gridded fields (Pascual *et al.* 2007) were also used for the analysis of circulation patterns in the Mediterranean Sea. However, considering that the barriers to gene flow were only detected along the Spanish coast (see below in Results section), the simulations were only carried out along this area. Simulated surface dispersion fields are presented using numerical particles (proxy of larvae) released at specific key points at the beginning of the *S. cabrilla* spawning season (April) until the end of the larval period (3 months later) at eight different sites on the Spanish coast. However, because particles arriving at the land boundary of the model are eliminated, the release area was not placed too close to the coast. This can result in some near-shore currents not being represented in the dispersion model.

The simulations were performed for April, May and June 2001 and 2004, as most of the collected comber samples would have been larvae in those years. The particles were released in 50 km × 50 km squares at 5 km separation on the 1st day of each month and allowed to drift during 30 days with the daily currents provided by the simulation. Square matrices for each month were computed quantifying the number of particles recorded at any of the eight studied localities during the 30 days drift. Thus, one particle can be recorded at several sites. Finally, one matrix for 2001 and another for 2004 were generated by adding the matrices of April, May and June of each year. We also ran the simulations changing the initial distance separation of particles between 1 km (7803 particles) and 5 km (363 particles). Matrices obtained were then standardized by particle number and could thus be compared. The standard deviation of the different simulations computed over all the elements of the matrices were lower than 1% both for 2001 (0.22%) and 2004 (0.88%). This shows low sensitivity of the model to small changes in parameters, which in turn demonstrates that the method is robust and nondependant of the number of particles. Here, we present the data obtained from the simulation with 5 km distance, as the graphs can be more clearly interpreted and the parameter of 5 km distance adjusts better to the original resolution of the model used.

These data were then correlated with the genetic migration data calculated with GENECLASS (Piry *et al.* 2004) and BAYESASS (Wilson & Rannala 2003) with a nonparametric Spearman's rho test in SPSS. To assure that our results were not only encountered due to chance, the correlation was repeated after the randomization of the data.

Results

Genetic variability

High genetic variability was found in *S. cabrilla* in terms of extensive polymorphism per population and locus (mean allelic richness 8.3 ± 0.4), as well as high expected (0.712 ± 0.04) and observed (0.676 ± 0.04) heterozygosities (Table 1, Table S1, Supporting Information). No linkage disequilibrium between loci was observed in any of the populations; thus, the eleven loci were considered statistically independent. Private alleles were present in all populations, with a mean percentage per population of 2.88 ± 0.58%.

Significant departures from HWE were observed in most localities, with the exception of COR, C, FOR and CP (Table 1), which could be a result of selection, Wahlund effects or null alleles. When all loci where analysed separately, departures from HWE was caused mainly by locus Sc05. For this locus, null alleles were detected with MICRO-CHECKER, as well as sporadically for some loci in different localities (Table S1, Supporting Information). Null alleles appear when one allele is unamplified because of mutations in the sequence where one of the primers was designed, or when technical problems associated with amplification and scoring arise (Hoarau *et al.* 2002). Technical issues could be ruled out because all failed amplifications for loci Sc05, Sc06, Sc07, Sc08 and Sc14 were reamplified twice, lowering the annealing temperature to 50°C, verifying that the non-amplified individuals were homozygotes for null alleles. All analyses were carried out including the Sc05 locus as well as without the locus and results were almost identical. Therefore, results presented in this study include the Sc05 locus, as assignment tests have a better performance with larger number of loci (Carreras-Carbonell *et al.* 2006b).

Population differentiation

The global multilocus F_{ST} revealed significant genetic differentiation ($F_{ST\ global} = 0.032$, $P < 0.001$). Pairwise F_{ST} values ranged between −0.004 and 0.135 (Table 2). Different degrees of structuring were found between all 13 populations of *S. cabrilla*. HE was the most genetically differentiated from all other localities. Other two locations with more proximity to Atlantic waters, CP and CG, were also significantly differentiated to all other locations, but revealing a close relationship between them. The rest of the populations with greater influence from the Mediterranean Sea were only weakly genetically differentiated. Some population pairs show relatively high F_{ST} values (for instance GR-MA F_{ST}: 0.013; C-MA F_{ST}: 0.011, see Table 2) that are not significant, which could suggest a limitation of the power because of lower sample sizes in GR and C. The Jost *D* values showed similar results to the F_{ST} values and ranged between −0.018 and 0.291 however seem to be less sensitive to population sizes (Table 2).

No significant association between genetic differentiation (F_{ST} or *D*) and geographic distance was revealed by a Mantel test ($P = 0.116$ and $P = 0.147$ respectively) when all locations were included, however, the Mantel test was significant when only Spanish locations were considered (F_{ST}: $P < 0.01$, *D*: $P < 0.01$) (Fig. S1, Supporting Information).

Putative barriers to gene flow were computed with BARRIER and all loci and are represented in Fig. 1. The strongest barrier (a) with a distance value produced by BARRIER of 0.03 was placed between CP and FOR and C, which is situated at the IC. The second barrier (b) would be situated at the AOF (distance value of 0.013), and the third barrier (c) separates ME from the other localities (distance value of 0.01). All barriers were supported abundantly by the different loci: Barrier a was present in 10 of the 11 loci; Barrier b was represented by eight loci, and Barrier c was shown by seven loci.

Two genetically differentiated clusters were detected when STRUCTURE was computed with all samples, as the peak in ΔK was for $K = 2$. The height of ΔK was used as an indicator of the strength of the signal ($\Delta K = 129.1$

Table 2 Multilocus JOST *D* distances between population pairs below the diagonal and F_{ST} values above the diagonal

	GR	OT	CER	COR	CC	BL	C	ME	MA	FOR	CP	CG	HE
GR	–	0.002	−0.003	0.003	**0.011**	0.006	0.003	**0.018**	0.013	0.005	**0.058**	**0.069**	**0.135**
OT	0.023	–	0.000	0.000	**0.013**	**0.012**	**0.011**	**0.012**	**0.004**	0.000	**0.050**	**0.056**	**0.122**
CER	−0.010	0.008	–	−0.004	**0.005**	0.000	0.001	**0.009**	0.002	−0.003	**0.040**	**0.049**	**0.110**
COR	0.013	0.005	0.001	–	**0.006**	0.002	0.009	0.007	0.001	−0.001	**0.042**	**0.049**	**0.111**
CC	0.024	**0.049**	0.006	0.040	–	−0.002	**0.003**	**0.006**	**0.009**	**0.008**	**0.042**	**0.055**	**0.100**
BL	0.023	**0.048**	0.003	0.016	−0.013	–	**0.003**	**0.009**	0.008	**0.007**	**0.040**	**0.053**	**0.105**
C	0.011	**0.051**	0.008	0.038	0.020	0.015	–	**0.015**	0.011	**0.009**	**0.044**	**0.053**	**0.108**
ME	0.040	0.030	0.017	0.029	0.011	0.037	**0.057**	–	0.008	0.008	**0.027**	**0.038**	**0.085**
MA	**0.049**	0.030	0.010	0.021	0.031	**0.042**	**0.044**	0.027	–	−0.004	**0.031**	**0.032**	**0.088**
FOR	0.005	−0.005	−0.018	0.004	0.019	0.025	0.027	0.029	−0.012	–	**0.028**	**0.031**	**0.088**
CP	**0.109**	**0.122**	**0.083**	**0.097**	**0.105**	**0.095**	**0.098**	**0.078**	**0.084**	**0.071**	–	−0.001	**0.022**
CG	**0.145**	**0.136**	**0.103**	**0.131**	**0.152**	**0.133**	**0.108**	**0.132**	**0.079**	**0.072**	0.001	–	**0.013**
HE	**0.276**	**0.291**	**0.237**	**0.277**	**0.240**	**0.243**	**0.237**	**0.244**	**0.211**	**0.210**	**0.083**	**0.049**	–

Bold and shaded values are significant after False Discovery Rate application ($P < 0.01012$). Population abbreviations are as in Fig. 1. CC, Cap de Creus; BL, Blanes; C, Columbretes; ME, Menorca; MA, Mallorca; FOR, Formentera; CP, Cabo de Palos; CG, Cabo de Gata; CER, Sardinia; COR, Corsica; OT, Otranto; GR, Greece, HE, Herradura.

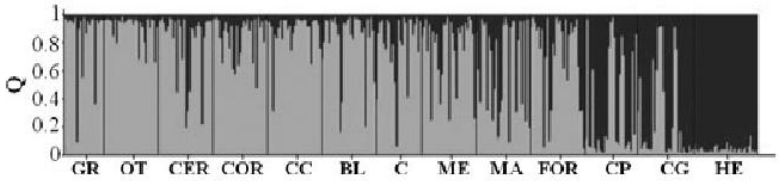

Fig. 2 Bar plot of the estimated membership fraction (Q) in each of the two genetically differentiated ($K = 2$) clusters identified by STRUCTURE. Each individual in the sampling locations on the x-axis is proportionally assigned to the clusters.

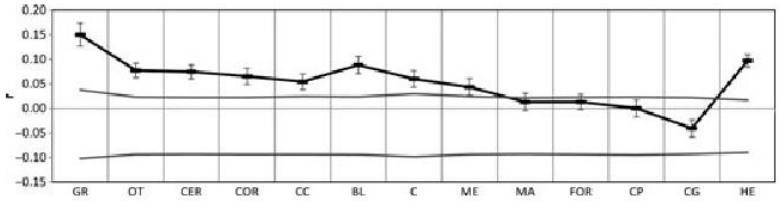

Fig. 3 Pairwise relatedness coefficient (Queller & Goodnight 1989) within each locality. Shaded area represent the 95% confidence interval of the null hypothesis of no differentiation across populations. Error bars are determined by permutations.

for $K = 2$) detected by STRUCTURE (Evanno *et al.* 2005). Furthermore, the highest likelihood was for $K = 2$ (Fig. S2, Supporting Information). The two populations (Fig. 2) could be identified by a Mediterranean unit and a more Atlantic-influenced group. This could be established by a hierarchical analysis with STRUCTURE, as running the program with different data subsets resulted in the separation of Mediterranean localities from CP, CG and HE (Fig. S2, Supporting Information). However, there was a slight substructure detected within the Mediterranean localities because L(K) was the same for $K = 1$ and $K = 2$. Furthermore, within each cluster, single individuals were genetically more related to the other cluster and could be considered as migrants. Seventeen individuals from the Mediterranean unit showed a Q value of 0.75 or more towards the other cluster with 11 of them found in the Balearic Islands (Table 1). All individuals of HE were assigned to the Atlantic group. The populations of CG and CP with mostly individuals of the Atlantic group had 21.7% of the individuals ascribed to the Mediterranean cluster (Table 1).

The mean pairwise relatedness values within localities revealed the individuals in GR to be most related with each other ($r = 0.150$) followed by BL ($r = 0.090$) and HE ($r = 0.089$) (Fig. 3). Low relatedness values close to zero could be seen for the Balearic Islands ME, MA and FOR. Negative relatedness values were found for the localities of CP and CG with the latter showing the lowest relatedness of all localities ($r = -0.0275$). Such negative values indicate that the relatedness value of the individuals tested was smaller than expected between random individuals (Queller & Goodnight 1989).

Focusing only on the western Mediterranean area to detect putative barriers on a smaller scale, PCoA were computed. The PCoAs with F_{ST} and Jost D values revealed similar structuring of localities, but the F_{ST} PCoA showed a higher resolution with 91.55% variation explained by the two coordinates in comparison with 82.6% variation if Jost D was applied (data not shown). The first axis of the PCoA clearly separated CG, CP and HE from the rest of the studied locations on the Spanish coast (Fig. 4). However, within those two groups, another division can be observed; a stronger separation of HE from CG and CP and a milder separation of the Balearic Islands (MA, ME, FOR) from the localities on the Spanish continental coast (C, BL, CC). The three separations correspond geographically with the three previously described barriers: AOF, IC and the BF (Figs 1 and 3).

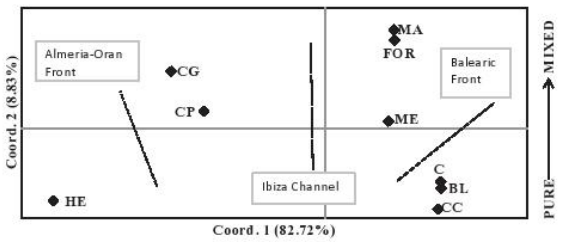

Fig. 4 Principal coordinates analysis calculated by F_{ST} values of nine Spanish locations. The positions of the oceanographic discontinuities in the area are plotted for comparison.

Assignment tests with GENECLASS resulted in 123 individuals being correctly assigned with over 80% of probability. Assignment values ranged from 0 to 0.33 for comparisons between different sampling sites (Table S2, Supporting Information), meaning that 33% of successfully assigned individuals collected in CG are from HE. When looking at self-recruitment, the values ranged between 0.21 and 0.68, with the highest two self-assignment scores found on the NSC and HE. Values from the BAYESASS analysis for pairwise comparisons were much lower between 0 and 0.04 and self-recruitment values larger, ranging between 0.78 and 0.83 (Table S2, Supporting Information). The results obtained with BAYESASS are significantly correlated with those of GENECLASS (Spearman's $\rho = 0.341$, $P < 0.01$).

Oceanography

Suspended particle circulation along the coast in the region between HE, CG and CP was almost exclusively south and eastwards (Fig. 5 and Table S2, Supporting Information), as observed by numerical particles (proxy of larvae) released at HE (green) and CP (blue) monthly from April to June at a depth of 20 m. The particles circulated through the Alboran Sea with the so-called Alboran Gyres (Tintoré *et al.* 1991) towards the coast of Morocco to then spiral back to the Spanish coast around the Almería-Oran Front (Allen *et al.* 2001). The particles released eastward of CG did not flow westward as they were stopped by the Almería-Oran Front. Although presenting monthly variation, some particles from CP crossed the IC. The particles released just south of the IC (red) partially moved slightly southwards; however, the majority flew north and once past the channel encountering the North Current which causes most of the water to be circulated towards the Balearic Current (Bouffard *et al.* 2010). Therefore, water coming up through the IC will rarely reach the east coast of the Spanish mainland. North of the IC particle movement (turquoise and yellow) was predominantly determined by the North Current, as it flows down the Spanish coast and circulates towards the Balearic Islands at the

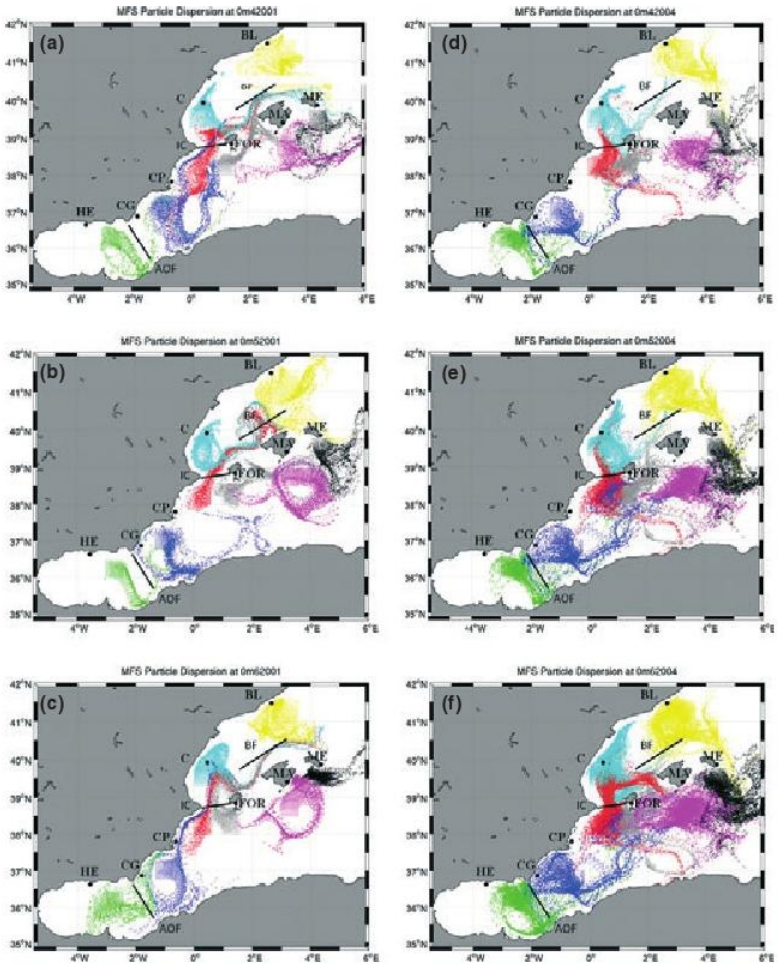

Fig. 5 Numerical particles (proxy of larvae) released at Herradura (green), Cabo de Gata (blue), South of the Ibiza Channel (red), Columbretes area (turquoise) and Cap de Creus (yellow) in (a) April 2001, (b) May 2001 and (c) June 2001 and (d) April 2004, (e) May 2004 and (f) June 2004 and allowed to drift for 30 days using the currents provided by the Mediterranean Forecasting System numerical model.

height of the channel. Simulation results for the months of April, May and June, the reproduction period of *S. cabrilla*, showed the same general pattern of dispersal. This can be also observed for different years, as simulations of 2001 and 2004 showed that these dispersal patterns are similar (Fig. 5).

The area south of Ibiza/FOR is characterized by a well-known spatial and temporal variability of the surface circulation (Pinot *et al.* 2002), a feature confirmed in our simulations that show a north-eastward flow in April 2001 crossing the MA channel, while flowing eastwards in May and northwards in June crossing the IC. In 2004, the situation was also highly variable, although with less well-defined trajectories. South of MA is characterized by weak surface circulation modulated by the presence of large-scale intermittent anticyclonic eddies detached from the Algerian Current (Puillat *et al.* 2002). In April 2001, a large and well-defined anticyclonic eddy was present in the area trapping the particle trajectories around the eddy. In 2004, the circulation was less well defined, particles being dispersed in the area (Fig. 5). In other words, in this area, the absence of a clear dominant current inhibits north/south or east/west exchanges. South of ME is also a highly variable zone as shown in 2001 simulations where an eastward/south-eastward and northeast/southward flow were obtained in April and May, while a very weak flow with low dispersion was obtained in June. The 2004 patterns from April to June were also showing predominant northeast and southward flows as in May 2001.

Correlations between genetic and oceanographic data were highly significant. The correlation between GENECLASS assignment tests and 2001 oceanographic data resulted in a Spearman's ρ of 0.691, and for the 2004 data, it was 0.685 (both $P < 0.001$) (Table S2, Supporting Information). If self-recruitment values were excluded, Spearman's ρ was also significant with both years (2001: $\rho = 0.338$, $P = 0.011$; 2004: $\rho = 0.444$, $P = 0.003$). When the data were randomized, the correlations were not significant (2001: $\rho = 0.171$, $P = 0.271$; 2004: $\rho = 0.126$, $P = 0.403$), showing that the significant results found with the original data are not a result of just chance. Correlation with the data obtained with BAYESASS was also highly significant (2001: $\rho = 0.456$, $P < 0.001$ and 2004: $\rho = 0.520$, $P < 0.001$), but revealing a slightly higher correlation in 2004 than in 2001.

Discussion

Genetic population network

Within the Mediterranean Sea, two genetically distinct clusters were revealed by the data set. Only two genetically differentiated clusters of the comber *S. cabrilla* were distinguished despite Jost *D* and F_{ST} differences between pairs of samples within the clusters. Surprisingly, samples from the eastern part of the Mediterranean, like GR, were grouped together with the central Mediterranean localities, the Balearic Islands and the NSC. The break between the two clusters is positioned geographically on the Spanish coast between the IC and the AOF with the intermediate locations representing an admixed population. The IC has rarely been considered in biological or genetic population analyses, although it has been a subject of oceanographic studies for many years (Pascual *et al.* 2002; Pinot *et al.* 2002; Fernandez *et al.* 2005). Within the western Mediterranean, most research on genetic structure of marine species focused on the effects of the Strait of Gibraltar and the AOF (Naciri *et al.* 1999; Patarnello *et al.* 2007; Bargelloni *et al.* 2008; Galarza *et al.* 2009a), and only few studies have looked at other barriers such as the BF (Galarza *et al.* 2009b; Schunter *et al.* 2011). The dusky grouper (*Epinephelus marginatus)* is the only fish species for which an influence of the IC on the genetic structure was directly tested, but no significant effect was found (Schunter *et al.* 2011). Nonetheless, the species composition of epibenthic crustaceans was found to differ between the south and the north coast, with the dividing location being around the IC (Abelló *et al.* 2002) and for the crab *Liocarcinus depurator* it has a profound effect on population differentiation (García-Mechán VH, personal communication). Furthermore, for the red gorgonian *Paramuricea clavata*, the IC and BF combined are revealed as the most relevant barrier (Mokhtar-Jamaï *et al.* 2011). Besides, it has been stated that the oceanographic processes around the IC are of special interest, as it is the location where the Mediterranean surface and intermediate waters meet with the less saline Atlantic waters (Fernandez *et al.* 2005). For several species, it has been observed that genetic structuring is defined by the influence of Atlantic waters vs. more Mediterranean water conditions of higher salinity and water temperature (Patarnello *et al.* 2007; Palero *et al.* 2008). Seemingly, the separation of the populations of *S. cabrilla* is influenced by the two different water masses.

Fine-scale subdivision of clusters could not be detected with programs like STRUCTURE, which often fail to identify differences when genetic differentiation is low (Latch *et al.* 2006; Hedgecock *et al.* 2007). However, pairwise F_{ST} and Jost *D* values in combination with measures such as PCoA can provide a more detailed insight into the structure within clusters. For the nine Spanish coastal and Island localities, *S. cabrilla* was genetically divided into four units: the Spanish north coast, the Balearic Islands, the Spanish south coast and HE. Apart from the clear discontinuities caused by the

IC and the AOF, the comber apparently is also affected by the BF. The Balearic Islands appear to be a genetically admixed group with influence from the Spanish continental coast reaching the islands because of the deviation of the North Current at the IC and individuals arriving from the South coast over the IC. However, as can be seen in the particle simulations as well as the hierarchical and putative barrier genetic analysis, ME appears to be more isolated than the other Balearic Islands.

The correlation between population structure and oceanographic processes suggests that genetic exchange is not driven by adult migration but rather by larval dispersal. For several other closely related species, behavioural site fidelity of adults has been identified (*Epinephelus coioides*, Antoro *et al.* 2006; *E. marginatus*, Hereu *et al.* 2006). Based on a study with acoustic telemetry, which demonstrated that the comber spends 95% of its time in an area of 0.77 km^2 (Alos *et al.* 2011), and by the present results, we can conclude that *S. cabrilla* is a territorial species with elevated adult site fidelity. Owing to the clear effect of oceanographic discontinuities on the genetic distribution of the species, larvae of the comber most likely are transported passively by the predominant currents.

Unidirectional and seasonal barriers

Despite the clear genetic structuring and effects of oceanographic discontinuities, values of pairwise Jost *D* and F_{ST} within the Mediterranean cluster were low, suggesting elevated levels of genetic exchange between the different units. Now, the challenge is to determine the origin and destination of larval migrants for a full understanding of the dispersal patterns. Possible migrants were revealed by assignment tests and relatedness values from which several trends in directional movement can be inferred. Individuals collected at the southern localities of CP and CG belonged to an admixed population with 21.6% having mixed genotypes. This population was comprised of individuals that were genetically related to and influenced by the Mediterranean cluster, while the majority were from the Atlantic-influenced cluster and mainly arrived over the AOF from HE. This suggests that the AOF, which has been described as a relatively permanent and strong barrier to gene flow (Patarnello *et al.* 2007), allows for genetic exchange from the west side of the barrier to the east side. However, counter directional movement could not be detected, because no individuals of Mediterranean influence were detected in HE and relatedness of individuals within this locality was high showing a comparatively high self-assignment value. There can be other possible explanations for isolation between two localities besides geographic isolation because of oceanographic discontinuities. Local adaptation to certain factors, such as temperature (e.g. for Atlantic salmon, *Salmo salar*, Dionne *et al.* 2007), environmental conditions (for European flounder, *Platichthys flesus*, Hemmer-Hansen *et al.* 2007) or possibly even anthropogenic impacts as an effect of differences in fishing pressure could decrease the dispersal success from other locations to HE. Furthermore, historical colonization is reflected on present-day genetic structuring; however, little is known about *S. cabrilla* in this regard. Nonetheless, the findings of gene-flow patterns were supported by oceanographic simulations of particle suspension, suggesting that the oceanographic front plays an important role in the determination of the observed genetic flow. There was retention of particles on the west side of the AOF and a small fraction moved eastwards crossing the AOF, while the westward movement of particles across the AOF was not observed. This predominant circulation pattern would encourage *S. cabrilla* larvae from HE to flow eastwards, but *S. cabrilla* larvae found east of CG would encounter a strong counter current when moving westwards. The majority of analyses validating the AOF as the cause of genetic breaks for species were based solely on significant genetic differentiation between different locations on either side of the AOF. Hereby, numerous species were divided into genetically distinct populations by the front, whereas other species with similar life-history traits but possibly different reproduction seasons were not affected (Patarnello *et al.* 2007; Bargelloni *et al.* 2008; Galarza *et al.* 2009a). However, no direct measures of gene flow across the barrier were taken nor was the direction of (the sometimes limited) gene flow considered. In this study, F_{ST} and Jost *D* between the two locations on either side of the AOF were high and an effect of the barrier was established; moreover, we detected a unidirectional genetic movement across the AOF from the west to the east along the coast. Even though oceanographic simulations reveal consistency in the flow pattern around the AOF over the years, further analysis in multiple years between genetic flow and current patterns should be undertaken to validate the long-term relationship.

The IC, as previously mentioned, is one of the main dividing barriers for *S. cabrilla*; however, a directional movement of individuals across this oceanographic discontinuity was also discovered. Several individuals collected at the admixed South Spanish coast (CP and CG) were genetically related to the Mediterranean cluster, illustrating the southwards movement of migration. Meanwhile, few individual collected on the NSC (BL, CC and C) originated from the Atlantic-influenced cluster and more individuals from the Balearic Islands

were assigned to the Atlantic cluster (Table 1). Hence, the main south–north genetic flow detected was from the southern locations towards the Balearic Islands (e.g. FOR) following the current deflection towards the Balearic Islands as simulated in the particle suspension data (see also Fernandez *et al.* 2005). It is expected that these conditions would facilitate the movement of *S. cabrilla* larvae from the southern localities across the IC towards the Balearic Islands, but rarely towards the Spanish continental coast. The encountered gene flow from the NSC to the South Spanish coast across the IC could be a result of the seasonality of this oceanographic discontinuity. Monserrat *et al.* (2008) demonstrate that the deviation of North Current towards the Balearic Islands occurs in spring and early summer and is caused by the formation of an anticyclonic gyre at the IC which hinders the southward flow. This gyre, however, is not formed after relatively mild winters with temperatures above 13 degrees (Lopez-Jurado *et al.* 2008), in the absence of which the North current can flow southward through the channel. Hence, the southward movement detected by the genetic analyses in this study could be caused by periodic mild winters. Nevertheless, it is possible that there is southward larval transport just along the coastline, and more refined modelling allowing the inclusion of more coastal regions could resolve this issue.

Epinephelus marginatus, the dusky grouper, has a similar average pelagic larval duration (24. 6 days) as *S. cabrilla* (24. 3 days; Macpherson & Raventós 2006), and it would be expected for both species to have a similar dispersal potential, especially because both show high adult site fidelity. Regardless, *E. marginatus* did not reveal any effect of the oceanographic processes around the IC on its genetic structure (Schunter *et al.* 2011). It has long been discussed why similar species show different genetic patterns, and there are many plausible reasons, one of them being the evolutionary history of each species. In this case, another explanation could be the differences in reproductive periods. The dusky grouper has been shown to spawn in August on the Spanish coast (Zabala *et al.* 1997) with the larvae dispersing in the plankton from late summer to fall. On the contrary, *S. cabrilla* reproduces in spring and the larvae are suspended in the water column from April to June (Sabatés 1990) and therefore susceptible to the discontinuity occurring at the IC in spring. These results could indicate that gene exchange through the IC is possible for species with reproduction periods in late summer or fall, whereas species that reproduce in winter (as *L. depurator*, García-Merchán, personal communication) or spring (as in the present work) are affected by the IC and genetic exchange might only occur in selected years after warm winters or in waters entrained along the coast.

The influence of oceanography on the genetic structure of marine species is now widely recognized (Galindo *et al.* 2003; Selkoe *et al.* 2006; Banks *et al.* 2007; Galarza *et al.* 2009a; Serra *et al.* 2010; White *et al.* 2010), whereas seasonal changes in current flow or temporal variability in oceanographic processes are rarely considered (Astraldi *et al.* 1995; Stenseth *et al.* 2006). On the west coast of the United States, the California and Alaska Current are influenced by a warm and a cold phase, and these alterations in current direction appear to lead to alterations in the biology of a wide range of fish and crustacean species in the Eastern North Pacific (Brodeur *et al.* 1996; Shanks & Eckert 2005). Carson *et al.* (2010) indicate that predominant gene-flow patterns for two congeneric mussel species (*Mytilus californianus* and *M. galloprovincianus*) are nearly opposite owing to the changing oceanographic patterns in different reproductive seasons. Shanks & Eckert (2005) even suggest that different life-history traits of a variety of species might have evolved in response to the seasonal changes in the California Current. On the Spanish coast, such an adaptation to oceanographic events is unlikely because *E. marginatus* and *S. cabrilla* have very similar life-history traits but reveal shifted reproduction times which could be one reason for the different genetic population structure of the two species. This emphasizes the importance of seasonality of oceanographic processes in combination with the species' reproductive period on its genetic structure and dispersal patterns. Furthermore, temporal sampling would be of interest as inter-annual variation in marine current patterns could introduce temporal genetic variation which can also be traced with more direct methods, such as mark-recapture or parentage analyses (Lowe & Allendorf 2010, Saenz-Agudelo *et al.* 2011).

Conclusions

The population structure and gene-flow pattern of a species can be influenced by predominant current patterns and oceanographic processes. Four units were identified, separated by three oceanographic barriers to gene flow: the IC, the AOF and the BF. The direction of the currents also plays a determining role in the population connectivity of *S. cabrilla*. The species genetic structuring is closely related to the present currents regimes across the oceanographic barriers. At the AOF, the comber larvae only dispersed from west to east of the barrier, indicating unidirectional movement similar to simulations of passive particles. Around the IC, connectivity was found between southern locations and the Balearic Islands as well as northern coastal localities

with the southern population. The latter can only be accomplished owing to a weakening of the gyre formed at the IC after mild winters or near coastal southward flow, allowing the current to flow south through the Channel. This phenomenon is a seasonal effect and coincides with the reproduction period of *S. cabrilla*, which might explain why other species with different reproduction times are not genetically divided by the IC. It is essential to include oceanographic data into population genetic studies, to not only understand the division of the connectivity but also to study directional dispersal patterns.

Acknowledgements

We thank all the people who helped us to collect samples and are grateful to Katie Munkres and anonymous reviewers for great suggestions on the manuscript. This work was partially funded by the Spanish Ministry of Science and Innovation through the BENTHOMICS (CTM2010-22218-C02-01) project and the FBBVA project (BIOCON 08 – 187/09). The authors are part of the research group 2009SGR-636, 2009SGR-665 of the Generalitat de Catalunya.

References

Abelló P, Carbonell A, Torres P (2002) Biogeography of epibenthic crustaceans on the shelf and upper slope off the Iberian Peninsula Mediterranean coasts: implications for the establishment of natural management areas. *Scientia Marina*, **66**(Suppl. 2), 183–198.

Allen JT, Smeed DA, Tintore J, Ruiz S (2001) Mesoscale subduction at the Almeria-Oran front. Part 1: agesotrophic flow. *Journal of Marine Systems*, **30**, 263–285.

Alos J, March D, Palmer M, Grau A, Morales-Nin B (2011) Spatial and temporal patterns in *Serranus cabrilla* habitat use in the NW Mediterranean by acoustic telemetry. *Marine Ecology Progress Series*, **427**, 173–186.

Antoro S, Na-Nakorn U, Koedprang W (2006) Study of genetic diversity of orange-spotted grouper, *Epinephelus coioides* from Thailand and Indonesia using microsatellite markers. *Marine Biotechnology*, **8**, 17–26.

Astraldi M, Bianchi CN, Gasparini GP, Morri C (1995) Climatic fluctuations, current variability and marine species distribution: a case study in the Ligurian Sea (north-west Mediterranean). *Oceanologica Acta*, **18**, 139–149.

Banks SC, Piggott MP, Williamson JE, Bové U, Holbrook NJ, Beheregaray LB (2007) Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology*, **88**, 3055–3064.

Bargelloni L, Alarcon JA, Alvarez MC *et al.* (2008) Discord in the family Sparidae (Teleostei): divergent phylogeographical patterns across the Atlantic–Mediterranean divide. *Journal of Evolutionary Biology*, **16**, 1149–1158.

Benjamini Y, Yekutieli D (2001) The control of false discovery rate under dependency. *Annals of Statistics*, **29**, 1165–1188.

Bouffard J, Pascual A, Ruiz S, Faugère Y, Tintoré J (2010) Coastal and mesoscale dynamics characterization using altimetry and gliders: a case study in the Balearic Sea. *Journal of Geophysical Research*, **115**, 1–17. doi: 10.1029/2009JC006087.

Bradbury IR, Bentzen P (2007) Non-linear genetic isolation by distance, life history and dispersal estimation in aquatic organisms. *Marine Ecology Progress Series*, **340**, 245–257.

Brodeur R, Frost B, Hare S, Francis R, Ingraham WJ (1996) Interannual variations in zooplankton biomass in the Gulf of Alaska, and covariation with California Current zooplankton biomass. *California Cooperative Oceanic Fisheries Investigations Reports*, **37**, 80–99.

Calderon I, Ortega N, Duran S, Becerro M, Pascual M, Turón X (2007) Finding the relevant scale: clonality and genetic structure in a marine invertebrate (*Crambe crambe*, Porifera). *Molecular Ecology*, **16**, 1799–1810.

Carreras-Carbonell J, Macpherson E, Pascual M (2006a) Primer note: characterization of 12 microsatellite markers in *Serranus cabrilla* (Pisces:Serranidae). *Molecular Ecology Notes*, **6**, 204–206.

Carreras-Carbonell J, Macpherson E, Pascual M (2006b) Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. *Molecular Ecology*, **15**, 3527–3539.

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.

Carson HS, López-Duarte PC, Rasmussen L, Wang D, Levin LA (2010) Reproductive timing alters Population Connectivity in Marine Metapopulations. *Current Biology*, **20**, 1926–1931.

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L (2007) Clinal variation in MHC diversity with temperature: evidence for the role of host–pathogen interaction on local adaptation in Atlantic salmon. *Evolution*, **61**, 2154–2164.

Estoup A, Largiadèr CR, Perrot E, Chourrout D (1996) Rapid one–tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. *Molecular Marine Biology and Biotechnology*, **5**, 295–298.

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.

Fernandez V, Dietrich DE, Haney RL, Tintore J (2005) Mesoscale, seasonal and interannual variability in the Mediterranean Sea using a numerical ocean model. *Progress in Oceanography*, **66**, 321–340.

Galarza J, Carreras-Carbonell J, Macpherson E *et al.* (2009a) The influence of oceanographic fronts and early-life history traits on connectivity among fish populations: a multi-species approach. *Proceedings of the national Academy of Science*, **106**, 1473–1478.

Galarza J, Macpherson E, Turner G, Rico C (2009b) Patterns of genetic differentiation between two co-occurring demersal species: ''the red mullet (*Mullus barbatus*) and the striped red mullet (*Mullus surmuletus*). *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 1478–1490.

Galindo HM, Olson DB, Palumbi SR (2006) Seascape Genetics: a Coupled Oceanographic-Genetic Model Predicts Population Structure of Caribbean Corals. *Current Biology*, **16**, 1622–1626.

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics! *Molecular Ecology*, **19**, 3845–3852.

Gilg MR, Hilbish TJ (2003) The geography of marine Larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology*, **84**, 2989–2998.

Goudet J (2002) *Fstat version 2.9.3.2. A program to estimate and test gene diversities and fixation indices*. Institute of Ecology, Lausanne, Switzerland. http://www. unil.ch/izea/softwares/fstat.

Goudet J, Raymond M, de Meeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1931–1938.

Guidetti P, Cattaneo-Vietti R (2002) Can mineralogiacal features influence distribution patterns of fish? A case study in shallow Mediterranean rocky reefs *Journal of Marine Biological Association of the UK*, **82**, 1043–1044.

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

Hedgecock D, Barber PH, Edmands S (2007) Genetic approaches to measuring connectivity. *Oceanography*, **20**, 70–79.

Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V (2007) Adaptive divergence in a high gene flow environment: *Hsc70* variation in the European flounder (*Platichthys flesus* L.). *Heredity*, **99**, 592–600.

Hereu B, Diaz D, Pasqual J, Zabala M, Sala E (2006) Temporal patterns of spawning of the dusky grouper *Epinephelus marginatus* in relation to environmental factors. *Marine Ecology Progress Series*, **325**, 187–194.

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL (2002) Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Molecular Biology and Evolution*, **19**, 2261–2264.

Johansson ML, Banks MA, Glunt KD, Hassel-Finnigan HM, Buonacorrsi VP (2008) Influence of habitat discontinuity, geographical distance, and oceanography on fine scale population genetic structure of copper rockfish (*Sebastes caurinus*). *Molecular Ecology*, **17**, 3051–3061.

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

Kimura M (1953) ''Stepping-stone'' model of population. *Annual Report of the National Institute of Genetics Japan*, **3**, 62–63.

Knutsen H, Jorde PE, Andre C, Stenseth NC (2003) Fine-scaled geographic population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology*, **12**, 385–39.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OEJ (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.

Leis JM (2007) Behaviour as input for modelling dispersal of fish larvae: behaviour, biogeography, hydrodynamics, ontogeny, physiology and phylogeny meet hydrography. *Marine Ecology Progress Series*, **347**, 185–193.

Levin L (2006) Recent progress in understanding larval dispersal: new directions and digressions. *Integrative and Comparative Biology*, **46**, 282–297.

Levinton JS (2001) *Marine Biology: Function, Biodiversity, Ecology*, 2nd edn. Oxford University Press, New York, USA.

Lopez-Jurado JL, Marcos M, Montserrat S (2008) Hydrographic conditions affecting two fishing grounds of Mallorca island (Western Mediterranean): during the IDEA Project (2003–2004). *Journal Marine Systems*, **71**, 303–315.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038–3051.

Macpherson E, Raventós N (2006) Relationship between pelagic larval duration and geographic distribution of Mediterranean littoral fishes. *Marine Ecology Progress Series*, **327**, 257–265.

Manly BFJ (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall, London.

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.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Millot C (1999) Circulation in the Western Mediterranean Sea. *Journal of Marine Systems*, **20**, 423–442.

Mokhtar-Jamaï K, Pascual M, Ledoux J-B *et al*. (2011) From global to local genetic structuring in a red gorgonian *Paramuricea clavata*: the interplay between oceanographic conditions and limited larval dispersal. *Molecular Ecology*, **20**, 3291–3305.

Monserrat S, López-Jurado J, Marcos M (2008) A mesoscale index to describe the regional circulation around the Balearic Islands. *Journal Marine Systems*, **71**, 413–420.

Naciri M, Lemaire C, Borsa P, Bonhomme F (1999) Genetic study of the Atlantic/Mediterranean transition in sea bass (*Dicentrarchus labrax*). *Journal of Heredity*, **90**, 591–596.

Narum SR (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics*, **7**, 783–787.

Neumann G (1968) *Ocean currents*. pp. 1. Elsevier, Amsterdam. p. 1.

Palero F, Abello P, Macpherson E, Gristina M, Pascual M (2008) Phylogeography of the European spiny lobster (*Palinurus elephas*): influence of current oceanographical features and historical processes. *Molecular Phylogenetics and Evolution*, **48**, 708–717.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, **13**, 146–158.

Pascual A, Nardelli BB, Larnicol G, Emelianov E, Gomis E (2002) A case of an intense anticyclonic eddy in the Balearic Sea (western Mediterranean). *Journal of Geophysical Research-Oceans*, **107**, 3183.

Pascual A, Pujol MI, Larnicol G, Le Traon PY, Rio MH (2007) Mesoscale mapping capabilities of multisatellite altimeter missions: first results with real data in the Mediterranean Sea. *Journal of Marine Systems*, **65**, 190–211.

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

Peakall R, Smouse PE (2006) Genalex 6.1: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pinet PR (2009) *Invitation to Oceanography*, 5th edn. Jones and Bartlett Publishers, Burlington, MA, USA.

Pinot J-M, Lopez-Jurado JL, Riera M (2002) The CANALES experiment (1996–1998). Interannual, seasonal, and mesoscale variability of the circulation in the Balearic Channels. *Progress in Oceanography*, **55**, 335–370.

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.

Planes S (2002) Biogeography and larval dispersal inferred from population genetic analysis. In: *Coral Reef Fishes. Dynamics and Diversity in a Complex Ecosystem* (ed. Sale PF), pp. 201–220. Academic Press, San Diego, California, USA.

Planes S, Jones GP, Thorrold SR (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Science*, **106**, 5693–5697.

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

Puillat I, Taupier-Letage I, Millot C (2002) Algerian eddies lifetime can near 3 years. *Journal of Marine Systems*, **31**, 245–259.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.

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

Raventós N, Macpherson E (2001) Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Marine Biology*, **138**, 1115–1120.

Rio MH, Poulain PM, Pascual A, Mauri E, Larnicol G, Santoleri R (2007) A Mean Dynamic Topography of the Mediterranean Sea computed from altimetric data, in-situ measurements and a general circulation model. *Journal of Marine Systems*, **65**, 484–508.

Robinson A, Leslie W, Theocharis A, Lascaratos A (2001) *Encyclopedia of Ocean Sciences, chap. Mediterranean Sea Circulation, 1689–1706*. Academic Press Ltd., London.

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

Sabatés A (1990) Distribution pattern of larval fish populations in the northwestern Mediterranean. *Marine Ecology Progress Series*, **59**, 75–82.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2009) Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis using different gene flow scenarios. *Molecular Ecology*, **18**, 1765–1776.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2011) Connectivity dominates larval replenishment in a costal reef fish metapopulation. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 2954–2961.

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

Schunter C, Carreras-Carbonell J, Planes S *et al.* (2011) Population connectivity in a commercial and endangered species: the dusky grouper (*Epinephelus marginatus*). *Journal of Experimental Marine Biology and Ecology*, **401**, 126–133.

Selkoe KA, Gaines SD, Caselle JE, Warner RR (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology*, **87**, 3082–3094.

Selkoe KA, Watson JR, White C *et al.* (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, **19**, 3708–3726.

Serra IA, Innocenti AM, Di Maida G *et al.* (2010) Genetic structure in the Mediterranean seagrass *Posidonia oceanica*: disentangling past vicariance events from contemporary patterns of gene flow. *Molecular Ecology*, **19**, 557–568.

Shanks AL, Eckert GL (2005) Population persistence of California current fishes and benthic crustaceans: a marine drift paradox. *Ecological Monographs*, **75**, 505–524.

Stenseth NC, Jode PE, Chan K-S *et al.* (2006) Ecological en genetic impact of Atlantic cod larval drift in the Skagerrak. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1085–1092.

Thorrold SR, Jones GP, Planes S, Hare J (2006) Transgenerational marking of embryonic otoliths in marine fishes using barium stable isotopes. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 1193–1197.

Tintoré J, Gomis D, Alonso S, Parrilla G (1991) Mesoscale dynamics and vertical motion in the Alboran Sea. *Journal of Physical Oceanography*, **21**, 811–823.

Tonani M, Pinardi N, Dobricic S, Pujol I, Fratianni C (2008) A high-resolution free-surface model of the Mediterranean Sea. *Ocean Science*, **4**, 1–14.

Tonani M, Pinardi N, Fratianni C *et al.* (2009) Mediterranean Forecasting System: forecast and analysis assessment through skill scores. *Ocean Science*, **5**, 649–660.

Torcu-Koc H, Turker-Cakir D, Dulcic J (2004) Age, growth and mortality of the comber, *Serranus cabrilla* (Serranidae) in the Edremit Bay (NW Aegean Sea, Turkey). *Cybium*, **28**, 19–25.

Uriz MJ, Turon X, Mariani S (2008) Ultrastructure and dispersal potential of sponge larvae: tufted versus evenly ciliated parenchymellae. *Marine Ecology*, **29**, 280–297.

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

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

White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ (2010) Ocean current help explain population genetic structure. *Proceedings of the Royal Society of London Series B, Biological Sciences*, **277**, 1685–1694.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Wright S (1969) *Evolution and the Genetics of Populations*. University of Chicago Press, Chicago, IL.

Zabala M, Louisy P, Garcia-Rubies A, Gracia V (1997) Sociobehavioral context of the reproduction in the Mediterranean dusky grouper *Epinephelus marginatus* (Lowe, 1834) (Pisces, Serranidae) in the Medes Islands Marine Reserve (NW Mediterranean, Spain). *Scientia Marina*, **61**, 79–98.

Zitari-Chatti R, Chatti N, Fulgione D *et al.* (2009) Mitochondrial DNA variation in the caramote prawn Penaeus (Melicertus) kerathurus across a transition zone in the Mediterranean Sea. *Genetica*, **137**, 439–447.

C.S. studies population connectivity and sexual selection for her PhD thesis. E.M., M.P. and J.C-C., are interested in marine population genetics and phylogeography. J.T., E.V-V., and A.P. are interested in mesoscale dynamics combining modelling with satellite and in situ observations. P.G. studies patterns of connectivity and dispersal of marine fish species.

Data accessibility

Microsatellite data: DRYAD entry (Data identifier: doi:10.5061/dryad.2rr59).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Summary statistics for the 11 microsatellites (Sc03–Sc15) and 13 populations of *Serranus cabrilla*.

Table S2 Migration rates and particle dispersal from source (rows) to sink (columns).

Fig. S1 Dispersion plot of the relationship of distance (ln Km) and $F_{ST}/(1 - F_{ST})$. Comparisons including only Spanish locations are represented in black while comparisons including other Mediterranean locations (GRE, LE, CER, COR) are represented in grey. Mantel test is non significant when comparing all populations ($R^2 = 0.1873$, $P = 0.1148$) and significant when using Spanish localities only ($R^2 = 0.4664$, $P < 0.001$).

Fig. S2 Graph representing the Structure output of the most likely number of populations (K) including the standard deviation between runs for three different data sets: All populations, all populations except HE, all populations expect CP, CG and HE and only CG, CP and HE.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supplementary data Table 1. Summary statistics for the 11 microsatellites (Sc03-Sc15) and 13 populations of *Serranus cabrilla*.

		Sc03	Sc04	Sc05	Sc06	Sc07	Sc08	Sc11	Sc12	Sc13	Sc14	Sc15
GR	**N**	22	21	22	21	18	22	22	22	20	18	22
	Na	5	8	17	13	9	9	3	5	10	11	3
	Ho	0.636	0.857	0.773	0.952	0.722	0.636	0.091	0.545	0.850	0.778	0.182
	He	0.582	0.788	0.925	0.865	0.745	0.707	0.088	0.533	0.813	0.742	0.168
	F_{IS}	-0.071	-0.064	**0.187**	-0.077	0.060	0.122	-0.012	0.000	-0.021	-0.019	-0.057
	F_{NA}	---	---	0.086	---	---	---	---	---	---	---	---
OT	**N**	30	30	30	30	30	30	30	30	30	30	30
	Na	8	10	20	13	9	10	4	4	12	16	4
	Ho	0.667	0.767	0.767	0.800	0.800	0.400	0.200	0.533	0.700	0.900	0.267
	He	0.689	0.834	0.928	0.892	0.759	0.669	0.242	0.507	0.846	0.843	0.241
	F_{IS}	0.049	0.097	**0.190**	0.120	-0.037	**0.417**	0.189	-0.035	0.189	-0.050	-0.092
	F_{NA}	---	---	0.087	0.204	---	---	---	---	0.088	---	---
CE	**N**	30	30	30	30	30	30	30	30	30	30	30
	Na	7	9	25	11	8	8	4	7	10	18	3
	Ho	0.600	0.767	0.800	0.867	0.733	0.667	0.200	0.633	0.767	0.833	0.333
	He	0.658	0.773	0.941	0.858	0.776	0.709	0.186	0.571	0.823	0.836	0.292
	F_{IS}	0.105	0.025	**0.167**	0.007	0.071	0.077	-0.058	-0.092	0.086	0.020	-0.126
	F_{NA}	---	---	0.073	---	---	---	---	---	---	---	---
COR	**N**	30	30	30	30	30	30	30	30	30	30	30
	Na	10	9	18	12	8	12	5	4	13	14	4
	Ho	0.667	0.700	0.867	0.933	0.767	0.767	0.200	0.600	0.733	0.767	0.300
	He	0.657	0.811	0.923	0.889	0.801	0.679	0.243	0.554	0.768	0.791	0.366
	F_{IS}	0.003	0.153	0.078	-0.032	0.059	-0.112	0.193	-0.065	0.062	0.047	0.196
	F_{NA}	---	---	**---**	---	---	---	---	---	---	---	---
CC	**N**	30	30	30	30	29	30	30	30	30	30	30
	Na	9	10	22	13	8	11	3	4	8	12	4
	Ho	0.667	0.767	0.867	0.733	0.862	0.567	0.133	0.633	0.867	0.867	0.400
	He	0.674	0.771	0.944	0.840	0.792	0.728	0.155	0.668	0.741	0.830	0.434
	F_{IS}	0.029	0.022	**0.099**	0.144	-0.071	**0.238**	0.156	0.069	-0.153	-0.027	0.095
	F_{NA}	---	---	0.042	---	---	0.116	---	---	---	---	---
BL	**N**	30	30	30	30	30	30	30	30	30	30	30
	Na	7	8	21	12	6	9	3	5	9	12	4
	Ho	0.533	0.700	0.833	0.633	0.767	0.567	0.200	0.500	0.767	0.733	0.333
	He	0.545	0.758	0.937	0.828	0.751	0.667	0.212	0.611	0.779	0.808	0.366
	F_{IS}	0.038	0.094	**0.127**	**0.251**	-0.005	0.167	0.072	0.198	0.033	0.110	0.105
	F_{NA}	---	---	0.054	0.121	---	---	---	---	---	---	---
C	**N**	25	25	25	25	25	25	25	25	25	25	25
	Na	7	8	20	12	6	10	3	6	8	15	4
	Ho	0.680	0.840	0.760	0.880	0.800	0.520	0.080	0.640	0.720	0.840	0.240
	He	0.702	0.764	0.934	0.856	0.789	0.755	0.078	0.654	0.806	0.818	0.252
	F_{IS}	0.051	-0.079	0.206	-0.008	0.006	0.330	-0.011	0.042	0.126	-0.007	0.068
	F_{NA}	---	---	0.094	---	---	0.160	---	---	---	---	---
ME	**N**	30	29	29	30	30	30	30	29	28	30	30

	Na	8	10	20	12	5	10	5	5	12	18	5
	Ho	0.833	0.724	0.655	0.833	0.767	0.867	0.433	0.690	0.786	0.767	0.633
	He	0.659	0.766	0.934	0.869	0.694	0.741	0.387	0.593	0.828	0.852	0.517
	F_{IS}	-0.249	0.072	**0.314**	0.058	**-0.088**	-0.153	-0.104	-0.146	0.070	0.117	-0.208
	F_{NA}	---	---	0.149	---	---	---	---	---	---	---	---
MA	**N**	30	30	30	30	30	30	30	30	30	30	30
	Na	7	9	21	14	9	9	5	6	12	18	6
	Ho	0.700	0.833	0.767	0.967	0.800	0.700	0.333	0.600	0.867	0.867	0.433
	He	0.677	0.826	0.937	0.890	0.790	0.716	0.345	0.544	0.806	0.877	0.453
	F_{IS}	-0.018	0.008	**0.198**	-0.069	0.004	0.039	0.051	-0.085	-0.059	0.028	0.061
	F_{NA}	---	---	0.091	---	---	---	---	---	---	---	---
FOR	**N**	29	30	30	30	30	30	29	30	30	30	30
	Na	7	10	22	12	7	15	4	5	11	18	5
	Ho	0.655	0.933	0.767	0.900	0.700	0.867	0.345	0.467	0.867	0.833	0.467
	He	0.650	0.817	0.926	0.889	0.750	0.823	0.328	0.543	0.797	0.853	0.412
	F_{IS}	0.009	-0.126	0.188	0.005	0.084	-0.036	-0.035	0.157	-0.070	0.040	-0.115
	F_{NA}	---	---	0.085	---	---	---	---	---	---	---	---
CP	**N**	29	28	29	28	27	29	29	29	29	29	29
	Na	5	10	19	13	10	11	6	5	14	21	5
	Ho	0.586	0.821	0.724	0.893	0.593	0.621	0.690	0.690	0.828	0.966	0.724
	He	0.576	0.822	0.932	0.840	0.796	0.693	0.679	0.592	0.855	0.910	0.612
	F_{IS}	0.000	0.019	**0.239**	-0.045	0.273	0.122	0.002	-0.148	0.050	-0.043	-0.167
	F_{NA}	---	---	0.111	---	0.128	---	---	---	---	---	---
CG	**N**	31	31	31	31	31	31	31	31	31	31	31
	Na	8	11	20	12	8	11	6	4	11	16	6
	Ho	0.774	0.871	0.710	0.742	0.806	0.645	0.613	0.355	0.871	0.742	0.581
	He	0.673	0.817	0.931	0.857	0.800	0.771	0.705	0.530	0.858	0.894	0.671
	F_{IS}	-0.135	-0.049	**0.253**	0.151	0.008	0.179	0.147	0.345	0.001	0.186	0.151
	F_{NA}	---	---	0.117	---	---	---	---	0.162	---	0.086	---
HE	**N**	35	35	35	35	35	35	35	35	35	35	35
	Na	8	7	19	11	7	8	5	4	13	17	8
	Ho	0.657	0.743	0.686	0.886	0.743	0.686	0.714	0.571	0.886	0.800	0.629
	He	0.631	0.783	0.928	0.843	0.786	0.727	0.664	0.560	0.855	0.871	0.627
	F_{IS}	-0.028	0.066	**0.274**	-0.036	0.069	0.071	-0.061	-0.005	-0.022	**0.096**	0.013
	F_{NA}	---	---	0.131	---	---	---	---	---	---	---	---

Population abbreviations as in Figure 1. N: Number of individuals. Na: Number of alleles. Ho: Observed Heterozygozity. He: Expected Heterozygozity. F_{IS}: Fixation Index with indication of significance after control of False discovery rate in bold ($p<0.00902$). F_{NA}: Frequency of null alleles identified with MICROCHECKER, --- no null alleles present.

Supplementary data Table 2. Migration rates and particle dispersal from source (rows) to sink (columns). Upper values in each cell correspond to GENECLASS assignment/ BAYESASS migration rate. Lower values correspond to 2001 particle dispersal/ 2004 particle dispersal. Abbreviations are as in Fig.1, NSC= North Spanish coast (CC and BL).

		SINK							
		NSC	**C**	**ME**	**MA**	**FOR**	**CP**	**CG**	**HE**
SOURCE	**NSC**	0.43 / 0.79 363 / 363	0.13 / 0.02 0 / 0	0 / 0.02 3 / 8	0.09 / 0.03 0 / 0	0.08 / 0.02 0 / 0	0.04 / 0.02 0 / 0	0 / 0.02 0 / 0	0 / 0.03 0 / 0
	C	0.15 / 0.02 0 / 0	0.6 / 0.79 363 / 363	0.1 / 0.03 1 / 0	0 / 0.03 0 / 0	0 / 0.03 0 / 21	0 / 0.02 0 / 0	0 / 0.02 0 / 0	0 / 0.03 0 / 0
	ME	0 / 0.02 0 / 0	0.4 / 0.02 0 / 0	0.5 / 0.78 363 / 363	0.1 / 0.02 0 / 0	0.2 / 0.02 0 / 0	0.1 / 0.02 0 / 0	0 / 0.02 0 / 0	0 / 0.02 0 / 0
	MA	0.12 / 0.03 0 / 0	0 / 0.02 0 / 0	0.08 / 0.02 35 / 0	0.2 / 0.78 363 / 363	0.38 / 0.03 13 / 23	0 / 0.02 0 / 0	0 / 0.02 0 / 0	0 / 0.02 0 / 0
	FOR	0 / 0.03 0 / 0	0.1 / 0.03 3 / 5	0 / 0.03 0 / 4	0.3 / 0.03 98 / 8	0.2 / 0.78 363 / 363	0.2 / 0.02 8 / 39	0.2 / 0.02 0 / 0	0.1 / 0.03 0 / 0
	CP	0.07 / 0.03 0 / 0	0 / 0.03 34 / 125	0 / 0.04 0 / 0	0.07 / 0.03 0 / 0	0.13 / 0.04 0 / 109	0.27 / 0.82 363 / 363	0.2 / 0.03 10 / 5	0.2 / 0.03 0 / 0
	CG	0.04 / 0.03 0 / 0	0 / 0.03 37 / 0	0 / 0.03 0 / 0	0 / 0.03 0 / 0	0 / 0.03 0 / 43	0.17 / 0.03 41 / 41	0.33 / 0.83 363 / 363	0.42 / 0.03 0 / 0
	HE	0 / 0.02 0 / 0	0 / 0.02 0 / 0	0 / 0.02 0 / 0	0 / 0.02 0 / 0	0 / 0.02 0 / 1	0 / 0.02 0 / 3	0.28 / 0.02 25 / 0	0.68 / 0.78 363 / 363

Supplementary data Figure S1:

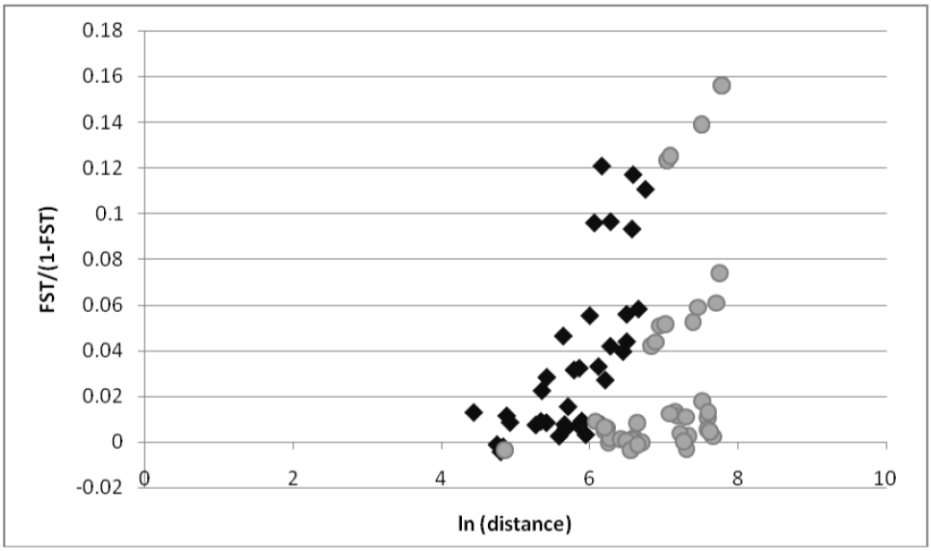

Fig. S1: Dispersion plot of the relationship of distance (ln Km) and $F_{ST}/(1-F_{ST})$. Comparisons including only Spanish locations are represented in black while comparisons including other Mediterranean locations (GR, LE, CER, COR) are represented in grey. Mantel test is non significant when comparing all populations (R^2= 0.1873, P=0.1148) and significant when using Spanish localities only (R^2=0.4664, P<0.001).

Supplementary data Figure S2:

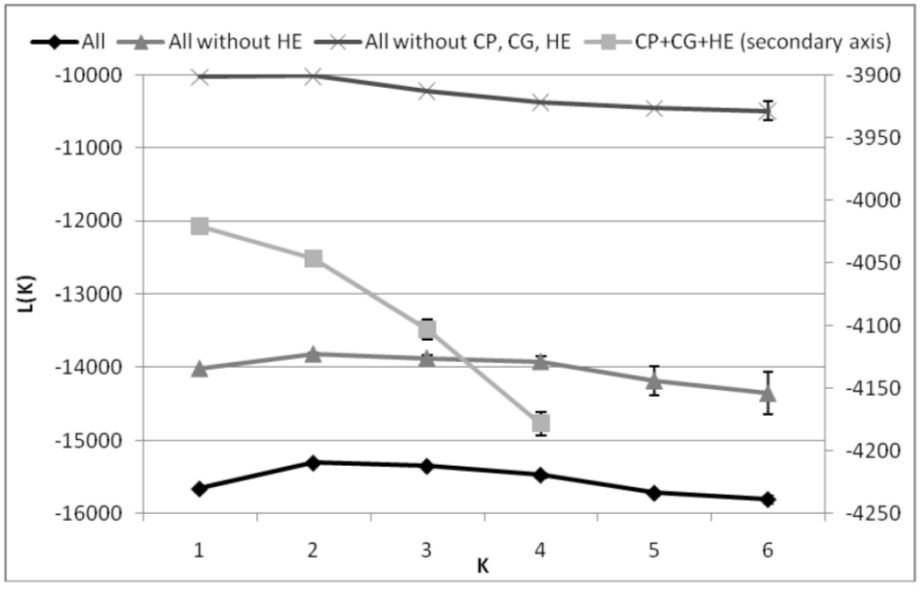

Fig. S2: Graph representing the Structure output of the most likely number of populations (K) including the standard deviation between runs for three different data sets: All populations, all populations except HE, all populations expect CP, CG and HE and only CG, CP and HE.

3.2. Comparative transcriptomics among reproductive strategies

3.2. Comparative transcriptomics among reproductive strategies

Publication 3: Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics

Publicación 3: Análisis transcriptómico y de expresión diferencial para una especie no-modelo con tácticas de reproducción alternativas

RESUMEN

Antecedentes
La dominancia social es importante para el éxito reproductivo de los machos en muchas especies. En el caso de *Tripterygion delaisi*, durante la época de reproducción, algunos machos cambian de color e invierten recursos en la construcción de un nido y defensa del territorio, mientras otros machos no cambian de color, pero pueden depositar su esperma cuando las hembras ponen los huevos. Utilizando RNAseq, describimos la expresión genética diferencial del cerebro entre machos dominantes, machos secundarios y hembras para estudiar la señal molecular del dimorfismo masculino.
Resultados
Encontramos que hay más genes que se expresan diferencialmente entre los dos fenotipos masculinos que entre machos y hembras. Esto sugiere que la plasticidad fenotípica es un factor más importante en la expresión genética diferencial durante la reproducción que el dimorfismo sexual. El macho territorial sobreexpresa genes relacionados con la plasticidad sináptica y el macho secundario sobreexpresa genes involucrados en la diferenciación y el desarrollo.
Conclusiones
Los genes previamente sugeridos como candidatos de la dominancia social en el contexto de las estrategias alternativas de reproducción, parecen ser específicos de cada especie y en este estudio presentamos una lista de genes nuevos, los cuales están diferencialmente expresados en *Tripterygion delaisi*. Este es el primer estudio involucrando todo el genoma de una especie no-modelo en el contexto de estrategias alternativas de reproducción y aporta información esencial para estudios futuros sobre la base molecular de la dominancia social.

Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics

C. Schunter[‡†*], S. Vollmer[§], E. Macpherson[‡], M. Pascual[†]

[‡] Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Car. Acc. Cala St. Francesc 14, Blanes, 17300 Girona, Spain

[†]Dept. Genètica, Univ. Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain

[§]Marine Science Center, Northeastern University, 430 Nahant Road, MA, USA.

[*]Corresponding author: Telephone: +34-972-33-61-01 Fax: +34-972-33-78-06.

e-*mail address*: cschunter@ceab.csic.es

Keywords: RNAseq, differential expression, alternative mating tactics, social dominance, phenotypic plasticity, *Tripterygion delaisi*.

Running title: Differential gene expression in wild breeders

Abstract

Background

Social dominance is important for the reproductive success of males in many species. In the black-faced blenny (*Tripterygion delaisi*) during the reproductive season, some males change color and invest in nest making and defending a territory, whereas others do not change color and 'sneak' reproductions when females lay their eggs. Using RNAseq, we profiled differential gene expression between the brains of territorial males, sneaker males, and females to study the molecular signatures of male dimorphism.

Results

We found that more genes were differentially expressed between the two male phenotypes than between males and females. This suggests that phenotypic plasticity is a more important factor in differential gene expression during the reproductive period than sexual dimorphism. The territorial male overexpresses genes related to synaptic plasticity and the sneaker male overexpresses genes involved in differentiation and development.

Conclusions

Previously suggested candidate genes to social dominance in the context of alternative mating strategies seem to be mostly species-specific and here we present a list of novel genes which are differentially expressed in *Tripterygion delaisi*. This is the first genome-wide study for a non-molecular model species in the context of alternative mating strategies and provides essential information for further studies investigating the molecular basis of social dominance.

Background

Polygamous mating systems are often defined by social dominance where territorial individuals top the social hierarchy. Alternative mating strategies in fish species are commonly associated with a dominance hierarchy including dominant or territorial males and secondary males or so-called sneaker males [1]. Being the dominant individual comes at a cost, as the territorial male has to invest in defending the territory, attract the female and guard the nest [2]. In some fish species those dominant individuals even do not feed during the reproductive period of several months [3]. The sneaking individual on the other hand can obtain reproductive success by sneaking into nests or mimicking female behavior and phenotype [4].

The term 'social dominance' suggests that dominance depends on the social setting. The territorial male is often the largest male present in many fish species but there are exceptions [5, 6]. This is especially true for fish species where the change from sneaker to territorial male is not fixed for life, but temporary and therefore plastic. Here, the switch to becoming territorial could depend on the presence of plausible nest sites, the number of mature females or even the presence or absence of other males [4]. For instance, in the goby *Gobius niger* [7] and the blenny *Trypterigion delaisi* [8, 9] a sneaker male can switch into a territorial male after removing the previous territorial male.

Phenotypic plasticity, the ability of a genotype to adapt to external conditions by changing its phenotype, has received considerable attention in evolutionary ecology [10]. The assimilation of an initially plastic response to an altered environment and therefore the maintenance of the phenotype has long been understood as a fundamental component in selection and evolution [11, 12]. More recently, short-term and non-adaptive phenotypic changes have been the focus of several studies [13, 14]. In many species reproduction is a temporal or a short-term event, taking place for example only once a year in the reproductive period. This means that the alterations occurring during this period, being behavioral and/or phenotypic, are plastic and mostly reversible changes [15]. In male alternative mating tactics, the different tactics are linked with differences in behavior often leading to phenotypic dimorphism with secondary sexual traits. Social interactions, in these cases, have been shown to trigger the behavioral and phenotypic change [16].

Social influences and behavioral changes lead to alterations in gene expression in the brain [17–19]. More specifically, social stimuli can lead to short term deviation from the baseline of gene expression in the brain [17]. While it is unclear how these changes are transmitted to the other organs, clearly the brain plays a vital part [20]. The neural basis of social status has been studied in the brain in cichlid fish species, due to their extreme diversity and the facility to be handled and kept in the laboratory (see [4] and references cited therein). Also, the Atlantic salmon (*Salmo salar*) exhibits alternative reproductive tactics and the plasticity in the development of these reproductive phenotypes has been investigated [21, 22]. Gene expression patterns related to phenotypic plasticity in different mating strategies have been analyzed either by targeting single genes or via microarray analyses [15, 19, 23], but the non-dominant male was either not reproductively active [19] or the two male types reproduce at different ages [22, 23] or the studies focused on male aggression without the presence of females [24, 25]. To our knowledge no attempt to characterize phenotypic plasticity in wild male phenotypes from the same population, both reproductively active, has been carried out using a genome-wide approach.

Our study species *Tripterygion delaisi* (Tripterygiidae), also called the black-faced blenny, is a common small rocky shore fish from the Mediterranean Sea and the east Atlantic coast [26, 27]. The black-faced blennies live camouflaged with the rock or algae they inhabit for most of the year. The sneaker males as well as the females exhibit the same camouflaged phenotype throughout the whole year, but in spring when the reproductive period starts, some males change their phenotype to a black head and a bright yellow coloring across the rest of the body. These males start protecting a small territory, which is referred to as their nest, against predators and other secondary males [8]. This coloration and behavior is transitory and only observed during the reproductive period. If a territorial male is removed from its nest in 20% of the cases a sneaker male takes over the nest changes its coloration and behavior and becomes territorial [9], hence, territoriality is a plastic trait. After female courtship from the territorial male, the female lays the eggs directly into the nest and leaves. The territorial male fertilizes the eggs by ejecting the sperm directly on them and is left to protect the eggs until the larvae hatch. The sneaker male can dart by and ejaculates its sperm over the nest from a distance [8]. Thus, in *T. delaisi* alternative mating tactics can be observed for two types of reproductively active males with different phenotypes.

In non-model species, it is often difficult to study molecular differentiation of plastic traits especially in absence of a reference genome. Here we used RNAseq to detect genes differentially expressed between males with alternative mating strategies in *Tripterygion delaisi* as well as between males and females. Wang and coauthors [28] stated that this method is a revolutionary approach to transcriptomics and due to the advances in technology such next generation sequencing approaches are now also feasible for non-model organisms. By sequencing and generating a *de novo* transcriptome assembly it is possible to look at a huge variety of expressed genes and demonstrate key genes which are expressed at a particular moment [29]. By using RNAseq in this study we generate a genome-wide catalogue of genes expressed during the reproductive period of *T. delaisi* and analyze expression patterns and differential expression to identify differences across the brain transcriptome of territorial males, sneaker males and females.

Results and Discussion

De novo transcriptome assembly and annotation

The *de novo* assembly of the reference transcriptome was performed with 50,360,654 trimmed reads (Phred score 35) of eight pooled individuals of normalized cDNA libraries and 194,148,779 trimmed reads of fifteen separately sequenced normalized samples. (Supplementary Material Table S1, Fig S1). Normalizing libraries allowed for the detection of genes also expressed at low levels and therefore a more complete reference transcriptome. With the Trinity *de novo* assembler 334,973 contigs were produced including different isoforms per contig. This is the first reference transcriptome for a fish belonging to the Perciforms Suborder Blennioidei (which includes more than 800 species and 130 genera [30]) and therefore provides a valuable resource and a first step towards a comprehensive understanding of genome-wide gene expression.

The *de novo* assembly includes differentially spliced isoforms [31]. Most contigs in the reference assembly had one isoform, but many contigs combined differentially spliced isoforms, with the

number of isoforms increasing with contig length (Fig.1). The few published genome-wide expression studies on non-model species mostly focus on differential expression at the genes level and de novo assemblies do not include alternative splicing [e.g 30]. However, alternative splicing has been demonstrated to have fundamental effects in the development and maintenance in eukaryotes with 92-94% of human genes undergoing alternative splicing [33, 34]. Alternative splicing is shown to be responsible for a number of diseases by changing the biological function with differently spliced isoforms [35]. Hence, it is likely that even social behavior or phenotypic expression patterns can be influenced or even dominated by alternative splicing.

The transcriptome contigs were then blasted against the Uniprot database to detect orthologous proteins and identify the gene name and their gene function. Homology with known proteins was only detected (<E-values $1x10^{-10}$) for 71,513 contigs which represent 21.4% of the transcriptome. 67,165 contigs had blast results between $1x10^{-11}$ to $1x10^{-180}$ and 4348 contigs had an E-value of 0. The RNAseq method used allowed a global analysis of the general molecular expression patterns of a wild non-model species and the discovery of a large amount of expressed transcripts. However, many of the contig sequences could not be matched to a known gene or to a functional description in the genetic databases. Small percentages of successful detection of homologous genes (e.g. 37% in wasps [36]) is a common limitation of studying a non-model species at this moment. As the sequences without homology could be novel genes of great importance to our non-model species we divided our transcriptome contig set into two subsets for further analysis: contigs with protein homology (with Blast) and contigs without protein homology (without Blast). The 'with Blast' set had slightly larger contig sizes than the 'without Blast' set (Fig.2a). Similarly, median open reading frame (ORF) sizes in nucleotide bases differed slightly between sets, with larger ORFs in the 'with blast' than in the without Blast set ('with Blast': 114.9, 'without Blast': 80.31). Protein coding potential above 0.5 was estimated with CPAT [37] and the without blast set has lower protein coding potential (Fig.2b). Lower GC content as well as lower expression levels were encountered for the 'without Blast' set (Fig.2 c,d). Homology with protein domains was highest in the 'with Blast' set and almost no homology was found in the 'without Blast' set (0.04%). Overall, the described characteristics of the 'without Blast' sequences suggest that a large number of these contigs are non-coding

transcripts, as has also been suggested for the wasp [36, 38]. Nonetheless, these analyses are all based on homology with annotated data in databases and therefore lack of sequence conservation across species, gene fragmentation and the lack of genomic information for *T. delaisi* or closely related species prevents annotation of contigs. An assembly of a genome reference sequence or in general more genomic resources for *T. delaisi* or closely related species would increase the quality of the transcriptome assembly.

The most expressed genes throughout the genome (first 1% of mean normalized expression) regardless of phenotype were annotated for biological function in Blast2GO [39] and enrichment analysis resulted in overexpression of 134 slimmed GO-terms in biological function. The majority of these terms (about 1/3) are involved in processes of translation and transport such as translation elongation and SRP−dependent cotranslational protein targeting to membrane (Fig.3). Interestingly, one of the thirty enriched upper-hierarchy GO-terms is social behavior. When looking at the first ten genes with the highest overall expression count across the whole genome, we identified a gene called Ependymin (*epd1*). This gene protein is associated with neuroplasticity and regeneration, and a predominant protein in the cerebrospinal fluid of teleost fish [40]. With our results we confirm that this gene is highly important in the brain of *T. delaisi*.

Differential expression between phenotypes

To determine the molecular basis that distinguishes the three phenotypes of *Tripterygion delaisi*; the territorial male, the sneaker male and the female three pairwise comparisons were performed. The comparison between the territorial male and the sneaker male resulted in the highest number of differentially expressed contigs with 600 significant contigs after FDR correction and adjustment for individual variation and 209 annotated contigs (Table 1, for a list of all genes see Supplementary Materials Tables S1, S2 and S3). The territorial male differentially over-expressed more genes than the other two phenotypes (43%). In general, fewer differentially expressed genes were found for the male and female comparisons than between the two male phenotypes suggesting that during the reproductive period of *T. delaisi* phenotypic plasticity, rather than sexual dimorphism, defines the gene expression patterns.

Phenotypic plasticity, the ability of a genotype to adapt to external conditions by changing its phenotype, has received considerable attention in evolutionary ecology, whereas focus has been laid on adaptive and selective phenotypic changes [10, 41]. Some recent studies, based on model species, test the molecular response to environmental change such as temperature, light, and presence of pathogens or pheromones [21, 42, 43] and found the adaptation to the environmental change to be regulated by protein phosphorylation. Protein kinases, such as those involved in the mitogen-activated protein kinase (MAPK) signaling pathway, mediate phosphorylation changes in other proteins and have been implicated in the control of synaptic plasticity in the adult brain [44]. In *T. delaisi,* we identified differentially expressed genes associated with social phenotypic plasticity whose gene functions are involved with the map kinases pathway. For instance, in the territorial male *madd* was upregulated with two different isoforms against the sneaker male whereas the sneaker male overexpressed *mapkapk3* against the territorial male (Supplementary material Table S1). In the female we find regulation of MAP kinase activity to be enriched (Fig.4), due to several isoforms of *trib2* being overexpressed towards the sneaker male, a gene that interacts with MAPK kinases and regulates activation of MAP kinases. This might be worthwhile analyzing more profoundly in future studies.

Annotation, meaning contigs for which a successful Blast hit was found associated with a biological function, was the highest for contigs overexpressed in sneaker males (55.5%, Table 1). For females and territorial males only about half of the overexpressed contigs could be annotated (27% and 28% respectively). The enriched GO-terms for the sneaker male are mostly involved in differentiation and development (Fig. 4). This could suggest that the elevated annotation success for the upregulated genes in sneaker males might be related to more functional descriptions of these developmental genes in the databases. On the other hand, over 70% of differentially expressed genes are left unannotated for the territorial male and the female and especially low annotation success was found for the upregulated genes in territorial males against sneaker males (Table 1). This might indicate that the genes involved in social phenotypic plasticity are not known.

Candidate genes in the context of social behavior

Proposed candidate genes from previous studies were searched for in our *de novo* transcriptome assembly (Table 2). The only candidate gene that showed significant differences in expression after correction for individual variation was the Somatostatin receptor type 1 (*sstr1*). Somatostatin is a neuropeptide also known as a growth hormone-inhibiting hormone and therefore commonly studied in the context of growth. In the African cichlid (*Astatotilapia burtoni*) somatostatin and somatostatin receptors have been shown to play a role in social behavior [45–47]. Somatostatin prepropeptide and somatostatin receptor type 3 (*sstr3*) were elevated in the dominant cichlid males in comparison to the subordinate males. For *T. delaisi*, there was no differential expression in these two genes although we found in *sstr1,* which was not measured in cichlids, significant differential expression between territorial males and females (with intermediate expression levels for the sneaker male). It is probable that this elevation in *sstr1* levels for males and especially territorial males is correlated with aggression. For cichlids, aggression levels were correlated with *sstr* (2 and 3) expression in the gonad and it has been demonstrated that somatostatin has a significant effect on aggressive behavior [46]. Nonetheless, this is the first evidence for the role of *sstr1* in regulation of social behavior and as previously suggested the effects of somatostatin on the regulation of growth and behavior are complex [46].

No other previously suggested candidate genes for social dominance were differentially expressed at a significant level in *Tripterygion delaisi* with the exception of the brain aromatase enzyme (*cyp19a1*), for which multiple isoforms were differentially expressed before adjustment for individual variation (Supplementary Material Table S2).

Cyp19a1 was found at higher levels in territorial males in comparison to sneaker males in other fish species [e.g. 15]. The brain aromatase is duplicated in fish and one form is expressed in the ovaries and the other one is expressed in the brain. In general, aromatase activity in the brain was found to be lower in castrated males than in non-castrated males of *Salmo salar* [48] and *cyp19a1* brain mRNA were lower in individuals with developing gonads in comparison to individuals with fully developed gonads in Atlantic croaker (*Micropogonias undulates*; [49]). In peacock blennies aromatase activity was also suppressed in sneaker males and elevated in nesting males [1]. Although this shows that brain aromatase clearly is an important enzyme involved in the regulation of social status between males in several fish species, individual

variation of the expression of this gene was high in *T. delaisi* thus we don't consider this gene differentially expressed (Supplementary Material Table S2). The five sneaker males all expressed the cyp19a1 isoforms at an equally low level, whereas two of the five territorial males show very high levels resulting in differential expression between the male types due to these outliers. Individual gene expression variation has previously been pointed out to be an important factor for phenotypic plasticity [15] and outlier expression might bias the outcome [32]. This emphasizes the need of non-pooled biological replicates even in genome wide studies.

Hence, individual variation was accounted for and the final set of differentially expressed genes between the phenotypes was again evaluated for individual variation by hierarchical clustering of expression patterns (Fig. 5, Fig. S3 & S4). For the 600 differentially expressed isoforms between the two male phenotypes, territorial males and sneaker males are clearly separated into two clusters illustrated by the hierarchical distance tree (Fig. 5), demonstrating that these genes are involved in phenotypic plasticity of different social statuses.

A commonly measured neuropeptide in relation to aggression, arginine vasotocin, is differentially expressed between males in several fish species, but not in *Tripterygion delaisi* (Table 2). In Atlantic salmon (*Salmon salar*) vasotocin is one of the key genes that is differentially expressed in the brain with down-regulation for the precocious male [22] as in the peacock blenny (*Salaria pavo*) where arginine vasotocin (*avt*) levels were detected to be higher in territorial nesting male in the forebrain [50]. In the African cichlid (*Astatotilapia burtoni) overexpression of avt was found overexpressed by the dominant male in the posterior preoptic area* and in the anterior preoptic area *avt* mRNA levels were higher in the non-dominant male *[51]*. Such regional expression differences were also observed in three-spined stickleback (*Gasterosteus aculeatus*) in relation to territoriality [52]. The fact that whole brain expression was measured in *T. delaisi* might mask actual expression differences between different brain regions. Nonetheless, Santangelo & Bass point out that there might be a species and context dependency of *avt* regulation across teleost species as seen for tetrapods [53].

The fact that most of the previously mentioned candidate genes were not differentially expressed in *T.delaisi* could be due to slight differences in reproductive social system. In African cichlids, the subordinate males have undeveloped testes and need to become territorial to reproduce [19],

which is distinct to the sneaker male in *T. delaisi* which has proportionally greater testes and is reproductively active [54]. In Atlantic Salmon precocious males and adult males show differential expression in some of the candidate genes but the two male types reproduce at different ages and the adult males do not settle down and defend a nest [23]. This suggests that there might be no general candidate genes but rather species-specific or context-specific candidate genes for social reproductive behavior.

Novel genes in the context of social behavior

By RNAseq analysis we uncovered a large set of novel genes differentially expressed during the reproductive period of *Tripterygion delaisi*. As previously mentioned, the majority of these contigs have no homologs or no functional description in the databases. Nonetheless, key genes for each of the phenotypes could be detected (Fig. 6). When comparing the female with the two types of males only three (annotated) genes are upregulated in females against both males. Downregulated in females, which means upregulated in both male phenotypes are two subunits of the CCR4-NOT transcription complex, which function as a general transcription regulation. The function of the CCR4-NOT complex is involved in all aspects of mRNA biogenesis from the transcription of RNA to its export [55]. This could be correlated with the lower levels or lower transcriptional activity in females, as less overexpression is found for females in general. *Col1a2*, which encodes for Collagen Type I alpha, is overexpressed in *T. delaisi* males. This gene is a well-studied gene in the context of bone development [56]. In the brain tissue, collagen can only be found in the blood vessels of the brain, which either suggests development of blood vessels or a possible transport of collagen to other organs for skeletal development. Both of these biological functions are enriched for the sneaker male (Fig. 4).

The genes defining the sneaker male in *T. delaisi* mostly have functions related to transport. *Rab33a*, *arf3* and *mfsd3* are associated with protein transport, protein trafficking and vesicle transport. *Klc1*, kinesin light chain 1, is responsible for organelle transport along microtubules.

The territorial male differentially overexpresses some genes related to synaptic plasticity, *Gpsm1*, a G-protein signaling modulator and *ncs1*, a neuronal calcium sensor, both involved in the activation or deactivation of the G-protein cascade. The role of *gpsm1* is to act as a regulatory sink of the availability and stability of the Gα component of the G-protein cycle [57]. This Gα component mediates signaling from vasoconstrictive hormone, such as vasopressin (homologue

for vasotocin in fish) [58], a previously mentioned candidate gene for aggression in fish [59]. The neuronal calcium sensor (*ncs1*) is sensitive to cytosolic Ca^{2+} changes and contributes to G-protein-coupled receptor desensitization and increases vesicle release in the presence of calcium. As *ncs1* was found in the dendroids in mice it may allow for locally regulated protein synthesis, which is linked to long-term synaptic plasticity [60]. The protein kinase Cδ (*prkcd*) is a recently-detected PKC isoform that plays critical roles in various cellular functions such as the control of growth, differentiation, and apoptosis [61]. In rodents, though, *prkcd* is a gene involved in signal transduction that is correlated with behavior [62]. *Aldh1l1* encodes for an enzyme from the aldehyde dehydrogenase family and is overexpressed for the territorial male. A gene from this family (*aldh9*) is one of the few genes that was associated with dominance of African cichlid males in a microarray study [15]. Although in cichlids this gene was expressed in lower levels for the dominant male against the sneaker male, it might be interesting to study this gene family directly in relation to behavior.

Importance of alternative splicing in the context of social dominance

The importance of alternatively spliced gene isoforms has long been accepted [34] and with the development of RNAseq a more detailed understanding is possible. Although most studies using RNAseq focus on expression on the gene level, expression of differentially spliced isoforms might vary even though overall gene expression might not. For humans 10% of the protein coding genes reveal population-specific splicing [63]. For *Tripterygion delaisi* one of the three overexpressed enriched biological functions found for the territorial male was nuclear mRNA splicing via the spliceosome indicating the elevated importance of exon joining and possibly alternative splicing in territorial males (Fig. 4). Nonetheless, for the differentially expressed genes in the context of social dominance, most alternatively spliced isoforms showed the same expression pattern (for details see gene tables in Supplementary material Table S3, S4, and S5). However, five genes of the differentially expressed and annotated gene set were expressed in an opposing manner in two or more isoforms between phenotypes. For the comparison between sneaker male and female opposing expression results were found for *prkar2b* and *c7orf51*; between the territorial male and female the genes were *rap1gap* and *macf1* and for the comparison between the territorial male and sneaker male one isoform of *phf20* was overexpressed in the territorial male and another isoform in the sneaker males. For example, in

the case of *rap1gap* RefSeq provided evidence that differentially spliced isoform transcript variants encode distinct proteins leading to different functions [e.g. 64]. These genes would have been overlooked as not differentially expressed if expression at the gene level was considered. With advances of analysis methods, such as the assembly of full length of alternative spliced isoforms [31] and models accommodating isoform expression estimates uncertainties [65], RNAseq has allowed estimating differential gene expression in isoforms in our species. Even though more work is needed to fully understand the precision and accuracy as well as the possibilities and limitations of the methodologies used, this approach will allow for a huge range of molecular studies on non-model species of evolutionary and ecological importance.

Conclusions

Phenotypic plasticity resulted to be a more important factor in differential gene expression than sexual dimorphism as more genes were significantly expressed at different levels between the two male phenotypes than between males and females. Previously suggested candidate genes to social dominance in the context of alternative mating strategies seem to be mostly species-specific and here we present a list of novel genes which are differentially expressed in *Tripterygion delaisi*. The genes that were differentially expressed for the territorial male were mainly related to synaptic plasticity possibly indicating the drastic change in behavior and phenotype. The sneaker overexpresses genes associated with differentiation and development. This result suggests that although this type of male is reproductively active, it is not fully developed. This is the first study looking at transcriptome data and differential expression for a non-model species (*Tripterygion delaisi*) in the context of alternative mating strategies and provides essential information for further studies investigating the molecular basis of social dominance. Overall, RNAseq has proven to be a useful tool for the analysis of ecological and evolutionary questions for non-model species.

Materials and Methods

Sample Collection

Territorial males, sneaker males and females of *Tripterygion delaisi* were collected in June 2010 on the rocky shore of the Costa Brava close to the town of Blanes (41°67'N,2°30'E) in the northwest Mediterranean Sea. 15 individuals (five territorial males, five sneaker males and five females) were collected for individual expression analysis. Eight additional individuals (three territorial males, three females and two sneaker males) were collected and used as a pooled sample for the *de novo* transcriptome assembly. All specimens were caught on the same day with small nets and put into large containers for transport back to the laboratory under the same conditions. Territorial males were collected from their nests and females and sneaker males were collected in the surrounding area. Thus we can assume that individuals of each phenotype were in the same reproductive stage and can be considered biological replicates. Fish were euthanized immediately arriving to the laboratory, all individuals still showing the same phenotypic coloration, snap frozen in liquid nitrogen, and stored at −80°C. The sex was double-checked especially for sneaker males and females, as they are phenotypically similar, by verifying the presence of ovaries or testes. The method of euthanasia and all field and experimental procedures followed the Spanish Laws (Royal Executive Order, 53/2013) for Animal Experimentation, in accordance with the European Union directive (2010/63/UE).

Total RNA extraction and cDNA library construction

Fish brains were dissected out of the frozen heads, weighed and placed in TRI Reagent. All *T. delaisi* brains weighed between seven and twenty milligrams. Tissues were homogenized in a Retsch homogenizer (MM200) at 25Hz for two intervals of 30 seconds. Phase separation was done with BCP Phase Separation Reagent (1-Bromo-3-Chloropropane) and all centrifuge steps were performed at 4°C. The RNA was precipitated with isopropanol and the total RNA pellet was washed twice with 75% ethanol to then be dissolved in TE Buffer pH 8.0.

Poly-A mRNA was purified using Dynabeads (Invitrogen) coated in $Oligo(dT)_{25}$ following the manufacturer's protocol, but adding a second wash-step. First-strand cDNA was synthesized with Random Hexamer Primers and SuperScript reverse transcriptase and second strand cDNA synthesis was performed using RNaseH and DNA Polymerase I. cDNA was then purified with the QIAquick PCR purification kit and fragmented with dsDNA fragmentase (New England

Biolabs) for 28 minutes at 37°C to yet again be purified with the purification kit. Fragments were prepared for Illumina sequencing with New England Biolabs reagents following the manufacturer's instructions for each reagent module. Firstly, ends were repaired with the End Repair module using E.coli ligase and End Repair enzyme mix. A single A-base was then added using the dA-Tailing module with Klenow DNA polymerase. Homemade four-mer barcodes were added with the Quick T4 DNA ligase to be able to multiplex the samples in the same Illumina lane. After each step the DNA was purified either with the PCR purification kit or the Qiagen MinElute PCR purification kit. Subsequently, 200-300bp fragments were gel purified in Invitrogen ultra-pure agarose and cleaned with the QIAquick gel extraction kit. Finally, cDNA fragments were enriched using Phusion polymerase for 15 cycles and once more purified with the PCR purification kit with a final volume of 30µl. Concentration and purity was measured several times throughout the process with the Agilent 2100 Bioanalyzer (Agilent RNA 6000 pico & DNA 1000 Kit). For each Illumina sequencing lane six barcoded prepared cDNA libraries were then mixed at equal concentrations to reach a final molarity of 10nM. Three individuals (one territorial male, one sneaker male and one female) as well as a normalized pooled library of 8 mixed samples (three territorial males, three females and two sneaker males) were sequenced at the length of 109bp, single-end, for a more solid base for the transcriptome assembly. The other twelve samples (four territorial males, four sneaker males and four females) were sequenced at a length of 52bp, also single end. Sequencing of the individuals used for the expression analysis was conducted on an Illumina Hiseq 2000 and raw fastq generation was performed at the FAS Center for systems biology at Harvard University (USA). Sequencing of the eight individuals' pooled-sample was done at DNAVision S.A. (Belgium).

De novo Transcriptome assembly and annotation

Reads sequenced at the length of 109bp were sorted by their individual sample barcode and the four-mer barcodes were removed. Read quality was checked and visualized with FastQC (Andrews 2010) and low quality reads were eliminated or trimmed in CLC Genomics Workbench 4.7 so that all base reads were superior to the Phred quality score of 35. Reads with the length below 20bp were removed. *De novo* assembly was performed with Trinity [66] , which allows for the detection of differentially spliced contig isoforms, using the default

program settings and contigs shorter than 200bp were eliminated. The transcripts/contigs were then annotated against the UniProt protein database using BLASTx and an E-value cut-off of $1x10^{-10}$. Furthermore, transcripts were also compared to the annotated proteins available on the NCBI Unigene database of *Danio rerio*.

The de novo assembly contigs were divided into two sets: 'With Blast', which are the contigs that resulted in a successful blast hit and 'without blast', which are the contigs without homology in the protein databases. Both sets were compared by measuring various parameters. Open reading frame presence and length was measured by using the getorf tool in EMBOSS [67]. We detected protein families homology with hidden Markov models implemented in the program Hmmscan3 by comparing our contigs to the PfamA protein family database [68]. To detect the protein coding potential of the contigs we used the program CPAT [37] and accepted contigs with a potential above 0.5 to be 'more likely' a protein coding. For this we used the zebrafish databases implemented in the program.

RNAseq analysis and differential expression

The transcriptome contigs were used as a reference for the evaluation of the expression values. The three samples (one territorial male, one sneaker male and one female) of 109bp length were cut down to the maximum length of 48bp to avoid a length bias in the expression value calculation, as the other 12 samples (four territorial males, four sneaker males and four females) used for differential expression analysis were sequenced at a shorter length. Quality trimmed reads for each of the 15 individuals were mapped against the reference with Bowtie as a short read aligner [69] and then processed with RSEM [70] which accurately quantifies transcripts and maximizes the use of ambiguously mapped reads. Differential expression values were computed with EBseq by using a Bayesian approach to accurately estimate isoform expression [65]. Three comparisons were performed to find the particular genes which distinguish each phenotype: territorial males versus sneaker males, territorial males versus females and sneaker males versus females. Count data were normalized by estimating a scaling factor for each contig in EBseq, which has been demonstrated to be a very robust method [71], and then data dispersion was

evaluated. It was tested for differences between the normalized base means with an empirical Bayes hierarchical model resulting in Posterior probabilities of differential expression. Comparisons were accepted to be significant at an FDR adjusted value of 0.05 which is a Posterior probability of 0.95 or higher. To avoid outlier expression bias due to great individual variation, we accepted only contigs for which standard deviation was smaller than the mean expression value within phenotype (SD<Mean). For visualization of the significant comparisons, heatmaps of the significant genes after FDR adjustment were produced with the heatmaps2 package in R. Hierarchical clustering of individual samples with 1000 bootstrap replications was performed with the R package pvclust [72] and heatmaps were sorted accordingly. Gene clusters were determined with k-means in R.

Functional annotation

The *de novo* assembled transcriptome was annotated with Blast2GO [39]. We performed several enrichment analyses. Firstly, the top 1% contigs which were expressed at the highest level regardless of phenotype were compared with the whole transcriptome via a Fisher's Exact test. Enriched GO-terms were then slimmed in REVIGO and treemaps produced [73]. Secondly, enrichment analyses were performed for the differentially expressed genes for the three comparisons by testing the upregulated genes in one phenotype against upregulation in the other phenotype.

Genes which were repeatedly mentioned as important in multiple studies focusing on alternative reproductive strategies in fish or were stated to be 'candidate genes' in the context of social dominance, were searched for in our transcriptome. The sequences of the candidate genes *gnrh* (gonadotropin releasing hormone receptor; [15, 19]), *epd* (ependymin; [24]), *avp* (arginine vasopressin; [22, 51, 59, 74]), *somatostatin* ([15]), *egr*1 (early growth response protein, [19]), *galn* (galanin, [15]) and *cyp19a1* (brain aromatase enzyme, [15]) were Blasted (Blastn) against the *de novo* assembly contigs using Blast 2 sequence in NCBI.

Availability of Supporting Data

Data have been deposited at the National Center for Biotechnology Information (NCBI) Transcriptome Shotgun Assembly database under BioProject PRJNA186408. Raw sequence reads can be found in the SRA database.

Acknowledgements

We would like to thank Brain Haas for his help with the *de novo* transcriptome assembly and Dan Barshis for very helpful input on the analysis of the RNA-seq data. Also, we are thankful to Cinta Pegueroles Queralt for her input on the manuscript. This work was partially funded by the Spanish Ministry of Science and Innovation (CTM2010-22218-C02-01) project. The authors are part of the research groups 2009SGR-636 and 2009SGR-665 of the Generalitat de Catalunya.

Competing interests

The authors declare that they have no competing interests.

Author contributions

CS, SV, EM & MP conceived the idea and designed the research. CS & EM conducted the field work. CS conducted the lab work with help of SV. CS & MP analyzed the results and CS wrote the paper with support and comments by all the authors.

References

1. Gonçalves D, Teles M, Alpedrinha J, Oliveira RF: **Brain and gonadal aromatase activity and steroid hormone levels in female and polymorphic males of the peacock blenny Salaria pavo.** *Hormones and behavior* 2008, **54**:717–25.

2. Alonzo SH, Warner RR: **A trade-off generated by sexual conflict: Mediterranean wrasse males refuse present mates to increase future success**. *Behavioral Ecology* 1999, **10**:105–111.

3. Munehara H, Takenaka O: **Microsatellite markers and multiple paternity in a paternal care fish, Hexagrammos otakii**. *Journal of Ethology* 2000, **18**:101–104.

4. Taborsky M: **Alternative reproductive tactics in fish**. In *Alternative reproductive tactics: an integrative approach*. Edited by Oliveira RF, Taborsky M BH. Cambridge University Press, Cambridge; 2008:251–299.

5. Taborsky M: **Sneakers, Satellites, and Helpers: Parasitic and Cooperative Behavior in Fish Reproduction**. *Advances in the Study of Behavior* 1994, **23**:1–100.

6. Taborsky M: **Sperm competition in fish: `bourgeois' males and parasitic spawning**. *Trends in Ecology & Evolution* 1998, **13**:222–227.

7. Immler S, Mazzoldi C, Rasotto MB: **From sneaker to parental male: change of reproductive traits in the black goby, Gobius niger (Teleostei, Gobiidae).** *Journal of experimental zoology Part A, Comparative experimental biology* 2004, **301**:177–85.

8. Wirtz P: **The Behaviour of the Mediterranean Tripterygion Species (Pisces, Blennioidei)**. *Zeitschrift für Tierpsychologie* 1978, **48**:142–174.

9. Jonge J, Videler JJ: **Differences between the reproductive biologies of Tripterygion tripteronotus and T. delaisi (Pisces, Perciformes, Tripterygiidae): the adaptive significance of an alternative mating strategy and a red instead of a yellow nuptial colour**. *Marine Biology* 1989, **100**:431–437.

10. Pigliucci M: **Evolution of phenotypic plasticity: where are we going now?** *Trends in Ecology & Evolution* 2005, **20**:481–6.

11. Lande R: **Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation.** *Journal of evolutionary biology* 2009, **22**:1435–46.

12. Thompson JD: **Phenotypic plasticity as a component of evolutionary change.** *Trends in ecology & evolution* 1991, **6**:246–9.

13. Thomas ML, Simmons LW: **Short-term phenotypic plasticity in long-chain cuticular hydrocarbons.** *Proceedings Biological sciences / The Royal Society* 2011, **278**:3123–8.

14. Gabriel W: **How stress selects for reversible phenotypic plasticity.** *Journal of evolutionary biology* 2005, **18**:873–83.

15. Renn SCP, Aubin-Horth N, Hofmann H a: **Fish and chips: functional genomics of social plasticity in an African cichlid fish.** *The Journal of experimental biology* 2008, **211**(Pt 18):3041–56.

16. Maruska KP, Fernald RD: **Behavioral and physiological plasticity: rapid changes during social ascent in an African cichlid fish.** *Hormones and behavior* 2010, **58**:230–40.

17. Robinson GE, Fernald RD, Clayton DF: **Genes and social behavior.** *Science (New York, NY)* 2008, **322**:896–900.

18. Whitfield CW, Cziko A-M, Robinson GE: **Gene expression profiles in the brain predict behavior in individual honey bees.** *Science (New York, NY)* 2003, **302**:296–9.

19. Burmeister SS, Jarvis ED, Fernald RD: **Rapid behavioral and genomic responses to social opportunity.** *PLoS biology* 2005, **3**:e363.

20. Fernald RD: **Social control of the brain.** *Annual review of neuroscience* 2012, **35**:133–51.

21. Aubin-Horth N, Renn SCP: **Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity.** *Molecular Ecology* 2009, **18**:3763–80.

22. Guiry A, Flynn D, Hubert S, O'Keeffe AM, LeProvost O, White SL, Forde PF, Davoren P, Houeix B, Smith TJ, Cotter D, Wilkins NP, Cairns MT: **Testes and brain gene expression in precocious male and adult maturing Atlantic salmon (Salmo salar).** *BMC genomics* 2010, **11**:211.

23. Aubin-Horth N, Landry CR, Letcher BH, Hofmann H a: **Alternative life histories shape brain gene expression profiles in males of the same population.** *Proceedings Biological sciences / The Royal Society* 2005, **272**:1655–62.

24. Sneddon LU, Schmidt R, Fang Y, Cossins AR: **Molecular correlates of social dominance: a novel role for ependymin in aggression.** *PloS one* 2011, **6**:e18181.

25. Sneddon LU, Margareto J, Cossins AR: **The use of transcriptomics to address questions in behaviour: production of a suppression subtractive hybridisation library from dominance hierarchies of rainbow trout.** *Physiological and biochemical zoology : PBZ* 2012, **78**:695–705.

26. Carreras-Carbonell J, Macpherson E, Pascual M: **Population structure within and between subspecies of the Mediterranean triplefin fish Tripterygion delaisi revealed by highly polymorphic microsatellite loci.** *Molecular Ecology* 2006, **15**:3527–39.

27. Carreras-Carbonell J, Macpherson E, Pascual M: **Rapid radiation and cryptic speciation in mediterranean triplefin blennies (Pisces: Tripterygiidae) combining multiple genes.** *Molecular Phylogenetics and Evolution* 2005, **37**:751–61.

28. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nature reviews Genetics* 2009, **10**:57–63.

29. Ekblom R, Galindo J: **Applications of next generation sequencing in molecular ecology of non-model organisms.** *Heredity* 2011, **107**:1–15.

30. **WoRMS - World Register of Marine Species** [http://www.marinespecies.org/]

31. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nature biotechnology* 2011, **29**:644–52.

32. Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR: **Genomic basis for coral resilience to climate change.** *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**:1387–92.

33. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes.** *Nature* 2008, **456**:470–6.

34. Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H: **Function of alternative splicing.** *Gene* 2005, **344**:1–20.

35. Modrek B, Lee C: **A genomic view of alternative splicing.** *Nature genetics* 2002, **30**:13–9.

36. Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldon T, Guigo R, Sumner S: **Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes.** *Genome biology* 2013, **14**:R20.

37. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W: **CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model.** *Nucleic acids research* 2013:gkt006–.

38. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R: **The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.** *Genome research* 2012, **22**:1775–89.

39. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics (Oxford, England)* 2005, **21**:3674–6.

40. Suárez-Castillo EC, García-Arrarás JE: **Molecular evolution of the ependymin protein family: a necessary update.** *BMC evolutionary biology* 2007, **7**:23.

41. Oliveira RF: **Social plasticity in fish: integrating mechanisms and function**. *Journal of Fish Biology* 2012:no–no.

42. Bent AF: **Plant mitogen-activated protein kinase cascades: Negative regulatory roles turn out positive.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:784–6.

43. Oliveira RF: **Social plasticity in fish: integrating mechanisms and function**. *Journal of Fish Biology* 2012:no–no.

44. Thomas GM, Huganir RL: **MAPK cascade signalling and synaptic plasticity.** *Nature reviews Neuroscience* 2004, **5**:173–83.

45. Trainor BC, Hofmann HA: **Somatostatin and somatostatin receptor gene expression in dominant and subordinate males of an African cichlid fish.** *Behavioural brain research* 2007, **179**:314–20.

46. Trainor BC, Hofmann HA: **Somatostatin regulates aggressive behavior in an African cichlid fish.** *Endocrinology* 2006, **147**:5119–25.

47. Hofmann HA, Fernald RD: **Social status controls somatostatin neuron size and growth.** *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2000, **20**:4740–4.

48. Mayer I, Borg B, Berglund I, Lambert JGD: **Effects of castration and androgen treatment on aromatase activity in the brain of mature male Atlantic salmon (Salmo salar L.) parr**. *General and Comparative Endocrinology* 1991, **82**:86–92.

49. Nunez BS, Applebaum SL: **Tissue- and sex-specific regulation of CYP19A1 expression in the Atlantic croaker (Micropogonias undulatus).** *General and comparative endocrinology* 2006, **149**:205–16.

50. Grober MS, George A a, Watkins KK, Carneiro L a, Oliveira RF: **Forebrain AVT and courtship in a fish with male alternative reproductive tactics.** *Brain research bulletin* 2002, **57**:423–5.

51. Greenwood AK, Wark AR, Fernald RD, Hofmann H a: **Expression of arginine vasotocin in distinct preoptic regions is associated with dominant and subordinate behaviour in an African cichlid fish.** *Proceedings Biological sciences / The Royal Society* 2008, **275**:2393–402.

52. Sanogo YO, Band M, Blatti C, Sinha S, Bell AM: **Transcriptional regulation of brain gene expression in response to a territorial intrusion.** *Proceedings Biological sciences / The Royal Society* 2012, **279**:4929–38.

53. Santangelo N, Bass AH: **New insights into neuropeptide modulation of aggression: field studies of arginine vasotocin in a territorial tropical damselfish.** *Proceedings Biological sciences / The Royal Society* 2006, **273**:3085–92.

54. Jonge J, Ruiter AJH, Hurk R: **Testis-testicular gland complex of two Tripterygion species (Blennioidei, Teleostei): differences between territorial and non-territorial males**. *Journal of Fish Biology* 1989, **35**:497–508.

55. Kerr SC, Azzouz N, Fuchs SM, Collart MA, Strahl BD, Corbett AH, Laribee RN: **The Ccr4-Not complex interacts with the mRNA export machinery.** *PloS one* 2011, **6**:e18302.

56. Karsenty G, Wagner EF: **Reaching a genetic and molecular understanding of skeletal development.** *Developmental cell* 2002, **2**:389–406.

57. Siderovski DP, Willard FS: **The GAPs, GEFs, and GDIs of heterotrimeric G-protein alpha subunits.** *International journal of biological sciences* 2005, **1**:51–66.

58. Kimple AJ, Soundararajan M, Hutsell SQ, Roos AK, Urban DJ, Setola V, Temple BRS, Roth BL, Knapp S, Willard FS, Siderovski DP: **Structural determinants of G-protein alpha subunit selectivity by regulator of G-protein signaling 2 (RGS2).** *The Journal of biological chemistry* 2009, **284**:19402–11.

59. Larson ET, O'Malley DM, Melloni RH: **Aggression and vasotocin are associated with dominant-subordinate relationships in zebrafish.** *Behavioural brain research* 2006, **167**:94–102.

60. Paterlini M, Revilla V, Grant AL, Wisden W: **Expression of the neuronal calcium sensor protein family in the rat brain.** *Neuroscience* 2000, **99**:205–16.

61. Kikkawa U, Matsuzaki H, Yamamoto T: **Protein Kinase C (PKC): Activation Mechanisms and Functions**. *Journal of Biochemistry* 2002, **132**:831–839.

62. Berger A, Roberts MA: **Dietary Effects of Arachidonate-Rich Fungal Oil and Fish Oil on Murine Hippocampal Gene Expression**. In *Unraveling Lipid Metabolism with Microarrays*. Edited by Berger A, Roberts MA. New York: Marcel Dekker; 2005:69–94.

63. Gonzàlez-Porta M, Calvo M, Sammeth M, Guigó R: **Estimation of alternative splicing variability in human populations.** *Genome research* 2012, **22**:528–38.

64. Mochizuki N, Ohba Y, Kiyokawa E, Kurata T, Murakami T, Ozaki T, Kitabatake A, Nagashima K, Matsuda M: **Activation of the ERK/MAPK pathway by an isoform of rap1GAP associated with G alpha(i)**. *Nature* 1999, **400**:891–4.

65. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, Haag JD, Gould MN, Stewart RM, Kendziorski C: **EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments.** *Bioinformatics (Oxford, England)* 2013:btt087–.

66. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nature biotechnology* 2011, **29**:644–52.

67. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends in genetics : TIG* 2000, **16**:276–7.

68. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD: **The Pfam protein families database.** *Nucleic acids research* 2012, **40**(Database issue):D290–301.

69. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome biology* 2009, **10**:R25.

70. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN: **RNA-Seq gene expression estimation with read mapping uncertainty.** *Bioinformatics (Oxford, England)* 2010, **26**:493–500.

71. Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G, Jagla B, Jouneau L, Laloë D, Le Gall C, Schaëffer B, Le Crom S, Guedj M, Jaffrézic F: **A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis.** *Briefings in bioinformatics* 2012:bbs046–.

72. Suzuki R, Shimodaira H: **Pvclust: an R package for assessing the uncertainty in hierarchical clustering.** *Bioinformatics (Oxford, England)* 2006, **22**:1540–2.

73. Supek F, Bošnjak M, Škunca N, Šmuc T: **REVIGO summarizes and visualizes long lists of gene ontology terms.** *PloS one* 2011, **6**:e21800.

74. Miranda J a., Oliveira RF, Carneiro L a., Santos RS, Grober MS: **Neurochemical correlates of male polymorphism and alternative reproductive tactics in the Azorean rock-pool blenny, Parablennius parvicornis**. *General and Comparative Endocrinology* 2003, **132**:183–189.

Tables

Table 1: Number of significantly expressed contigs and percentage of annotated genes between phenotypes and for a given phenotype with all comparisons combined.

	Total	**Annotated**
TM>SM	360	21.90%
SM>TM	240	54.20%
TM>F	144	43.10%
F>TM	104	28.80%
SM>F	162	57.40%
F>SM	177	26.00%
	Over-expressed	**Annotated**
TM	504	28.00%
SM	402	55.50%
F	281	27.00%

FDR adjusted significance value of 0.05. > indicates over-expression in the phenotype on the left. TM=territorial male, SM=sneaker male, F=female.

Figures

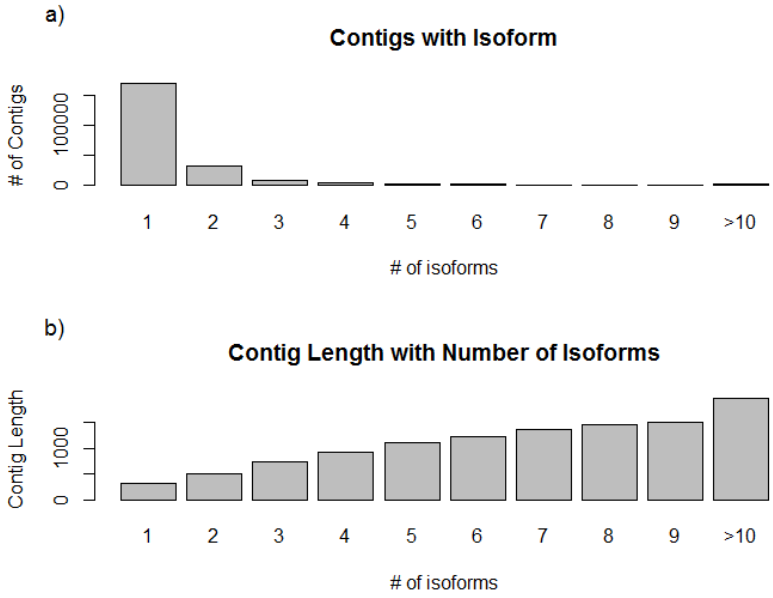

Fig.1: Frequency distribution of isoforms detected in the de novo assembly of the reference transcriptome. (a) Amount of contigs with different number of isoforms (b) length of contigs with different number of isoforms

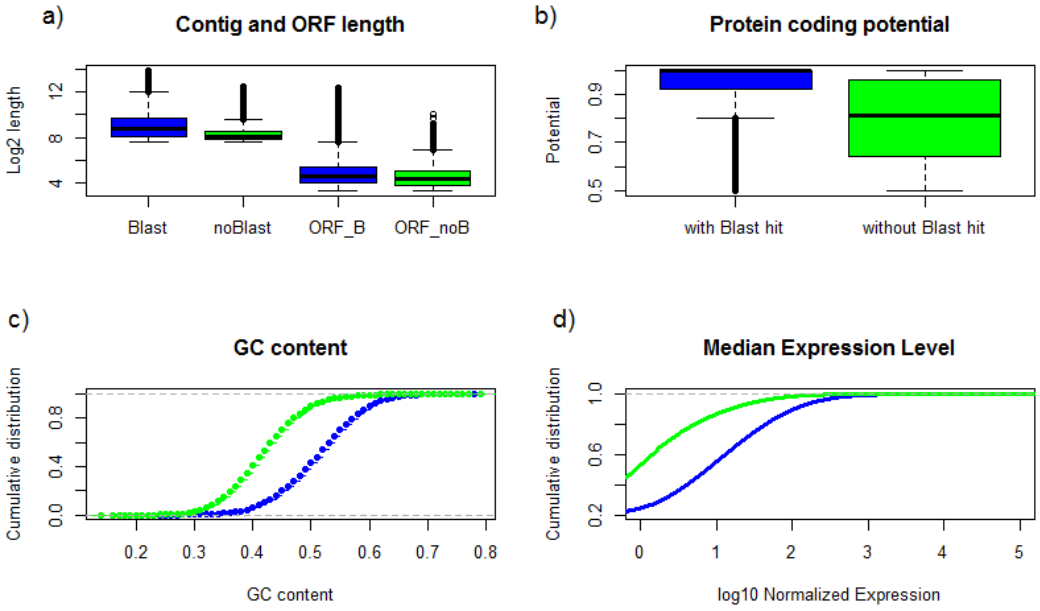

Fig.2: Comparison between the contigs with Blast-hits and the set of contigs without Blasthit. a) Overall length and ORF length, b) Protein coding potential determined by CPAT, c) GC content per contigs, d) Overall normalized expression.

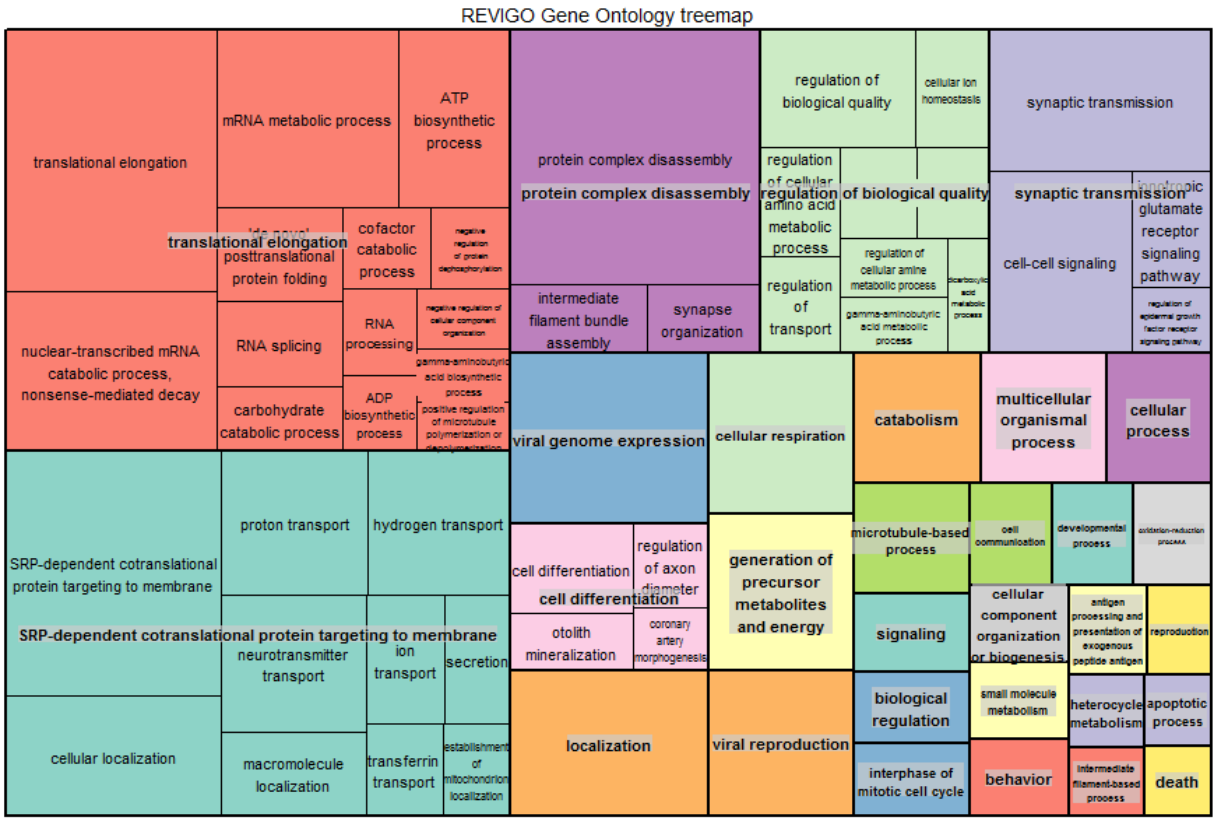

Fig.3: Gene Ontology treemap for the 1% highly expressed genes regardless of phenotype. The box size correlates to the –log10 p-value of the GO-term enrichment. Boxes with the same color can be grouped together and correspond to the same upper-hierarchy GO-term which is found in the middle of each box.

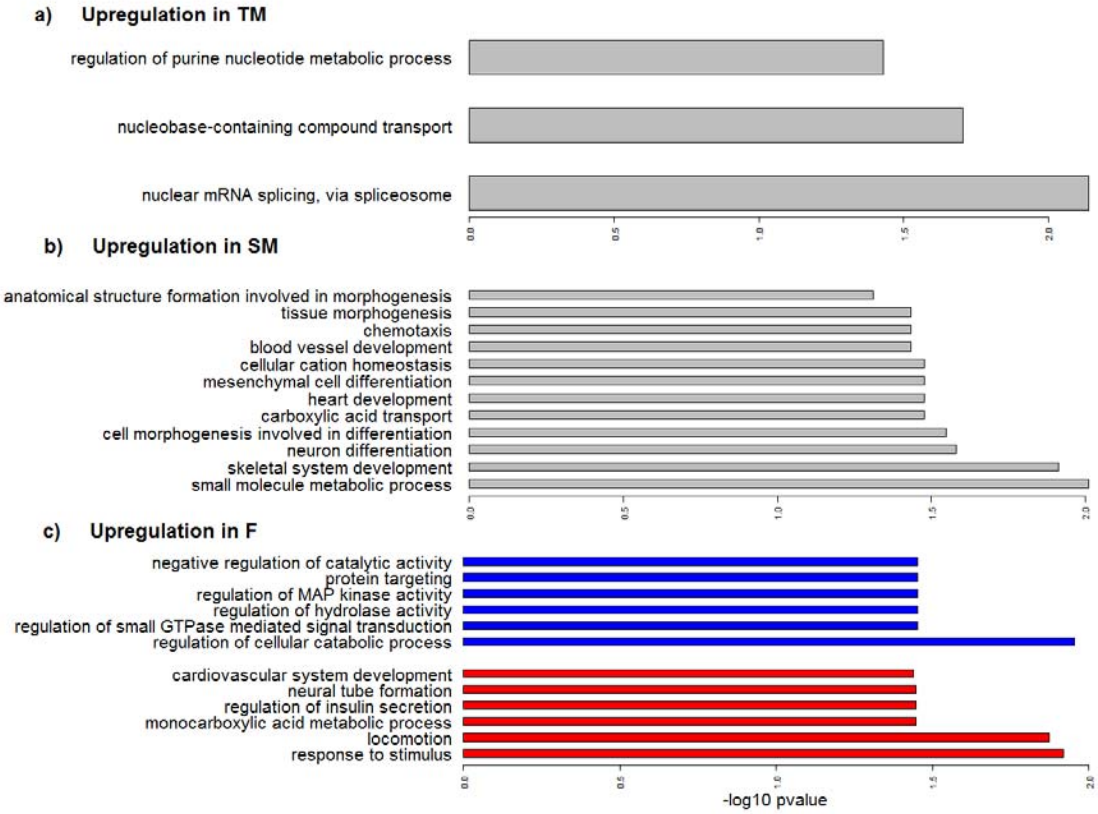

Fig.4: Barcharts representing the enriched biological processes associated with the upregulated genes in SM and TM (a, b), whereas no enrichment was found with upregulation against females. Upregulation in females (c) against the territorial male (red) and sneaker male (blue).

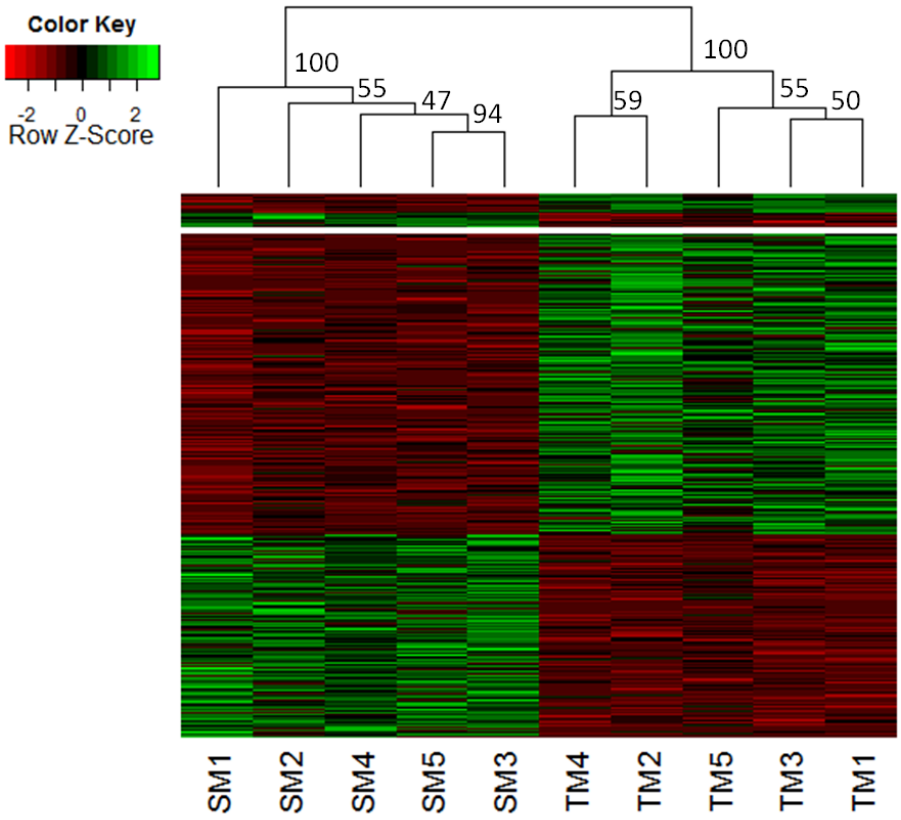

Fig.5: Heatmap comparing significant differentially expressed contigs, either annotated or not annotated, between five sneaker males and five territorial males. Intensity of color indicates expression levels. Similarity in expression patterns between genes is represented by kmeans clustering separating highly expressed genes above the white line and less expressed genes below. Similarity between individuals with hierarchical clustering can be seen above the heatmap with bootstraps.

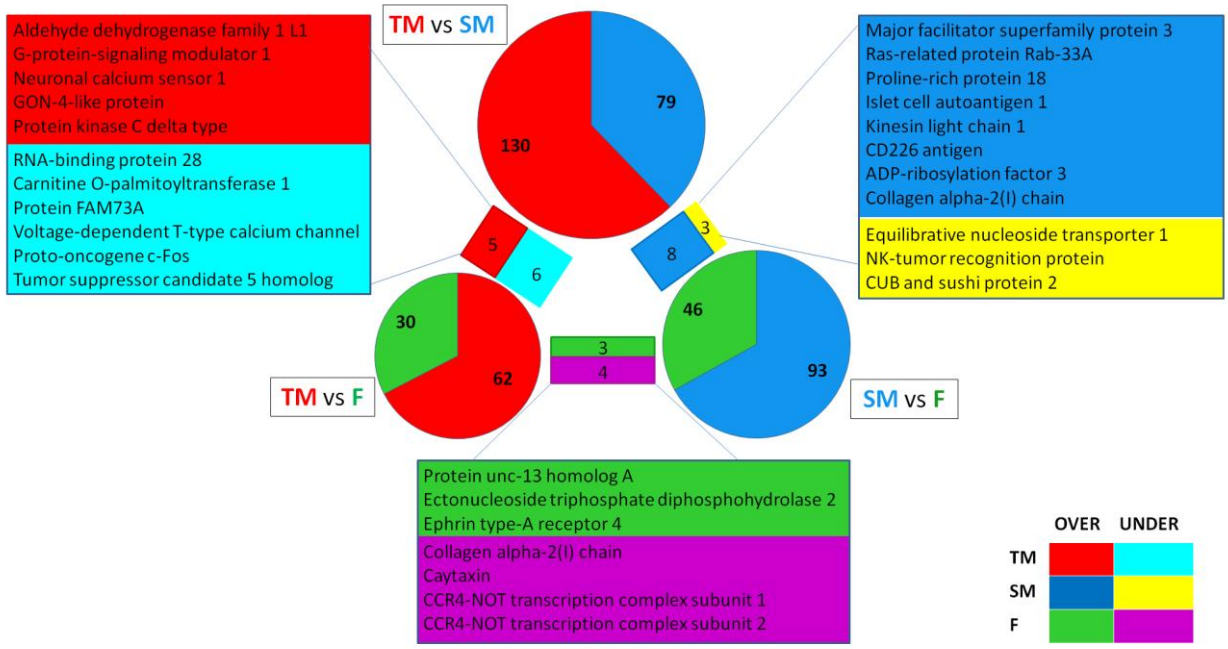

Fig.6: Venn diagram of the three comparisons: territorial male versus sneaker male (TM vs SM), territorial male versus female (TM vs F) and sneaker male versus female (SM vs F). The circle size is scaled to the number of differentially expressed genes for each comparison. Bars between piecharts and the corresponding gene lists identify contigs which are significantly expressed in adjacent comparisons and therefore representative of one phenotype. Colors identify the number of genes over or under expressed in the different phenotypes. In red over-expression in territorial male, in blue over-expression in sneaker male and in green over-expression in female. Significant under-expression in territorial males is shown in turquoise, pink are the genes under-expressed in females and yellow are under-expressed in sneaker males.

Supplementary material

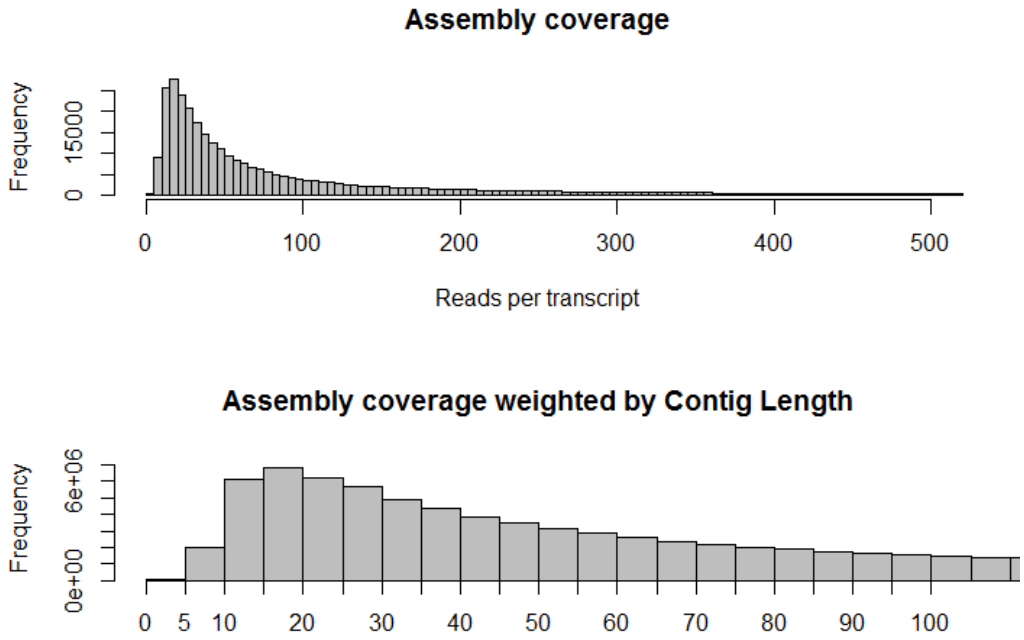

Fig.S1: Read coverage for the de novo transcriptome assembly contigs (above) and weighted frequency distribution of coverage by the length of the transcriptome contigs (below).

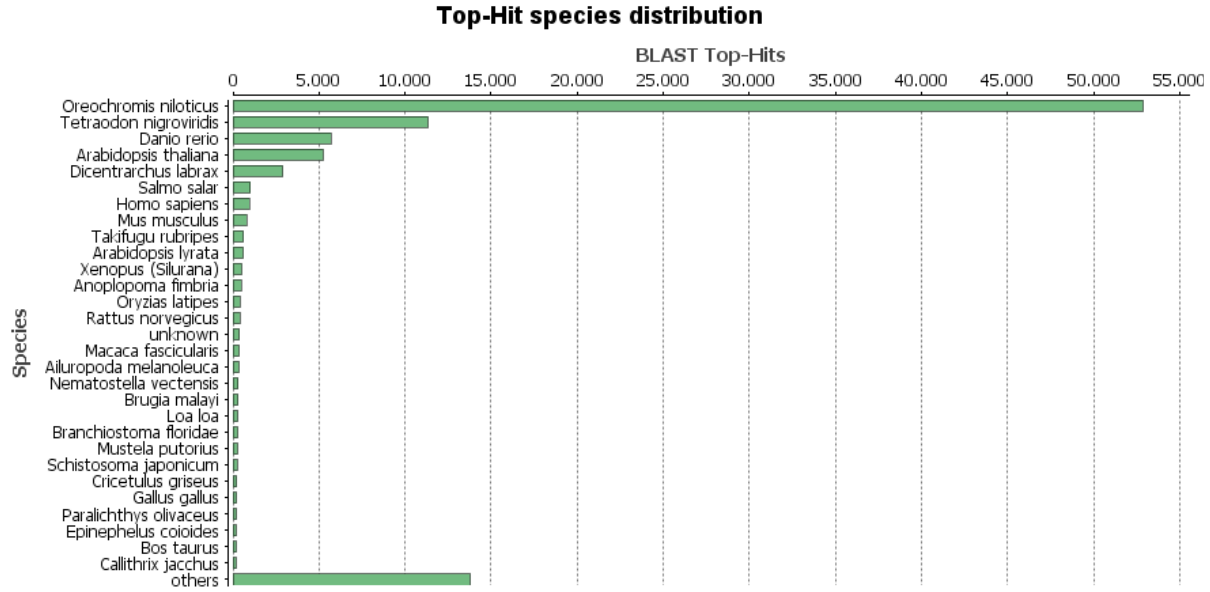

Fig S2: Top-hit species distribution for the Blast resultsof *T. delaisi* de novo transcriptome assembly.

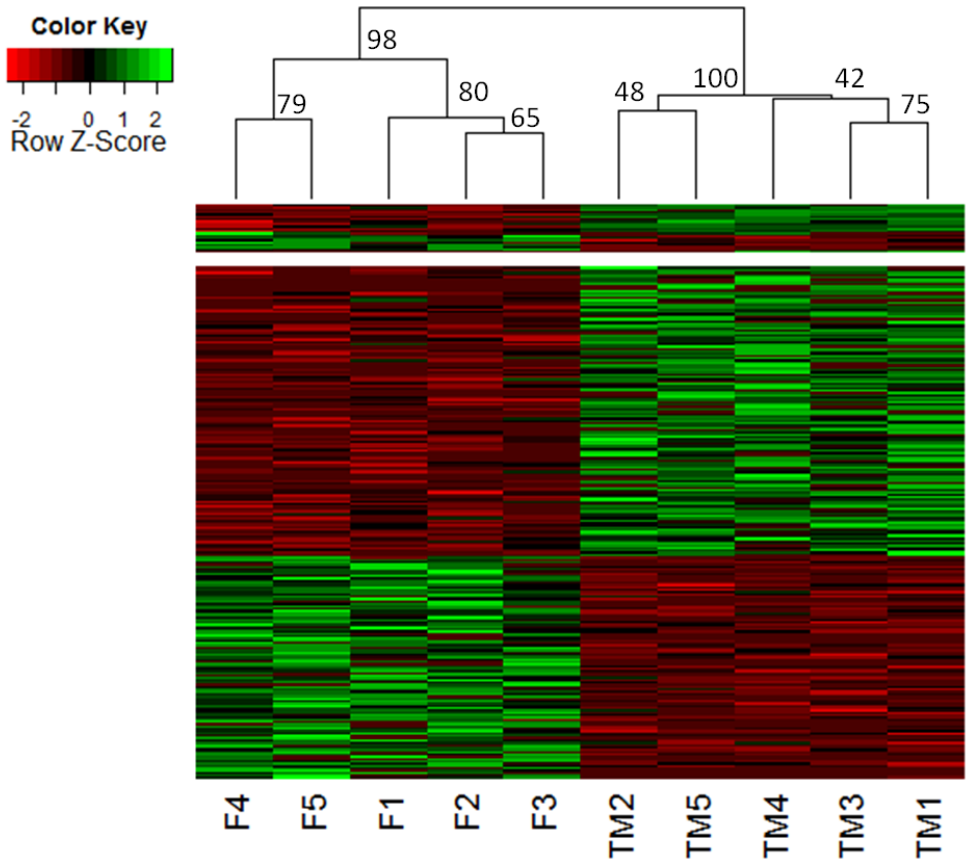

Fig.S3: Heatmap comparing significant differentially expressed contigs between five females and five territorial males. Intensity of color indicates expression levels. Similarity in expression patterns between genes is represented by kmeans clustering separating highly expressed genes above the white line and less expressed genes below. Similarity between individuals with hierarchical clustering can be seen above the heatmap with bootstraps.

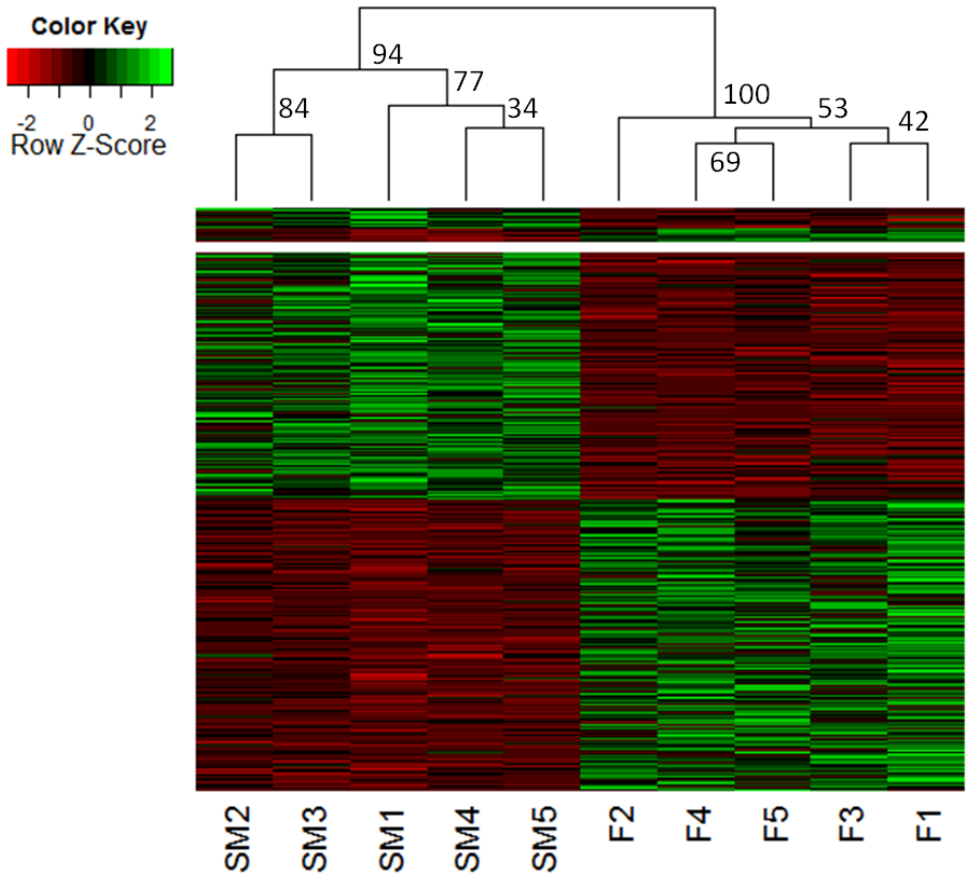

Fig.S4: Heatmap comparing significant differentially expressed contigs between five females and five sneaker males. Intensity of color indicates expression levels. Similarity in expression patterns between genes is represented by kmeans clustering separating highly expressed genes above the white line and less expressed genes below. Similarity between individuals with hierarchical clustering can be seen above the heatmap with bootstraps.

3.3 Fine scale genetic connectivity via parentage analysis

3.3 Fine scale genetic connectivity via parentage analysis

Publication 4: SNP development from RNA-seq data in a non-model fish: how many individuals are needed for accurate allele frequency prediction?

Publicación 4: Desarrollo de SNPs a partir de datos de RNAseq en una especie de pez no-modelo: ¿cuántos individuos son necesarios para la predicción de una frecuencia alélica correcta?

RESUMEN

Los polimorfismos de nucleótidos simples (SNPs) se están estableciendo como los marcadores preferidos en genética de poblaciones debido a una serie de ventajas en relación a otros marcadores, incluyendo su alta densidad en el genoma, la calidad de los datos, la reproducibilidad y la eficiencia en el genotipado, así como la facilidad de transferencia entre laboratorios. Los avances en la tecnología de secuenciación y los métodos para reducir la representación genómica, permiten un aislamiento factible de SNPs para especies no-modelo. La técnica por RNA-seq es una de las utilizadas para detectar SNPs y para el desarrollo de marcadores para un genotipado a gran escala. En este estudio, presentamos el desarrollo de 192 marcadores de tipo SNP en *Tripterygion delaisi,* un pez de las costas rocosas del Mediterráneo. Se obtuvieron datos de RNA-seq para 15 individuos y se utilizaron para detectar SNPs aplicando varios criterios de selección. Asimismo, se obtuvieron genotipos de 1599 individuos de la misma población con los loci resultantes. Se encontraron diferencias en heterocigosidad y frecuencias alélicas entre los dos conjuntos de datos. La heterocigosidad fue, en promedio, más baja en la muestra poblacional y la diferencia media entre la frecuencia de los alelos de los dos conjuntos fue de 0,135± 0,100. Se realizó un remuestreo por "bootstrap" con los datos de secuencias para predecir el tamaño de muestra apropiado para la detección de SNPs. Debido a que la producción de bibliotecas de cDNA consume mucho tiempo y recursos económicos, sugerimos que el uso de siete individuos para la secuenciación de RNA reduce la probabilidad de descartar SNPs altamente informativos debido a la falta de polimorfismos, y que más de 12 muestras no mejora considerablemente la predicción de la frecuencia alélica y, por lo tanto, se considera como un tamaño de muestra óptimo.

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) doi: 10.1111/1755-0998.12155

SNP development from RNA-seq data in a nonmodel fish: how many individuals are needed for accurate allele frequency prediction?

C. SCHUNTER,*† J. C. GARZA,‡ E. MACPHERSON* and M. PASCUAL†
*Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Car. Acc. Cala St. Francesc 14, Blanes 17300, Spain, †Department of Genètica, Universitat Barcelona, Diagonal 643, Barcelona 08028, Spain, ‡Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Road, Santa Cruz, CA 95060, USA

Abstract

Single nucleotide polymorphisms (SNPs) are rapidly becoming the marker of choice in population genetics due to a variety of advantages relative to other markers, including higher genomic density, data quality, reproducibility and genotyping efficiency, as well as ease of portability between laboratories. Advances in sequencing technology and methodologies to reduce genomic representation have made the isolation of SNPs feasible for nonmodel organisms. RNA-seq is one such technique for the discovery of SNPs and development of markers for large-scale genotyping. Here, we report the development of 192 validated SNP markers for parentage analysis in *Tripterygion delaisi* (the black-faced blenny), a small rocky-shore fish from the Mediterranean Sea. RNA-seq data for 15 individual samples were used for SNP discovery by applying a series of selection criteria. Genotypes were then collected from 1599 individuals from the same population with the resulting loci. Differences in heterozygosity and allele frequencies were found between the two data sets. Heterozygosity was lower, on average, in the population sample, and the mean difference between the frequencies of particular alleles in the two data sets was 0.135 ± 0.100. We used bootstrap resampling of the sequence data to predict appropriate sample sizes for SNP discovery. As cDNA library production is time-consuming and expensive, we suggest that using seven individuals for RNA sequencing reduces the probability of discarding highly informative SNP loci, due to lack of observed polymorphism, whereas use of more than 12 samples does not considerably improve prediction of true allele frequencies.

Keywords: minor allele frequency, non-model species, RNA-seq, SNP development, *Tripterygion delaisi*

Received 7 May 2013; revision received 13 July 2013; accepted 18 July 2013

Introduction

The application of single nucleotide polymorphisms (SNPs) as genetic markers for the study of model species has been common practice for more than a decade. The discovery of SNP markers has been primarily based on two methods: Sanger sequencing unigene-derived amplicons or mining existing EST databases followed by PCR-based validation (Picoult-Newberg *et al.* 1999). However, these methods are time-consuming and expensive and therefore restricted to small-scale discovery (Vignal *et al.* 2002). Although reduced representation shotgun Sanger sequencing enabled the discovery of a huge number of SNPs in the human genome (Altshuler *et al.* 2000), 'next generation sequencing' technology now provides much greater genomic resources for the discovery of these genetic markers and has made large-scale detection of SNPs in nonmodel species much more accessible.

Aside from the relative facility in discovery, SNPs also have multiple other advantages, including the large number of them in the genome of most organisms (Brumfield *et al.* 2003). For example, tens of millions of SNPs have been identified in humans (The 1000 Genomes Project Consortium 2012). In population genetics, evaluation of genetic differentiation generally employs putatively neutral markers such as microsatellites, but if genes under selection are responsible for population differentiation, such a focus can fall short (Morin *et al.* 2004). Indeed, non-neutral SNP markers have recently been shown to reflect genetic differentiation in a much more refined manner than neutral ones (De Wit & Palumbi 2012; Lamichhaney *et al.* 2012) and also to disentangle the components of differentiation that are due to adaptation and to drift (Jones *et al.* 2012). SNPs also have lower genotyping error rates and observer biases in

Correspondence: C. Schunter, Fax: +34-93-40434420;
E-mail: celiaschunter@gmail.com

comparison with microsatellite markers (Everett *et al.* 2011; Helyar *et al.* 2011), and the ease of standardization allows better comparison across studies and combination between laboratories (Hauser *et al.* 2011). Finally, SNP genotyping is more rapid and cost-effective than other markers, and therefore facilitates large-scale studies with greater numbers of samples and/or markers. Despite these advantages, the development of SNP markers for nonmodel species without a reference genome has really only become viable in recent years.

RNA-seq is a transcriptome profiling approach that uses high-throughput sequencing technology (Wang *et al.* 2009) and can also allow for SNP discovery. A few recent population genomic studies have demonstrated the utility of RNA-seq for the development of SNPs in marine species, such as the Atlantic herring *Clupea harengus* (Lamichhaney *et al.* 2012) and the red abalone *Haliotis rufescens* (De Wit & Palumbi 2012). A benefit of SNP discovery with RNA-seq is that the data obtained by sequencing can also be used for additional research questions, such as the study of adaptation to different environmental conditions (Barshis *et al.* 2013).

Tripterygion delaisi (the black-faced blenny) is a small near-shore rocky reef fish from the Mediterranean Sea and northeastern Atlantic Ocean. It has been the subject of phylogeographic and population genetic studies using microsatellites (Carreras-Carbonell *et al.* 2005, 2006; Domingues *et al.* 2007), as well as of behavioural ecology studies (Wirtz 1978; Jonge & Videler 1989). The species is common and easy to catch and is of interest for the study of alternative reproductive tactics and behaviour.

Here, we describe the application of RNA-seq for the development of SNP markers for parentage analysis with *T. delaisi*. Despite being bi-allelic, SNP markers have been shown to have great utility to accurately and powerfully identify kinship, parentage and population of origin. We use simulations to evaluate the polymorphism and allele frequency estimates that result from different number of individuals being included in the RNA-seq survey and compare these estimates with the corresponding values in a large population sample. This analysis provides guidance on strategies for cost-effective application of RNA-seq for SNP discovery. Finally, general considerations regarding the discovery of SNPs for parentage analysis using RNA-seq data are provided.

Methods

Sample collection and NGS sequencing

Fifteen individuals of *Tripterygion delaisi* were collected during the reproductive season in June 2010 from a near-shore location on the Mediterranean coast of Spain (near Blanes; 41°67′N, 2°30′E). The collection consisted of five territorial males with bright yellow colouring and black heads, five camouflaged secondary (sneaker) males and five females. Fish heads were removed and snap-frozen in liquid nitrogen. Whole brains were then dissected out and placed in TRI reagent for RNA isolation. cDNA was fragmented, and four-mer barcodes for each individual were ligated to allow multiplexing. High-throughput cDNA sequencing was performed in four lanes of Illumina Hiseq 2000. One dominant male, one sneaker male and one female were used for sequencing with a target length of 109 bp single end (SE) reads, and four more individuals of each phenotype were used for sequencing with a target length of 52 bp SE, for a total of five biological replicates per phenotype. Approximately 12.5 million reads on average were obtained for each of the 15 individual samples. The *de novo* transcriptome assembly was produced with the 15 individual samples as well as eight additional pooled samples and was assembled with Trinity (Grabherr *et al.* 2011).

In silico SNP development

High-quality reads from the 15 individual samples were mapped against the *de novo* assembled transcriptome with the programme BWA (Li & Durbin 2009) and using a conservative approach, by setting the error rate to 0.005 and the distance seed to 5. The output files of mapped reads (one from each individual sample) were then sorted, and individual IDs for each sample (read groups) were added, with duplicate reads removed. All of this was carried out with Picard (http://picard.sourceforge.net/) command line tools. Subsequently, the 15 files were merged and indexed with SAMtools (Li *et al.* 2009). Due to the presence of insertions and deletions in the data, individual reads could be misaligned against the reference and mistaken for SNPs. Therefore, we performed local realignments of regions with elevated numbers of insertions or deletions, to minimize mismatching across all reads, with the Genome Analysis Toolkit (GATK 2.0; De Pristo *et al.* 2011). High-quality variants were called with Unified Genotyper using a minimum Phred quality score of 30. Variant sites of all qualities were then recalibrated against the high-quality variants with VariantRecalibrator, which uses a Gaussian mixture model to better distinguish true variant sites from sequencing error. Finally, out of this recalibrated data set, we selected all variant sites with a Phred quality score of 30. Genotypes for the 15 individuals used for each of the variant sites were extracted and combined using a published python script (De Wit & Palumbi 2012). Functional enrichment was assessed using exact tests (FDR 0.05) in Blast2Go (Conesa *et al.* 2005) for

contigs with more than two alleles, as well as for the final SNP panel.

SNP selection, assay design and genotyping

Putative SNPs were selected for assay design after a series of filtering steps. The first filter retained only sites with no secondary variants within 25–30 bp of the target SNP. This is due to the requirements of the assay chemistry employed. To better estimate genotype and allele frequencies, we chose to retain only putative SNP sites with mapped reads for at least seven individuals. For the retained set of target sites, deviations from Hardy–Weinberg (H-W) equilibrium were evaluated, and sites not consistent with H-W equilibrium removed. Allele frequencies were calculated and SNPs sorted by frequency. As SNPs are most powerful for parentage analysis when the minor allele frequency (MAF) is 0.5 (Anderson & Garza 2006), we selected putative SNPs with frequencies between 0.3 and 0.7 for the allele found in the reference transcriptome.

Because the putative SNPs were identified in transcriptomic data, which only includes sequences from processed mRNA, there is a risk of assay failure due to proximity of target sites to an intron/exon boundary. To predict intron/exon boundary locations, we blasted (blast-n, e-value cut-off of 1×10^{-10}) the transcriptome contigs with retained putative SNPs against the reference genomes of *Danio rerio* and *Oreochromis niloticus*, which are the most closely related species with a reference genome. When a contig successfully aligned to a segment of another species' genome sequence, we determined whether the 60 bases on each side of the putative SNP (needed for assay design) were also present and contiguous in the alignment. This method was aimed at determining whether the targeted SNP site is too close to an intron and thus at risk of assay failure. However, it should be noted that this will bias SNP discovery towards homologous genes between different species. Sites for which we confirmed the presence of the flanking sequence in another species' genome are hereafter termed 'nonintron' SNPs. All putative SNP sites were visually inspected with the Integrative Genome Viewer (IGV 2.1) prior to assay design. We then designed and evaluated a total of 240 SNPtype genotyping assays (Fluidigm Corporation). For the first 96 assays, we selected 49 putative 'nonintron' SNPs and the 47 putative SNPs from nonannotated contigs with the allele frequency closest to 0.5. The subsequent 144 assays all targeted putative 'nonintron' SNPs. All 240 assays were initially evaluated with a set of 94 *Tripterygion delaisi* individuals and those with insufficient signal strength or allele clustering quality, or with deviations from H-W equilibrium, were dropped. This left a total of 192 SNP assays that were used for subsequent genotyping of 1599 *T. delaisi* individuals, all collected at the location described above from March to September 2010. DNA was extracted with DNeasy 96 Blood and Tissue Kits (QIAGEN Inc.) on a QIAGEN Bio Robot 3000. Primer pools were produced for two sets of 96 SNP loci and a pre-amplification performed with the following thermocycling conditions: 95 °C for 15 min, then 14 cycles of 5 s at 95 °C and 4 min at 60 °C. Pre-amplified DNA was then diluted 1:100 in suspension buffer and 2.5 μL combined with Biotium 2X Fast Probe Master Mix, SNPtype 20× Sample Loading Reagent (Fluidigm), and the reference dye ROX (Invitrogen Inc.). In parallel, each SNP type assay was mixed with 2X Assay Loading Reagent (Fluidigm). These were then loaded onto Fluidigm 96.96 Dynamic Genotyping Arrays, which use nanofluidic circuitry to combine 96 loci with 96 samples in 9216 reaction chambers, with thermocycling and fluorescence detection performed on the EP1 genotyping system (Fluidigm).

Data analysis

Genotypes for all 1599 individuals of *Tripterygion delaisi* were called using the Fluidigm SNP Genotyping software and allele frequencies calculated to obtain the 'true' minor allele frequency in the population for all 192 SNP loci. Genotype frequencies from 1599 genotyped individuals were tested for H-W and linkage equilibria with Genepop 4.2 (Rousset 2008). All SNP assays were also tested on 230 samples of *Tripterygion tripteronotus*, a very closely related and sympatric species, to evaluate species specificity.

As library preparation for high-throughput sequencing can be expensive and time-consuming, we evaluated how well we would have estimated the population allele frequencies of our SNP loci with different numbers of individuals used in the RNA-seq data generation. We used sequence reads for 15 individual samples for the reference allele frequency estimates, and resampled (with replacement) the consensus genotypes for these 15 individuals to produce bootstrap samples with different numbers of individuals. We generated 15 different data sets with 100 replicates each for 1–15 individuals, which provided bootstrap estimates of the reference allele frequency for each locus. The probability of observing no sequence variation at each locus with different numbers of individuals sequenced was then estimated. We also calculated the absolute differences between the bootstrap estimates of the allele frequencies for each data set and the 'true' allele frequencies in the population obtained from genotyping 1599 individuals. We present the mean variance of the difference between the estimated allele frequency and the 'true' frequency.

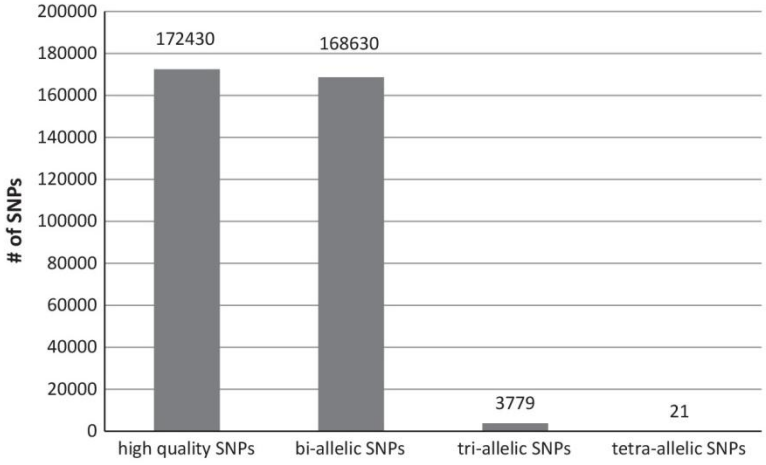

Fig. 1 Number of variant sites with different numbers of nucleotides observed in RNA-seq data.

Results

SNP discovery

A total of 172 430 high-quality polymorphic sites were found in the reference transcriptome, which translates to a high-quality SNP every 894 bp. Of these putative SNPs, 97.80% had one alternative allele, 2.19% had two alternative alleles and only 0.01% was tetra-allelic (Fig. 1). No bias in representation of any functional annotation was found for the contigs with three or four allelic variants.

For the bi-allelic SNPs, the majority of variant sites detected in the reference assembly were transitions (Fig. 2). Nevertheless, transversions represented approximately 40% of polymorphic sites observed, resulting in a transition:transversion ratio of 1.46:1. The frequencies of the four different possible transversions were significantly different ($\chi^2 = 110$, $P < 0.001$), with AT and CG variants less frequent.

To select sites for SNP assay design, we first excluded putative SNP sites with secondary variants within 30 bp, and 45 961 (26.65%) were excluded from further analysis due to the presence of such variants (Fig. 3). To better evaluate the allele frequency and decrease the probability of designing assays for sites that are not true SNPs, we further retained only those putative variants found in seven or more individuals, which reduced the total to 30 909. Of these sites, 6 919 (or 4.01% of all high-quality putative SNPs) had a reference allele frequency between 0.3 and 0.7 and did not deviate from H-W equilibrium. Almost all (99%) of the sites that did deviate from H-W equilibrium were found to have an excess of heterozygotes. While the selection of SNPs with intermediate allele frequency could skew this result, we found the same pattern in all putative bi-allelic SNPs, with an excess of heterozygotes in 90% of those not at H-W equilibrium. The 6 919 contigs containing the putative SNPs that met the allele frequency criterion and were in H-W equilibrium were then compared with reference genomes as described above. The number of 'nonintron' putative SNPs found when comparing against *Oreochromis niloticus* was 2 326 and when comparing against both *Oreochromis niloticus* and *Danio rerio* was 621. This final number of putative SNPs for assay design was just 0.36% of the total high-quality polymorphic sites initially identified.

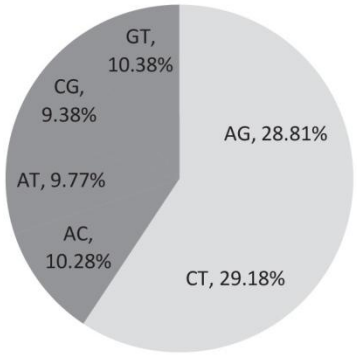

Fig. 2 Percentage of type of nucleotide substitutions. Transversions are indicated in dark grey and transitions in light grey.

SNP assay evaluation

The first panel of 96 assays targeted 49 'nonintron' SNPs and 47 randomly selected variant sites (all filtering criteria applied, except comparison with other genome sequences). The success rate of the 'nonintron' SNP genotyping assays was 80%, whereas only 68% of the

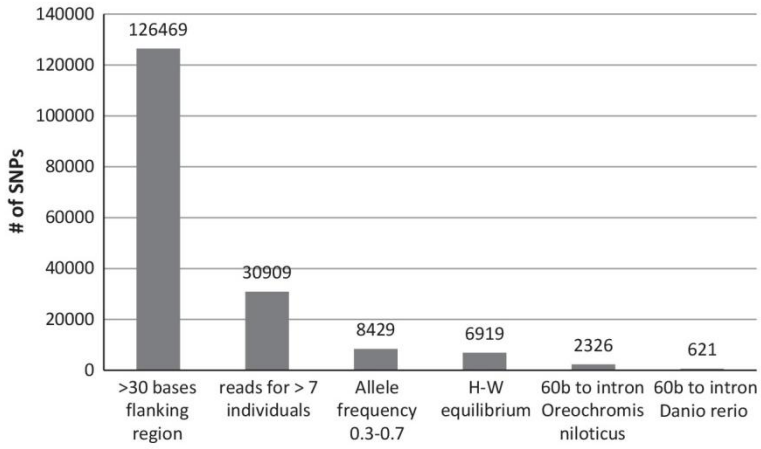

Fig. 3 Reduction in the number of putative SNPs by adding filters (from left to right). The *Oreochromis nilotus* and *Danio rerio* columns indicate the number of contigs successfully blasted against the reference genome with at least 60 bp to an intron on both sides of the candidate SNP site.

assays for the randomly selected loci yielded consistently scorable genotypes. Only four of the 96 assays had no fluorescent signal, all of which were from the randomly selected loci and most likely due to an intron/exon boundary close to the targeted polymorphic site. Additionally, four of the putative variant sites were monomorphic in the population samples, with two each from the 'nonintron' and the randomly selected loci.

After this initial evaluation, 144 more SNP assays were designed, all targeting 'nonintron' SNPs. None of these 144 assays failed completely (i.e. no fluorescent signal) and 84% successfully and consistently produced scorable genotypes. The final set of 192 SNP markers for *Tripterygion delaisi* can be found in Table S1 (Supporting information). Thirty-one of these loci (Tdel001 to Tdel059) are from the 'randomly' chosen sites, and the other 161 loci (Tdel101 to 334) are located in genomic regions with sufficient distance to introns in other fish species. Heterozygosity did not differ significantly between the two sets of loci ($Z = 1.388$, $P = 0.165$). Functional analysis of the annotated contigs used to develop the SNP markers did not reveal over- or under-representation of any functional gene ontology category when compared to the overall categories of the reference transcriptome.

Genetic variability

There were significant differences between the heterozygosity estimates for the 192 SNP loci from the sequence data of 15 individuals and from the genotype data of the 1599 individuals (Wilcoxon-matched pairs test; $Z = 10.14$; $P > 0.0001$; Fig. 4). The estimates from the RNA-seq data ranged from 0.27 to 0.75 and from the genotype data between 0.19 and 0.76 (Table S1, Supporting information). We identified eight outlier loci, as well as two extreme outliers with low heterozygosity and one

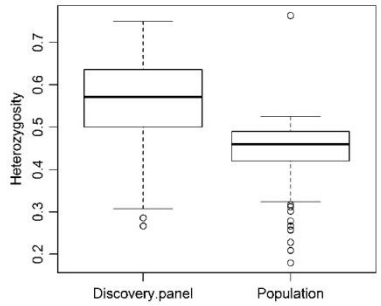

Fig. 4 Heterozygosity estimates from the RNA-seq SNP discovery data from 15 individuals and the 'true' heterozygosity from genotypes of 1599 individuals.

extreme outlier with high heterozygosity (Fig. 4). The latter (Tdel049) was found to be in the gene MRP-L51, which codes for the mitochondrial ribosomal protein.

Although all loci were consistent with H-W equilibrium in the sequence data, several loci were found to not be at equilibrium in the genotyped population (Table S1, Supporting information). There were six pairs of SNPs that had significant linkage disequilibrium: Tdel040 and Tdel267, Tdel273 and Tdel194, Tdel253 and Tdel216, Tdel264 and Tdel236, Tdel148 and Tdel209 and Tdel059 and Tdel250.

There were no significant differences between the allele frequency estimates from the RNA-seq data and from the genotype data ($Z = 1.052$, $P = 0.293$). However, the mean difference in the frequency of the same allele as estimated from the sequence data and from the genotypes was 0.135 ± 0.100. This is due to the fact that for some loci, the minor allele in the RNA-seq data was the more frequent allele in the genotypes and vice versa.

The 192 SNP assays were also evaluated with 230 samples of *T. tripteronotus* and 11 of them failed. The

other 181 assays successfully produced genotypes, but 94% of these loci were monomorphic. These SNP assays are therefore of limited utility for genetic analysis in nonfocal species, but the extremely limited variation does allow identification of sympatric *T. tripteronotus* individuals.

How many individuals used for RNA-seq are optimal for SNP development?

Library production for next generation sequencing can be expensive and time-consuming, so it would be helpful to be able to predict the number of individuals that must be sequenced for the discovery of a set of highly informative SNP markers. By resampling, we can evaluate whether <15 individuals would have provided similar results. The number of loci identified as monomorphic was higher when sequence data from very few individuals were used (Fig. 5). With data from one individual, 60 of the 192 loci would have been identified as monomorphic on average, whereas this value, and its standard deviation, declines steeply with increasing individuals. With data from seven individuals, the mean and the standard deviation remain below the 5% margin of error, with a maximum of six loci inferred as monomorphic, and with 11 individuals variation would have been observed at nearly all of the loci.

The variance of the difference in allele frequency between the sequence data and the population genotypes is highest when sequence from only one individual is used (Fig. 6), decreases substantially by adding data from a second individual, and continues to decline steadily up to 12 individuals. While this indicates that it may be possible to reduce the number of individuals for, and therefore the cost of, sequencing, it is still important that the individuals from which the SNPs are ascertained be as representative as possible of the whole population and species.

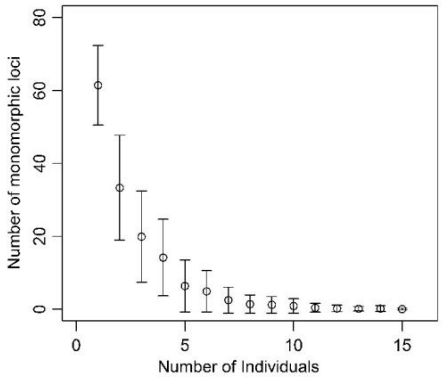

Fig. 5 The proportion of loci (of 192) which were inferred to be monomorphic depending on how many individuals are used for the RNA-seq SNP discovery.

Discussion

We report the discovery of 192 validated single nucleotide polymorphism (SNP) markers from RNA-seq data of a near-shore rocky reef fish, *Tripterygion delaisi* (the black-faced blenny). These SNP markers were selected from 172 430 putative high-quality SNPs identified from the reference *de novo* transcriptome assembly. SNP discovery in other nonmodel species using RNA-seq data has generally found fewer variable sites (6094 putative SNPs for lake whitefish, Renaut *et al*. 2010; 26 000 for sockeye salmon, Everett *et al*. 2011; and 21 579 for the red abalone, De Wit & Palumbi 2012), whereas efforts using exome sequencing (440 817 SNPs for the Atlantic herring, Lamichhaney *et al*. 2012) or whole genome sequencing (2.7 millon SNPs for the giant panda, Li *et al*. 2010) have found more. The transition:transversion ratio

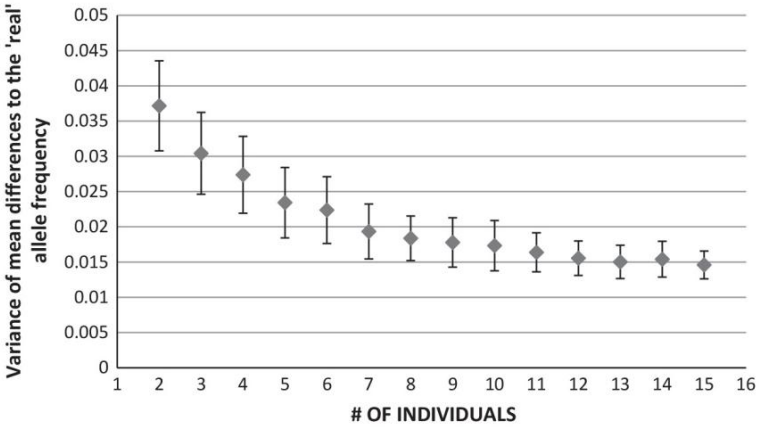

Fig. 6 Comparison between the mean allele frequencies estimated with different numbers of individuals sequenced and from the population genotypes. Mean variance across loci is presented with standard deviation across bootstraps.

found for *T. delaisi* was similar to the ratio found by pyrosequencing in lake whitefish (*Coregonus* spp. 1.65:1, Renaut *et al.* 2010).

SNP discovery via RNA-seq

SNP discovery from high-throughput sequencing data presents several important challenges, the most prominent of which are sequencing errors and discriminating them from true segregating polymorphisms (Ekblom & Galindo 2011). Olsen *et al.* (2011) sequenced a random genomic library of three ringed seal individuals for SNP discovery, and putative SNP regions were then resequenced. Over 54% of their 768 putative SNPs turned out to be false positives due to sequencing error. This can be partially avoided by augmenting sequencing coverage and increasing sequence quality thresholds for data retention. However, the number of individuals used for sequencing is critical, as the probability of a sequencing error occurring at the same site in sequence reads from different individuals is very small. Many SNP calling programmes, including GATK, use a Gaussian mixture model to estimate a probability of being a true SNP for each variant site (De Pristo *et al.* 2011). Another problem for SNP discovery with high-throughput sequencing methods, including RNA-seq and restriction-site-associated DNA sequencing (RAD-seq), is that variants in multicopy genomic regions can be mistaken for SNPs (Fredman *et al.* 2004). Allelic dropout is also an issue with RAD-seq (Gautier *et al.* 2012), but not RNA-seq, as amplification is not based on genomic restriction enzyme sites. However, the use of RNA-seq data to discover SNPs has its own set of challenges when a *de novo* reference transcriptome must be assembled, which is not trivial (Martin & Wang 2011).

Variant site preselection

Filtering putative SNP loci from RNA-seq data resulted in a large decrease in the number of candidate sites for design of genotyping assays, especially when putative SNP sites had to have reads from at least seven individuals. We further selected polymorphic sites that had an allele frequency between 0.3 and 0.7 and were in H-W equilibrium, as SNP loci are most informative for parentage analysis when alleles are equifrequent (Thompson 1987; Anderson & Garza 2006; Baruch & Weller 2008). Nearly all of the loci not in H-W equilibrium (99%) had an excess of heterozygotes, which likely means that duplicated genes or domains were sequenced (Hohenlohe *et al.* 2011), although some of these sites could be under balancing selection. To avoid designing genotyping assays in regions that cross an intron/exon boundary (IEB), candidate SNP loci can first be PCR amplified and evaluated by sequencing or electrophoresis, as has been carried outwith sockeye salmon (Storer *et al.* 2012). A more cost-effective approach, implemented here, is to choose putative SNP loci that successfully align to a genomic region in other species and for which no introns are present. There were no assay failures for loci selected with this procedure, whereas 8.5% of assays for 'randomly' chosen loci failed entirely. By selecting genes with blast hits against reference genomes of other fish species, we are not considering SNPs in novel, orphan or highly divergent genes. Nonetheless, genetic variability in the assayed population did not differ significantly between the SNPs selected with the two methods. Therefore, using data from reference genomes of related species to identify homologous regions and intron/exon boundaries can help reduce the failure of assays developed from RNA-seq data.

How well does RNA-seq data predict SNP variability in population genotyping?

We had sequence data from 15 individuals for SNP discovery and 1599 individuals were subsequently genotyped with the 192 SNPs markers that were successfully developed and validated. We evaluated how well 15 individuals represented the population in terms of marker variability and allele frequency. The mean allele frequency across all loci estimated from the sequence data and from the population genotypes were very similar, even though the average difference in the frequency of a particular allele was 0.135. This is in agreement with Milano *et al.* (2011), who did not find a significant difference between the MAF from high-throughput sequence data of a marine fish and that of the genotyped population. This indicates that 15 individuals is a sufficient number for accurate prediction of the usefulness and power of SNP loci, although not necessarily the exact population allele frequencies. Such prediction is particularly beneficial when developing SNPs for parentage analysis, where a high MAF is critical (Anderson & Garza 2006). Conversely, the sequence data did not predict observed heterozygosity of the resulting SNPs in the population, as was also observed by Milano *et al.* (2011). However, in their study, mean heterozygosity was higher for the population than in the sequence data used for discovery, which is contrary to our findings. Additionally, all of our SNPs that were successfully developed into markers were consistent with H-W equilibrium in the sequence data, but eight loci (4.17%) were not in the population sample, indicating that 15 individuals are not sufficient to fully predict whether a SNP locus is in H-W equilibrium.

Optimal data set for RNA-seq derived SNP discovery

The production of cDNA libraries for RNA-seq is time-consuming and expensive, and it would be optimal to reduce the number of libraries/individual samples to a minimum. The probability that a highly polymorphic site would be observed to be monomorphic in the RNA-seq data used for SNP discovery, and therefore removing from further consideration, is highly reduced when seven or more individuals are used. This would be especially important if a large set of SNP markers is needed, as the number of candidate SNP sites would be greatly reduced due to failure to observe polymorphism. If an elevated MAF is necessary, as for parentage analysis studies, the SNP discovery data should predict the true population allele frequency as well as possible. We found that about 12 individuals included in the RNA-seq data collection predict the mean population allele frequency as well as 15 individuals, although the true frequencies of individual alleles in the population may be quite different.

Acknowledgements

We thank S. Vollmer for the help with the *de novo* transcriptome assembly. We are grateful to the Southwest Fisheries Science Center Molecular Ecology Team, especially V. Apkenas, C. Columbus, E. Gilbert-Horvath, D. Barshis, E. Anderson, D. Pearse and A. Abadía-Cardoso for support and helpful comments. We also thank C. Pegueroles for bioinformatic expertise. This work was partially funded by the Spanish Ministry of Science and Innovation through the BENTHOMICS (CTM2010-22218-C02) project and the FBBVA project (BIOCON 08 – 187/09). The authors CS, EM and MP are part of the research group 2009SGR-636, 2009SGR-665 of the Generalitat de Catalunya.

References

Altshuler D, Pollara VJ, Cowles CR *et al.* (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513–516.

Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, **172**, 2567–2582.

Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR (2013) Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 1387–1392.

Baruch E, Weller JI (2008) Estimation of the number of SNP genetic markers required for parentage verification. *Animal Genetics*, **39**, 474–479.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, **18**, 249–256.

Carreras-Carbonell J, Macpherson E, Pascual M (2005) Rapid radiation and cryptic speciation in mediterranean triplefin blennies (Pisces: Tripterygiidae) combining multiple genes. *Molecular Phylogenetics and Evolution*, **37**, 751–761.

Carreras-Carbonell J, Macpherson E, Pascual M (2006) Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. *Molecular Ecology*, **15**, 3527–3539.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)*, **21**, 3674–3676.

De Pristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

De Wit P, Palumbi SR (2012) Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, **21**, 2884–2897.

De Wit P, Pespeni MH, Ladner JT *et al.* (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, **12**, 1058–1067.

Domingues VS, Almada VC, Santos RS, Brito A, Bernardi G (2007) Phylogeography and evolution of the triplefin *Tripterygion delaisi* (Pisces, Blennioidei). *Marine Biology*, **150**, 509–519.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

Everett MV, Grau ED, Seeb JE (2011) Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. *Molecular Ecology Resources*, **11**(Suppl. 1), 93–108.

Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ (2004) Complex SNP-related sequence variation in segmental genome duplications. *Nature Genetics*, **36**, 861–866.

Gautier M, Gharbi K, Cezard T *et al.* (2012) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, **21**, 3165–3178.

Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources*, **11**(Suppl. 1), 150–161.

Helyar SJ, Hemmer-Hansen J, Bekkevold D *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, **11**(Suppl. 1), 123–136.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**(Suppl. 1), 117–122.

Jones FC, Chan YF, Schmutz J *et al.* (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology*, **22**, 83–90.

Jonge J, Videler JJ (1989) Differences between the reproductive biologies of *Tripterygion tripteronotus* and *T. delaisi* (Pisces, Perciformes, Tripterygiidae): the adaptive significance of an alternative mating strategy and a red instead of a yellow nuptial colour. *Marine Biology*, **100**, 431–437.

Lamichhaney S, Martinez Barrio A, Rafati N *et al.* (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 19345–19350.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **25**, 1754–1760.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, **25**, 2078–2079.

Li R, Fan W, Tian G *et al.* (2010) The sequence and de novo assembly of the giant panda genome. *Nature*, **463**, 311–317.

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nature Reviews Genetics*, **12**, 671–682.

Milano I, Babbucci M, Panitz F *et al.* (2011) Novel tools for conservation genomics: comparing two high-throughput approaches for SNP discovery in the transcriptome of the European hake. *PLoS ONE*, **6**, e28008.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**, 208–216.

Olsen MT, Volny VH, Bérubé M *et al.* (2011) A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Arctic ringed seal (*Pusa hispida hispida*). *Molecular Ecology Resources*, **11**(Suppl. 1), 9–19.

Picoult-Newberg L, Ideker TE, Pohl MG *et al.* (1999) Mining SNPs from EST databases. *Genome Research*, **9**, 167–174.

Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology*, **19**(Suppl. 1), 115–131.

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

Storer CG, Pascal CE, Roberts SB, Templin WD, Seeb LW, Seeb JE (2012) Rank and order: evaluating the performance of SNPs for individual assignment in a non-model organism. *PLoS ONE*, **7**, e49018.

The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1092 human genomes. *Nature*, **491**, 56–65.

Thompson EA (1987) Parental and sib likelihoods in genealogy reconstruction. *Biometrics*, **43**, 585–600.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution*, **34**, 275–305.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57–63.

Wirtz P (1978) The Behaviour of the Mediterranean *Tripterygion* Species (Pisces, Blennioidei). *Zeitschrift fur Tierpsychologie*, **48**, 142–174.

C.S., E.M. and M.P. conceived the idea and designed the research. C.S. and E.M. conducted the field work. C.S. produced the RNAseq data. Bioinformatic analysis, SNP discovery and SNP genotyping was done by C.S. with the help of J.C.G. C.S. and M.P. analyzed the results and C.S., E.M., J.C.G. and M.P. wrote the manuscript.

Data accessibility

Data have been deposited at the National Center for Biotechnology Information (NCBI) Transcriptome Shotgun Assembly (TSA) database under BioProjectPRJNA186408. Raw sequence reads can be found in the NCBI Sequence Read Archive database and NCBI dbSNP accession numbers for the final set of SNP markers are specified in the Table S1 (Supporting information).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 List and data base SNP (dbSNP) numbers of 192 SNP loci sequences from *Tripterygion delaisi* with at least 60 base on each side of the variable site.

Supplementary Material Table S1: List and data base SNP (dbSNP) numbers of 192 SNP loci sequences from *T. delaisi* with at least 60 bases on each side of the variable site. The two SNP alleles are marked in brackets. Allele frequencies and heterozygosities estimated from the 1,599 genotyped individuals. *Not in H-W equilibrium in the genotype data.

Assay Name	dbSNP No.	Assay sequence [reference transcriptome allele/alternative allele]	Allele frequency of the reference allele	Heterozygosity
Tdel001	778235193	ATGCCTCTCKACTTGATCTTCCCTTCTTCTCCAAACCTCTCTGCTCAACAGGTTTGCGAGC[C/T]TCCCCTCTCCACTCCTCTCTTTGCCCGCCTGGTGTTGAACTGGCTTGTTTCCCCATAGTAA	0.459	0.515
Tdel005	778235212	CATCCAACGTCCAGATGTATTTCATCTCACAGTTCACCCTCAATAAATATGCCTTTGTTGC[T/C]TTTTCCTGGTTTCTTCTCCACCTTCAAGCTTTCTTTTYTTCGTCTAAGATCCCATCTCCAA	0.443	0.463
Tdel008	778235218	GTAGACCCTGCCCCCCCCCTCTCTCCACGCAAACCTCTGGAATCACACACACACATGCACA[C/A]ACACTTCCCTCTGTACCCATAGATATAACTAATATCTATGTTGATGCCTCCTTTATGACTT	0.420	0.468
Tdel009	778235219	AAAAAGAAGCTTGTTTTTCTTCCAAGGAAATGTTGCCACTCAGAACGTAAAAAGTGGTATT[G/T]TTTCCTGACAGCAGATGTTAGTGTTGATCACMTCTGACATTTCACTTTTGATATATCCAAT	0.596	0.453
Tdel010	778235223	GAATGTTATCTGCTCCCATTTGACTGTGTGGAAYATCTATGGACTGTTGGCCTCATCGTTC[T/C]TGAAAATATCTGCAGTTGACTGGAACATGGGAAGTTTTTGGTTGTGACTGTGCGATGCTGA	0.498	0.508
Tdel012	778235227	CTCTTCCTCTTCCTCACTCTGTTTCASAAAGACAAGAAGAATGGGAATAATCAGTCAAGAG[T/C]GGGAAACTGGTGGCAGACCTGAGCTTTATCAGTTAAAAGAAAGAAAAAAAAGAAAAAAAAA	0.468	0.494
Tdel014	778235230	ATGGACTGTGCAGATCGATCAGGCGCTTRTGGATCCTCATCTGGAAGCGATCCCAGGTTTT[G/A]GAACCTTCACCACAGGGGGTCTTTCTGGTGGTAATACGCAGAGTCTTGGTTGGCATGCGGA	0.691	0.418
Tdel015	778235232	GCGAGAGACCACCGGTGCTCATCCAGCTCTACCTCAGCTGGATCAGAAACGCTGATGCCGC[T/A]TCATATTTATACATATCTACAGTGCAAGTCGTCTCCAYTGCTATCAAGTCGCACTGCCTCA	0.289	0.399
Tdel017	778235235	TCTTCAGACTTTTCTCCTCCTCCTCTCTTTTCTCTCCTTCCTGTCTGCGGACGGCGTGCTG[G/T]TTTCTTAGCGGAGGGTGGGCGTGGCCGTGTTGGACGGAAGGATTCAGTCGCCAACCGA ATC	0.488	0.507
Tdel018	778235240	CATATGACTAAAACCCCAAATTCCAACTTCATATTTGCCATGAATGTAACATTATTCCCTT[C/T]AACGTGAACCTGCTCGTCTCCCCTAACGACCGGTCAGCATCAGCACGTGTCCGTCCACCTC	0.466	0.497
Tdel021	778235242	CACGTATGACTGTACTRCACAGACTTGGGATGAGCTGAGGTGGAGTGGAGGAATTAGCATG[A/T]GCAGAGCGTTATTATAGCTCATCAGGGTATTTCGCCAGTTTTCTTCACCTTTAAATCTCAG	0.558	0.490
Tdel022	778235243	TAGCGGAGGGGCCGTTGTGAGCGCCGTTGTTTTGGTCCTGGGCTTCAGGAGCCTCCTCCAC[G/T]GTTGGCTTCGGCCTGTTCATGACTGGATGACAYACATCCTCCAACTCCTGAATCTTGGCGA	0.507	0.476
Tdel024	778235244	TGTTATAGGCGTGGATGTATCCCCTTTGGAGATTGAAAGGAACACCTAAGCCTGCTCCGTT[C/T]TCGGCATCATCGCTGGTGGTCGAAGAAGTTGTGCTCGTGGTRGATGCGCACTGTACAAGCA	0.738	0.398
Tdel028	778235247	CGATGTACAGCGGGCCGGTGCTGCGGTTYCAGTACTGTATTTCCTGAGGGTACAGTAAAGC[G/A]TTCCAGGAATACTCTCAGGCCATTGGCCAGGTGTACCCGGAYATCCGCATYGAGGGAGAAA	0.497	0.503
Tdel032	778235251	TGGTACTCTCACTCTCAGAAGGTTTAGATGGAGTTGCATCACGATTTGTGTCCTTACAGGG[G/A]TCTTGTTTGGTTGTCACGATGACTTTGAAAACTGGGACTTGGTCRGATGTTGTTTCMTCTT	0.273	0.376
Tdel033	778235254	ATCAGAGCAGCCTCACTGCTGTGAAGTATAGATTGGAATAATTGTTGCTGCTAGCTCTCAT[A/T]TCTCGTTATCTAAAGTTTTGGTGGATTGTTCAAACCACAACACGCCTCTTGAGGTAGACAC	0.727	0.410
Tdel037	778235263	CTGCAGGGTCCAAAATAAMATGCTTTGGCTTATTTTGTTCACCAGGGGGGTTAAAAAGTCA[C/A]AAACCAGCATGGTTTGATTTTGCATYGTGTTTTGTATCGAAAAATGGGACTCAATGAGGTA	0.600	0.433
Tdel039	778235264	AGAGGCTTCCTGGCTGGAGCAGCGGAGGTGGTGGTGGTGCGGCTGGCCGTTGTTGTGCGCA[T/C]TGACGTGGTGGTTTTACTGGCTCCTGCGGTGGTAGTCGTTCCAGTTTTGGTGGCAGCGGTG	0.301	0.407
Tdel040	778235267	SCCACTTCTGTGCTTACTGYGAGTGACTTAACATTGAATGTTAATGACCACAAATCTTGAG[C/T]TTTTTCCTTTAATTTGTTGCCCATCTGTGTGTAAATGCACAGCTGGAGGCTGAATAGTTAT	0.689	0.434

Tdel042*	7782352 73	GCATCATAAMAAGTAGTGAACCTTCGATGGTGACCTGCTGGATTCCCATTAACA AGAGTGC[A/G]TTTTGTCAAAGGAACGTACGGACAGGGCAGCGGGAGTCTTCT CAGCAGTAATGACTWGATG	0.575	0.482
Tdel043	7782352 75	GCATCATAAMAAGTAGTGAACCTTCGATGGTGACCTGCTGGATTCCCATTAACA AGAGTGC[A/G]TTTTGTCAAAGGAACGTACGGACAGGGCAGCGGGAGTCTTCT CAGCAGTAATGACTWGATG	0.574	0.317
Tdel044	7782352 80	CGTCTTGAACACACGATAAGTGTGATCGCTCCTTTTCTCCTCTATCTTCTTCTC AAAGAAG[T/C]TGTCGTAGTGGAAGTGGGGCATTTTTYCTGACAAGTTGTCCT TCAGCAAGTGGGTGACCCT	0.369	0.477
Tdel048*	7782352 82	TWAAAAGTTATCTTTTAGTTTGCTGTCTTTTGAAGACAGGGAACATTGGCATAA GCCAATA[T/C]GAAAAGAAGCATCAATGTTGGGTTTATTGAAGGCAAATATAC AAAAATATAGATATTTCAT	0.271	0.339
Tdel049*	7782352 85	AATCCTCAGACTGGTTTGGAGTTGGTTCAGTTCAGTGGGTGGTTGGTACAGTCT GGTTGGT[G/T]CTGGTTATCGGTAGTTCTGGTTATCGGTGTTTCCCGACGCGG TTGAAGCGTCGGTAGAGGA	0.541	0.764
Tdel051	7782352 86	GGGCCAGCGGAGTGGTCGCCTCCACCCTGACGTTCTCACTGCGTCCATCCTCCT CCTCTTC[T/A]GCTCCCAACGGCCACAGTGTTGTTGTTTTCTTTTTTTAAATA TGTGGGGTAGGGGGGGGTT	0.434	0.483
Tdel052	7782352 87	TCTCTGGGCCCTTGGTGTCCAGGGCGATACCGATAGGCCTGTACTGGATGCTGC CGGGCTC[A/G]AAACTCTCACAGGCCTCACGCACGTTCTTAATTGTTTGTTCA TGGTACTCGTGCGTGCCGT	0.476	0.498
Tdel053	7782352 90	AGAAGAAAAAACGTCAGACCCAGAAGAGCGGCCACAAGCAGAATGCTCCCGTCC ATCATCA[T/C]CACCACCACCAGCAGTACCACCACCCCTCGGCCTCGCACCCT CATCACCAATCCCAGAGCT	0.560	0.486
Tdel054	7782352 93	ATCATGCATCGTTAGTTTGCAGGCTAAATGCGAAGCTGTTCTGTGATGTGAAAT GGCGTTG[A/G]TGCAAGCGTTTAACATGAAGTTATATATGTCTGTAGCCTACG TCTCGATGAGGAATAAGAC	0.476	0.479
Tdel055	7782352 97	CAGCTTTGGCTTGGCACGGCCCACCTGGGTGGACGGTCTGAACAATGCTGGACG ACAACTG[C/G]TGGACAACAGGGGGCGCTCTCTTCTCTGGACATCTGCGTGGG GCTTTAACAGTGCCAATGC	0.479	0.509
Tdel057	7782352 99	CCTGAACGACTCCTTCACCTCCTCCCACTTAACTCTTTACGTCTGAAAGAAAAC GTGGACC[G/A]GTGTGATTCTGTACAAGGTATTTACAGTATTTACAGCCRCTT GTYACACRGCGTGGACTCTCTAGTGGGACTT	0.516	0.496
Tdel059	7782353 02	TTGTAGCACKACTYTTCCCTGGRCTGCTTGGTGCATCAAGATCACTCAGAGCTC CGAGATC[A/G]TGAAGTTCACTTGATCGCAGACGTTTGATCGAAGACTTCAAC TCTYCCTCTKCCAGCTTKG	0.359	0.463
Tdel101	7782353 05	GGTGGTGGTCGTCTTCAGGCTAGGGAATGACAGACGGTGGCTGTCTTTTGTGTT AACTAC[A/C]GACACAGATGTGGACAGCTTAACCTGACCACCATGGAAACTGG AGGGCCTTATTGGGGAT	0.441	0.501
Tdel102	7782353 09	AGCCTGTYGTCTTGCCGCCACCAAACTGAGTCCTGAAGCCAAAGACGAACACAA CGTCGGG[G/T]GTGGTCTTGTACATCTTGGCAAGTTTCTCCCTGATTTCAGTC TTGGGGACTGTGGCTTTGC	0.449	0.478
Tdel103	7782353 11	GTGATGTAGAGTCCGATCTTGCCTCATACACAATGACAACATCCACAGGGGGCA CCTGTAC[G/A]GTTGTGTCCAGATCGAAGAAGCTTCAGGATGATGGTAGCTAT GGGGTGAAGAAAAATGCTT	0.506	0.525
Tdel105	7782353 14	GAACTGTGCCGAGTCCTTCACAGAGTACTGGACCTGCCTGGATTATACCGGCCT GGCAGAG[C/A]TGCGTCACTGCCGTAAACAGCAGAATGCCTTYGACAGCTGCA TACTTGACAAACTGGGYTG	0.496	0.472
Tdel106	7782353 16	CCATGGGCGTGGGCTTGTCCTCCTCCTTGGTCCCTGAGCTGGGGTTAGCTGTGC TGTTAGC[A/G]TTAGCATTGCTCTGCTCAGCGTCTTTGTTTGCGGACGCCTCC TGCTGCTGCCGCTGTTGCT	0.470	0.493
Tdel108	7782353 19	AGGCCGGCTGGATCGAAGGTGCTGCCATCCTGTTCTCAGTCATCATTGTGGTCC TTGTGAC[A/T]GCTTTTAATGACTGGTCCAAGGAGAAGCAGTTCAGAGGGCTG CAGAGCCGCATCGAGCAGG	0.431	0.507
Tdel111	7782353 21	TTGCCGTCTTCACCTTTGTTGATGATGCCAAGTACTTTGTAAGAAGGAATCCAT GGACATA[T/C]TACGTCTCCTATGCCGTCTTCTTTGTGTGTCTGATCGTCCTG AGCTGCTGTGGAGACTTCCGCC	0.324	0.424
Tdel112	7782353 25	CCACCTTCTYCCGGATCTTGCTAGCCCTCAGCGGGATCAGGTTCTTGCGTTTCC CGCTGAC[G/T]AGAGCGTTCTGCTTGTACTTGCCCTCCTCTTTGGTGCCGTCG GGGAAGGTCGTGCAGCCGT	0.653	0.418
Tdel113	7782353 30	ACAACCCAACCTACTTCTCCCGCCGACTCTCWCGCTTCGCAGACATCTACATGG CGTCGAT[C/A]AGCTGCATGCTCAACTACGACTTCCAGCACACCTTCTTCCCT CGMCGCACCCCCCTGCAGC	0.367	0.447

Tdel117	7782353 41	CTCRGATGAAGATGCTGAGRATCTGCCAGTAGACAAATCATCTGCTTCTGGAGC TGGTGGG[A/G]GCCACCATCAGTCATCGCCACGGACTTCTCAGAAGACCCAYG ATCCCCCTCCCAMCCCCAT	0.274	0.405
Tdel122	7782353 45	AGCGCGCTTTGAGGGATGACAGCCGGTTAAGGAGGGGAACTGAACATCACTTCA CACGCCA[G/A]ATGTCAAGTGTTGGGTTTCCTTCATCCACTGTCACTTCTCAG CCAGTGTCGTCTGCTGCCC	0.614	0.475
Tdel123	7782353 48	TCTTTTCTGCTTTGTCCTTGTCGTCCTGGGCGGAGCTCTTTGGCTTGTTTTCGT CGTCACT[C/T]AGATATTCGCTGTCGCTCGAGTCGGACTTTTCAGAGTCCTCT GAGTCACTGTGATCCACGT	0.539	0.486
Tdel124	7782353 51	ATCTGACGGACATTTTCACCACGCTGGTGGACCTCAAGTGGAGGTTTAACCTCT TCATCTT[C/T]GTGCTGGTGTACACGGTGACATGGCTCTTCTTCGGCTTCATG TGGTGGCTTATTGCATACC	0.810	0.279
Tdel126	7782353 55	TCTTTAAGGTGGACTACTCTCAGTTTCATGCCACATTTGAGGTTCCCACCCCTC CTTACAG[T/C]GTAAAAGAGCAGGAGGAGGCCCTGCTCCTCTCCTCGCCCCTC ATGGCTCCGTCCCTGTGCA	0.517	0.465
Tdel127	7782353 59	GACCTTCAAGACGATGCTGGGCCTTCACATCTGTGCCGTTCACCTTCCATGACG CCGACGC[A/G]TCATTATTCAGGGAGCCGCACTGCATGGTGACATCRCTGCCC ATCCGCTCGTACTGGTTCC	0.641	0.462
Tdel128	7782353 64	CTGCTTTCCTGTTTCTCGGCCCCTCGTCGTCAGAGTTCTCTCTTTCCTCATCGG AGTAATG[G/A]CGTTTCTCCTCCTCGTCCTCTGACTGGTCGCTCTTCGCTCCT CCGTCCCACTTCTCGTCAT	0.272	0.379
Tdel132	7782353 68	GGTTCTCTTCAAGATGAAGGGTGGTCAGGTGCTTCATGCTAGTGAGGCCAACAG CTTCTAC[G/A]GTACTGAAATTGTTCTGGGAAAGATCGAGCTCTGTYAAGTTG AACAGCACCTCCAGCTCCC	0.347	0.472
Tdel137	7782353 74	GGGTCTGGGATACGCTCAGTACCTTCCCAAGGAGCAGTACCTGTACACCAAGGA GCAGCTG[A/C]TGGACCGCATGTGCATGACTCTAGGAGGACGCGTCTCTGAGG AGATCTTCTTTGGACGCAT	0.545	0.501
Tdel139	7782353 81	TCCTGGTGTGGTGCATGGCTCCGACTCCCTCCAAYGGGTCGGTCCAGATTTACA ACCGCAT[T/C]GTTCGACCTTTCTTCCTCAAGAATGAAGCCAAGATTGACGAC GTGGTGAAGAATCTGAAGG	0.439	0.479
Tdel140	7782353 84	CAGTTCGGCCGAGCCRTCAGTGCGTTTCATCCTCTGCAGCCCGAGAACAGAGGG ATAGACC[T/C]CCTCCTCTTCTTCATCTTCAAGTCTCTTCACCCTGAGAAGGG GAGSGTCGCTGCTAGGTGA	0.298	0.428
Tdel142	7782353 91	AGGTGAACCCGTACACCTTGCGCTTTGAGGGTACATGGGCCACTGGCTTTGACA AACGAGG[C/A]GCAAGCAACGCCTTCATGGCCTGTGGCGTGCTCTACGCSGTG CGCTCTGTCTTTCAGGAYG	0.327	0.429
Tdel143	7782353 95	TGCAKGTGATGATGATGGTCTTCTCTCCATCCCATGAGGGGTTGTCGGCCATGA TTTTGGC[G/A]TGGGTAGTGACGTCCTGCGGGGAAAGCTGGGGTGACTCATTG GGCTGTGTGTGGATCCAGC	0.630	0.458
Tdel144	7782353 98	GGAAGAACYTGAAGCAGAAGCGAGACGAGAACTACCAGAAGATCATGTGTTGTC TGGTTCA[T/C]CGCTACCTGACCTCCACGCGGCAGAAGATGCAGAGCACGGAT CARGCCACGGTGGAAAAYC	0.517	0.490
Tdel145	7782354 01	AGACGAAGGGCCCCCTCCCAGCCGGAACTATCCTGAAGCTGGTGACTTCTGCAG ATGGCAA[A/G]CCCACAACCATCATCACCACCTCCCAGGCTGGAGGCACCGGA AACAAACCTACCATCCTCA	0.546	0.493
Tdel146	7782354 04	GTGCCCTCGAGAGGCAGTATGCAATCACGAGGACGTTGATGACGAAGTGGGCCA CGGTGCA[G/C]AGGAGGCTGAACTTGGAGACGGCCTTCAGGTTCTTCAGGAAG GCGCAGGGCAGCAGGGCTG	0.207	0.302
Tdel148	7782354 07	CACCCAACCCGGGCATYCTCCCCAGCTCCTCAAAGATGGACCTGAGGTTGGCGA GGCTCTG[A/C]GGGGTGTACTGCTCATCCAGGTCCTCGGTGGCCCGCTTGGAG GCGTCTGTATTCTCCTCAT	0.518	0.490
Tdel152 *	7782354 11	GGTTACTCTGAGGGCTGACGTCCACACTTACAGCTCTGCCGTCAGGAGGGGGGT GTTGATA[A/G]GTGTCCTGGCCTGGGTCGAAGTCCCTGGCAGGCAGCTCATCT CCAGTKGGCGGCTCCAGCT	0.430	0.411
Tdel153	7782354 14	GAGCAGCGCCAGTTGAGATCCACAAGTGTGGTGAAGATGTCCGTCAGGTAGCGA TAGGTCTCCC[T/G]CATGTTCCCGTGCTGCACATTGCAACGACCATTCTTCTC CACGTAACGCTGCCTCTTACGT	0.670	0.445
Tdel154	7782354 19	TGAGCAGCAGGAACTCGTTGGTGCTGGCGCGCAGAGTGTGGCCCAGCACTCGCA AACCCAC[G/A]AAGTGGCGGGTCAGCTTGAAGATACGCAGGATWCGGACGAAG CGCACCACGCGCAGGAAGC	0.329	0.444
Tdel155	7782354 23	CRGACCTCAAGAGAGACAAGTGCATTTCCCCAMCTAAACCCAAACGGGGCGAAA CATCATC[T/C]GCGCGAGAAGAAAAGGAGCAACGCCACTCTCACTCCCATCAC CACCACCATAAACACCGCC	0.372	0.429

Tdel156	7782354 27	GCACCTCCTACCTTGGAATGAGCCACCAGGTGGCCGGAGGTGGAAGTGACTCTG GTTTCTC[T/C]GCGCTGACGTCCCACAGCGGTGGAGCGGACTCCTCCACCTAC ATGGGGATCAACCATCGAA	0.391	0.485
Tdel157	7782354 32	TGATTCCCCCTCCCGTCACTGCCAACGGGTGCCTGTGTGCACCTTCCTTCACTC CCTCCAC[C/A]ACCACCTTTGTCCTCCCTGCATCCCACGTCATGCATGCAATA TAGTCCCTCTTGTGATGGC	0.474	0.511
Tdel158	7782354 36	TTTGATGCCGTTGTCGTGGTGCTTGCCGTTTGCGTGCGCTGCCGTGCCGTTCTG ATACTGC[C/T]TAATGTCTTGCTTGGGCAACCGTTTACCCTTCACRTAAGCCT GGACCCAGAAGTTRGAGAA	0.613	0.449
Tdel159	7782354 41	CACACAGCTRTTYAAGCTGGCTCACAAATACCGGCCAGAGACCAAGCAGGAGAA GAGGCAGAGGCTGCT[G/T]GCTCGTGCTGAGCAGAAAGCTGCTGGCAAGGGAG ACACTCCAACRAAGAGGCCTCCCGTCC	0.683	0.437
Tdel161	7782354 46	AATCCCAGGACAAGTACACGTTCTTGTCGTCATCAACGAAGCAGGACTTGACCT CATACTG[G/A]CCGCGAGACATCATTCCTTTAGGCGTCTCTTCAAGTGGGCAC ATRAACTCCTGCTCCTCCG	0.290	0.409
Tdel165	7782354 52	CCATCCTGGCCATGCACATTATGCCCCCTGARCGGTGGAGGACTTCCCGTCTAT ACTTCCT[G/A]CGACGACTCCTGGTCACGACACATGCCCGCAAGGTCTCRGCA GCCTCCACAAACAAAGTCA	0.240	0.364
Tdel166	7782354 56	CCCGGAGGACTTGGGCCATGCCTCCAACCATCCCATCATCCCCTCGCCGCTCTC TCTGGA[G/A]GACCACAGCAGCCTGTCAGTGAGCTTCCAGCTGCCTCCGCCTC CTCCCTCCCCACCRCTGCCCACTC	0.387	0.498
Tdel167	7782354 59	AGAGGATAAACATGACAGGAGCAGGGATAAGCTGCTTTATGAGGACAGGGTGGA TCAAAG[T/C]GTTAAAGAAGCTGAAGAAAAACTCAGTGAGGTATCKCAGTTCT TCCGGGACAAAACAGA	0.358	0.471
Tdel169	7782354 67	CTTTGCTGTGCGCGATGGGCGTGGCTATCTGCCGGGGTCAGAGGTCAGCCAACA CCTGCG[A/G]AGGCTTAGCGTGGTGGAGTTCAGCTTCTACGTGGGATTCCCGG CGCTGCAGATCGCAGAA	0.618	0.460
Tdel170	7782354 71	GTGGTCGAACACAGTGGGGACGTACTCCTCTGGGAAGGCGTCGTTGGCGTAGCT CATCAA[C/G]AGACACGTTTTCCCCACCGCTCCGTCTCCCACCACCACGCACT TTAACATGATAGTACCG	0.696	0.437
Tdel171	7782354 75	GTCCAGAAGCTGTCTAAGAACGAGGTTCTGATGGTGAACATCGGCTCTCTCTCC ACTGG[T/C]GGGCGAGTTAGTGCCGTTAAAGCTGATCTGGCTAAAATCGTGCT GACGAACCCCGTCTGC	0.545	0.517
Tdel172	7782354 80	CTCGGAGCAGGACCGGTCAATGAGTTTGGGAGAGCAGGGGTCGTTGGTCTCAAT GCTGTG[T/C]CGCCGTGACAACAGTGGTCGGACCGGCTCTGTGATGGACAGCC CATTCATTTCAGATATC	0.608	0.464
Tdel173	7782354 84	CACTCTGGACGAGGTGGAGCTGGTCCACTTCGAGCGGGTCCAGTTTCACCTGAA GAACTT[C/T]GACGTGGTCATCGTCTACAAGGACTACAGCAAGAAGGTCACCA TGATCAACGCCGTGCCG	0.704	0.412
Tdel174	7782354 90	TCACAGCTGGATTCGGTTTGGTGTCCTTCTGCGCTCTTTGCCGCTGGCAGTTCG TGAATA[C/G]TCCCGTTAATGTTGGAAGGCTGCTGCTCGCTGTCGGTGAACTG GGACACCCCGCTGTCAC	0.884	0.209
Tdel176	7782354 97	CACCACCAGACCTCTTCCTTTTCTGGCTCCTCGGTACACCGTTTCAACAATGTC TATCAT[C/T]TCCTGCTTGTCCTCCATCGTCCAGTTGATTTTATTGTTGTTAC CAGTGCCCAGATCAATC	0.712	0.411
Tdel177	7782355 01	CTCCTTCTATGACAACCTYCTGACGAAAGTCCCCACCCAGTCCCTGAGGAGAGT ACCAGG[A/G]TTGAAGTTTCTTGATCTCAATAAGAACCGAATAAAAATGATTG AGACAGGAGACTTTCAA	0.633	0.495
Tdel178	7782355 06	GAGGTTCTCCAGATCAGCAAAGGCATGAGGTGTGATGTAGGAAATTGTGTTTCT GGACAA[T/C]GTTAGATCCACCAGTCGTGTCATGTTTGCRAAATCTTTTCTTT TTATACTCGTAACAAAG	0.651	0.458
Tdel179	7782355 14	GTTGAGGAGGACGCGGACGGCCTTGCGGGGCTTCACGCCGCTGCGCATCACGGT GACCAG[T/C]TTGGGCCGAACAAAGTCCTTGGTCTCCCGCGGCTCGCAAGCGG CCTTGGCGGCTAAGGAC	0.424	0.468
Tdel180	7782355 25	AGGTGAGACAGTTCGTCTAGATCAGAGAAATGTGGCCTTTCAAGTGGACACTTC CAGTGA[G/C]AGCCCATTCCTCATCCTACCCCTCACCTTCTCCCATGTCATTG ACGACACCAGTCCTCTG	0.509	0.506
Tdel181	7782355 31	CATGCCCCCCCCCCGGGTACGCGCCGTGCTACGAGGCTCTGCTGCCTCGGGAGCG GATGCG[C/T]ATCGAGCCCATCAAGGAATACAAGCACGACTTCAGCGGCGTCC GCAACCTGCTGGGCCCG	0.505	0.491
Tdel182	7782355 35	GCCTCCTTCCTCACCTTCTCTTGGCCTCTGCGAGTCTTCACAGAATTTCGCACT GCATA[T/C]GTACACTATCGCGTCGAGAAGCTGTTTGGTCATGATTACCTGCC CGCAACCCCATGTGAC	0.329	0.433

Tdel183	778235539	TTTCGGACAAAGGRCCGTCAAGCGGCACCATTTTGGCTATTGGAGAGCGAAAAC GCAAGAC[G/A]GTTGATGGCGGCAAAGCACCGGCCAAGAGAGAACGTTCTGAG AGCAGCTCCAAGGCGGGC	0.810	0.257
Tdel184	778235550	GGAYAAAGATGCCCGCACCAAGTTGAAGACCTCAGAGCAACTGGACCAGCCCAT CACAGC[T/C]TGCTGCTTCAATCACAACGGCAACATCTTYGCRTACGCTTCCA GTTATGACTGGTCAAAG	0.548	0.440
Tdel185	778235557	ACAACCACCCTGACCCTGCCTTTGGCGTGGCCAGGATAAACACCGTGAAAGACA TGATGCT[G/A]GGAAAAGACGACAAAGCCTGGGCCATGTATGTGGACAACAAC CGCTCCTGGTTCATGCAC	0.384	0.448
Tdel186	778235561	TGCAGCGTCCACTGACCCGGCCACCGCCATCTCCATGGGGCTGGTGAGCGTCAG TTCTCC[T/A]GACGTCAGCTCCGTGTCCGAATCATCTGGGAAGGACGCTCCGT CCCAGAGGCCGCTGTGC	0.524	0.509
Tdel187	778235564	TCCTGAGCGAAGAGGYAACCTGATGTCCTGCATCTTGCAGGCTTGCAGCAGGTT GTGGCG[A/G]TCGCTGCGCTTCTGCTCCAGCTTAGTCTCAATGGCCGTCACCT CCTTCTGGAGCTGCGTC	0.281	0.400
Tdel188	778235568	GCGTCTGTATTCTCCWGCTTCGTCCTTGGAGCTTCTCACGGTGTCGCTGTGCTG GGCCAC[G/A]GTTTCGGTCATGGAGCCCACCTTCCCTCTRAACCACTCCTCCG CCGCCTGCATGTTCCTGGCTG	0.425	0.487
Tdel189	778235573	GGAGCTGGAGGGCCAAACTTGGTACCACGGCAAAATGAGCCGCAGAGACGCTGA AAAGTT[G/A]CTGAAAGACGATGGGGACTTTCTCGTTCGTAAGAGCACCACAA ACCCAGGGTCCTACGTA	0.484	0.502
Tdel191	778235574	CACCTCGTCCGTCCCCTCCGCCACCGTCACGCCCATCAACGAGGGCTTCATGTC ACACGC[G/A]AACTTCGACGCCACGCGAACCGGAAACCGCCGGAAAGGTGAAA AGCGCCTCAGCCGGAAC	0.402	0.469
Tdel192	778235576	GGGCGTCGCCAGCGAAGTCGCCCGCCAGACGGCCGAGGAGAGCGAGAGTGGGAA CTCGGC[G/C]TCCTCCGCCCTGCTGTCAGAGTACAGCATCACCAACACCATGA ACACMGGACTACGAACCCC	0.604	0.478
Tdel193	778235577	AACGGTGGCGTTGGACGGCGTATAGAAGTCCCCCATGGGCGACTTGGTCTTGAT CTTGAT[G/A]CAGGTCAGCTTGGAGCCTCCTGCTTTGGCTCGCTGGCGTGGGA GCGGCTGGCTGGTCTCG	0.331	0.445
Tdel194	778235579	GCRGAGCGCCTTACTGCAGAGCGGCTTCACGCTGAGAGGATGGCGTCTGTGTCT GATCCCGC[T/C]GCCAGGCTGCAGATGCTCAATGTGACACCGCAYCATCACCA ACACTCTCACATTCACTCCCACCT	0.557	0.500
Tdel195	778235587	CAGGGAGCTGGCGGAGCAGCGGGTCAGAGAGCAGATCCACATCGACGCAGCGCT GGAGCA[A/G]CAGAAGCTGGAGGACCAGAAAGCTCTGGAGAAGGCCGTCAGCA CCAGCCTGACGCACAGC	0.364	0.455
Tdel197	778235590	CTTCAACTTTCCCTTCAACGGCAAGAACCACAGAGACTTCACCGACAGCAAGTA CACGCT[G/C]CACTTTAACCAGTACGTGAGCGAGGTGGTCCCAGACAGCTACA AGAGACCCTACCTGATA	0.531	0.496
Tdel198	778235594	GGGGGGCAGCGGAGGAGGCCTCGGGAGCGGAGGGGAGATGAATGAGAAACACTG CTGGAA[A/G]CCCACCACCGTCCCGCAGCAGAAACAGGAAATGGGCTCCCAGC TGCAGTACAACATCATG	0.396	0.506
Tdel199	778235598	GATTTCCTGCAGGGACTTGGGTCAGGGGGACTGCCAGGGCTGGCTGTGGAAGAA GAAGGA[T/C]GCTAAAACCTATTTCTCACAGAAGTGGAAGAAGTACTGGTTCA TCCTGAAAGACACGTGC	0.622	0.471
Tdel200	778235603	TCTGGCTCTCCWCTCTGTTCCTTGATGTACTGAATACAGTCTCTGCCCGTCTTC TCTGAGTCAAC[T/A]ATGATGGCGTCCATGTTTTTGCCCAACACTTTGGTGAC AGCGATCTGATACTTCTTCTGG	0.486	0.497
Tdel201	778235606	GCAAACCTCCATTGCGTGCGAGGCCCCTGAGCATCTGAAAGGAATGATGGTTAC CAATAT[T/C]CCCAAGCTGGACTGCGAGGCTCCTACTGTCACAATAACATACG AGCCAAACATCGAAAGT	0.385	0.479
Tdel202	778235611	GGGGGACTGAGGCATGTGCTGCGTCCCAGTGCTGCCGTCCCACGTCCTCCAGTA GGCTTT[C/A]CGGCTGGACTCGGGGGACAGGTGCTGCATCACGTTCTCCGCCG AGTCGGGGGTGGCCGGYCCAGAG	0.299	0.409
Tdel207	778235623	GCTGCTGGCGTCGTCCTGCGAGGAGTTGGCGGTGCTGTTGCAGATGAAGGCGCG CGCGTA[G/C]AGCCAGTAGTCGGTGCCGATGGCCACCGTCATCAGGGCGAAGG CGGCGAACGCTCCCATG	0.553	0.509
Tdel208	778235628	CTGCGCCTCGTTGTGGAATGCTCGGAGTTTGGCCACCAGGTCTGTGGAGTCCTG CGCTGC[G/A]TTCACTGCCAACGTCTTCGGGATGACCAGCAGAGAGCGAGCAA ACTCAGCGATGGCCAAC	0.332	0.450
Tdel209	778235632	ACCCAACCCGGGCATYCTCCCCAGCTCCTCAAAGATGGACCTGAGGTTGGCGAG GCTCTG[A/C]GGGGTGTACTGCTCATCCAGGTCCTCGGTGGCCCGCTTGGAGG CGTCTGTATTCTCCTCA	0.502	0.500

Tdel210	7782356 35	GGCTTGGTATCCTTCTGCCATGGCTACGTGCATTCTCAGGTGATTGTCTGAGTG TCTATG[T/A]GCCGTGGCCAGCTCGGAGCTCTCCGTCATCTCAGAGGAGTTTG TCTCATTGCCAGAATCG	0.158	0.267
Tdel212	7782356 38	GTTCAGCTGGCGCTCCACTGCCTGGTGCTGCTCCCGCTTCTGATTGGACGAGCG TTCGTC[C/G]TGGTTTCGCTGTTTCTGGGCCGAACTTGAGTCACTGGTCACGT TTCGCGTTGCCATCTGG	0.542	0.515
Tdel213	7782356 39	CGGGACGATCTCCGAGCCCTGCCGGATGTACGTGCCCAGCAGCTCGCGCTTAGA CTCGGC[G/A]TCAATGAAAGTGATGCGCTCAATCATGGCCCAGATAATGATCC CCAGGGCGAAGATGTCC	0.251	0.376
Tdel215	7782356 40	CATCCCTCTTTCCCTCTCCTTCTTGCCTGCACCTTAAACCCAACAATTCCTTTC AATTATT[A/G]CGTCAGTAGTGAGCTTCTGTTTCGTTTTCACCTGTAAATTTC TGTACAAATAAAATATCAG	0.679	0.445
Tdel216	7782356 45	TCATCTCCAATGATGGTCTCCCTTCCTGATATCTGGGGCTCTATCCTCTGAAGC TGCTGG[A/C]TTTTTCCAATAAACTAAATGGAGCCTCTGACTTGGATAATGCC GCCTCTGACCAGTGGCTC	0.577	0.498
Tdel218 *	7782356 47	CAAAAAATAAATAAAATCAGTCCTCGCGTGATAAAGTACAGCTTACAGTTGACC CTCCCCC[T/C]AAAGAACGGTTAGCTCTGTTGTATCTTAAGTGCTCACTTTGA AGAARGTAGAGTCTTCTGT	0.468	0.412
Tdel221	7782356 57	CACCGGGTGATTTATTGATGAGTTTTATCATCTGTGTTTTTAAATAAGAACAAA AAGTGAC[A/C]AAGAGCCGTCCGCCTGTATGAGGCGAGAAGGCCGCGTGGATG GATRGGAGAGTTGGGGAGC	0.302	0.408
Tdel222	7782356 67	GATGGATAGAGAGGTGKAGTGACTGGGTGTAATGTGGTGGCACGAGAGGAACGA TCAAATCTTCTTGACC[C/T]TTGTCAGTGCCTTATCCAAATCTGCTTGTAAAC ACAAACGTTCACTGCCAAAGCGGAGAGC	0.645	0.457
Tdel224	7782356 70	CTGGTGCTAGTGAAGTTTTTGGTTACATTGTTTTTGAATGTTGTGAATGCACTA AATTATT[A/G]ACATCCTTCCATCCATCCTTTTCATTGCCTTGCCAAATTTAA AGCAAGTGATTTGTGATAA	0.649	0.457
Tdel225	7782356 71	AGAGGAGGAAGCAGCTTTGTCTTCTGGGTTTGGTTTGACTTTGTCTTTTTCTTC CGTTGTC[A/G]CACCCTTTTGATCGGGAGCTTTGTTTTTGCTCTTCTTAYGTT TGTGYTTCTTTTTCTTCTT	0.640	0.455
Tdel226	7782356 72	ATAAATATAGAGATATACTTGTCGAGTGGTTATGACTTCCAGCTGTGATATTTC AGTTCCT[C/T]TGAACCTTTTTTTGTTTGTGACTTTTGAAGTGAATGTTAAAA GTTCTTCGAAAAAAGAAAA	0.615	0.481
Tdel227	7782356 73	CAGGCCTGGTCACAGAGATTCTAACYGCAACAGAAGAAGACCAGGCCAACTGGG GTTGAAG[G/A]AAAACTGTACAGAAGAGGAGAGGACAGGGCTGGGATGCAGAA GGTTGGAAGGTGTGAAGTC	0.415	0.452
Tdel229	7782356 74	CTCCCAAAACCAGGCAGTGATGATGCCTTTTTGAGTCACTTATGTTGAGATCGC ACAGATT[C/T]TGAGGGACGAATTTGCCATTTGCTCACCAAACCAACACGAGC CAGTTGGTGAGGGCTGCGA	0.623	0.454
Tdel232	7782356 75	TCCCAGAACCATGCTTAAGGAACACTAGCGTGCCCAGAATCGATATGTTGTGTG GTGGAAC[T/C]CCCAAATTAAAGCTTTCGTACAGATGGTTTATGCTGCTCGTG TTTTTGTTTTTTTTTAAAC	0.565	0.499
Tdel233	7782356 76	AGCTGTATGCCTGCCTCATTTTCTAAGATCTGTCTTTTATCCGCCTGCAATGTT GCCTGC[T/C]TGTAAAACCATCTTCATGACCCCAGTTTCTGAGCAGWTSTTGA CAGAAGCGTTRTAAGGAGCTGCAAC	0.691	0.441
Tdel234	7782356 77	TCGTCTGTAAACRGCATGTGTCACCATCAACTGTCAGCCGGCTCAATATATCGC CAGAGAA[A/G]TCAGTGCTGATGGACTTCTTGTGCTGACATAGCTGTGTGAAA GCTGGAGCTCTCTCTCATT	0.656	0.453
Tdel236	7782356 78	ATAACTACCAAAATACAAGGGAAATGTGCAAATAGAAAGGAGGAAGAGCTTAAA AGATATA[T/A]TCTAGTGTGCAAAATAAAAGCTCGGGTGTACATTCTATCACA CTGAAGCAAAAMGAACCAG	0.526	0.494
Tdel237	7782356 81	GCTTCTCTCTGTCCTGTCCTCTCTCTGTCCCCGTCCTCCTCCTCCTCCACCATA CGAGCCT[C/T]TAGCATACAGTCTGCGTTAAATGTCTCGTTGYARTCTATACA TTTAGCCACGTTCGCTGCA	0.614	0.473
Tdel238	7782356 82	GYCTAGTGGTAAACTTGATACTTGATTGRTGGAAAGCAGCGTCTTTCTTCTGCG AAGTGTG[T/A]TATGCACGTTCTGGGTTTTTCGAAGACCGACTTGCAGTACTC CGAAGCTCTCYTACTGTAC	0.525	0.507
Tdel240	7782356 86	TAATTTACATMTTGGTCAGTGCTCAAACCTCAATACAGCTGATAGTCTAATATT GAAATGT[T/A]GGTTTTAGTCAATCACACAGTAAAGACCAATCAAACCRAGCA YGGGATTGTATGTTTGTGT	0.405	0.478
Tdel241	7782356 87	AAATAATATCATRTGAAGTGATATTCCTTAAAAAGTYCAAAACTAATTTGAGGC TTCGCTCTTC[G/A]AATCTCTCCTCTGTGTGTTTGTGATGAAATGAGAATAAA GTCACATGTTTGATCATCAACA	0.620	0.455

Tdel242	778235689	AGATTGTCAATCATTGAAGGCCKKGTGCTGTGTCGACCGAGCCTAACGGGTCAG CCAAGGT[A/C]ATGCTCCCACATGACACGTGCATTATCCCACATTTGAGAGGA CCCCTTGTTCAAMATGTAT	0.609	0.469
Tdel243	778235690	GCGGGGAGGCGGTGAAGGCCGACGAGCAGAAAGCTAACGTGATCCTGTGCGGAC TGCTGGA[C/G]AAATGGTTCCCGGAGGAGGTGACGAGGTCCAGGACGCTGGGC CAGGTGGACGAGCTGTGCA	0.600	0.474
Tdel247	778235691	ATAATTCCTCTGCCATTCTTTCGTCACTYGCAGTCCTTTTCACAACTTGTATTG AAAGTAA[A/T]GTCACTGTCACATGAACAAGCAGATACAACTCTTTAAAAACT AAAAATAAATAAATCACTT	0.491	0.494
Tdel248	778235692	CTAATTTTTTCTCAAGATTTTTTGCTACAAGCCCACTTCATTTGCCTTCTGAAA GTCCCA[C/T]TCTGATAGCCCAAATGAAAGACGTCAGCCACACCACATGTTGT ATTTGGGAATATACACTA	0.561	0.518
Tdel249	778235693	GCATGGAGGACTCAGACGCGTCCTTGTGACATAAATTTAGTTTTCAGGCCCCGC ATAGCTC[T/A]GAGAGATGATTCATTCCTGACAGCTTTGACCCAGSAGAGGAC AGTGTTGCTATCTGAGGCT	0.346	0.432
Tdel250	778235694	GAGTGAATTTGTAGCACKACTYTTCCCTGGRCTGCTTGGTGCATCAAGATCACT CAGAGCTCCGAGATC[A/G]TGAAGTTCACTTGATCGCAGACGTTTGATCGAAG ACTTCAACTCTYCCTCTKCCAGCTTKGATACAATTT	0.358	0.467
Tdel251	778235698	GAGTTTGAGRMCACGAGCCCCACAGTGGCCGTAAAGATGTGCCTGTAGTGTAGC GGTTAAGTCACAGT[G/A]ACGCCAGCACTCTCCTTGTTCTTGTCTATGTAGGT GGGAGAGTTGCTTTTGTCTGTTATGT	0.258	0.382
Tdel253	778235704	CATCCCATGTTGTACTGTACTATATCTGACTTTATCGTGGGCAAGGCTGATGAA CGAAGGC[T/G]CTCTAAAGGTGGAAAGGGAAAGAAATCAGTTTCATCCGATGG ACATATCTGTACAGTGAAG	0.431	0.480
Tdel254	778235709	TTTCGATCGCACCTCCTCCTWYGATTGTTTTGTCGTTGCACAATGGAAAACGCC ACACGAG[A/T]TCACCTGGTCCCCCAACAAGTGCAATTTCCCTCCACCACGGA CATCGAATCAACAGTCTTT	0.560	0.468
Tdel255	778235710	TGCAGAAAAGCAACAGGGTTTTGAAGTTGTACAATGTTTCAGTGCATCATGTGA GGGCAA[C/T]GAAATGTACTGTGAATCTGTTTTCCCCTCTTTTGTTCATCAAC GAACAGTTTCATCCCTAA	0.542	0.463
Tdel256	778235713	ATAAGCTTCAGCTCATATGTGACAGTAGTGCTCTGTCTCATTTATGACTTATTA TGGAAA[G/A]GTTTCTCACACGTCACTCTTTAGCTTGTATATAATGTGTGTAT CAGAAGACAGAATGTCAC	0.535	0.478
Tdel257	778235718	TGTCTCCATATAGTGTTTAGTGCTCAGTTCAACATGCCCTTCCGGTCACAGTCT TCCGCAT[A/T]GTTCAATATTCTTTGAGGCGGTTCGTGGTCGCAGGATTCATC ATGCTGGCCTCTCTCCGTG	0.371	0.454
Tdel259	778235721	AGTCGGTGACTCCGTCCAGGCAGCCGTTGCCGAAGGTTCCGCTGAGTCGTGAGT CTCGTCC[C/G]AGAGAGGAACTCCCGTACGGAGGAAACAGACTCAGATCGGCT CCGAGGATCGGCTCCATGA	0.509	0.482
Tdel261	778235723	ATAATAACGTGTCAGGATTCAATGGTTGGAAAACTAAGAGGTTGTTGCCCGTCC CTTTCTC[A/T]GCAATTAGTCAAATGGAACCCAGAGAGAACAAAACCCTGTCC TKAAACATGAACAATATGT	0.455	0.510
Tdel262	778235726	CTSCTTTCTGAGTGGCTGCYGTGGCGGAGCGAGATGGCGTTGAGTCCCGAGAAG AACCGCT[T/C]GGTCGATCTGCGCCATCACTGACTCCGCCTCCTGGAGGTCTG GCCACACCTACTGCTCTGG	0.686	0.423
Tdel264	778235728	TTTCAACACCAAGCGGAGGCTGTTTGYTTGAACTTTTCTTTGTTACTGAGGATA AACTTTC[A/G]GGAGGAGGGACCGGCAGATAAAATACAGGTGACTTTAACTGG ACTTCTTGGGACCACAAKGA	0.401	0.453
Tdel266	778235729	CTGCAGCAGGCTGTGTTCTGACGTCTCTCCACGCTCGTGCGTGAACCGAGACAG TTTGGAC[A/G]TGTGCTGCTCCAGCGGCCTGCTCCGCTTGCTGTCTGGYTCTG CGGAGYTGCTCCTCTCCTC	0.418	0.477
Tdel267	778235730	GCTCCCAATCTACTCTGACTTCCCACCCATGTGTCCATGTTGCTGTGTGTGGAT GTGAGAG[C/T]GTGAGACACACCATGTGGTTGTTGTGAAAAATGTGAACAATT CCTCTTTGTGTGGATCTCC	0.428	0.503
Tdel268	778235731	AGGATACATTTCGGTAATTCCCCCCCCCCACCCAAAAAATTGAATGAAAATGAG GCACAA[A/C]TTTTGCATCTAACATAAGGCATTATGATGGTCACGDATGGCAG CGTAATGTTAGTGTGCGT	0.676	0.397
Tdel270*	778235733	TGCTCTTCCGTSAATTTACAGCCGGCCTTCCTCTGCATCGACACAGAAAGCTCA TCACAAC[G/C]CCAAGAAGCAATGGCGTGTTGTCCTCCTCTTCATCATCATCG GCTTCATCATGGACCGAAG	0.614	0.258
Tdel271	778235734	AGAGTACGCACAGCTGCTGGAGGACATGCAGAACGGCTTCCGTACGTTGGAGGA ATAGGCT[T/C]CCGGCCATCACAGTCTTTCCAGCTTTCATCTCCGACCCCTCC ACTGCTGAATCACATTCCA	0.679	0.438

Tdel272	7782357 36	TGGCAAACTCTGTACACCTGTCTCTCCTTGGCCTCATTATGGGTGTCAGGGAGC TCATACG[T/C]AGGGGAGTCTAAACTCT TCCTCTCCAGCTGTTTCGTTACCTCAGCCATTGTCTCTGTGGTG	0.677	0.437
Tdel273	7782357 40	ACCCTCCAGGAGCCCCCCCTCCTCCTCTTCAGCTCATGCCACATCCTCTGCCCG TGCAGTC[A/T]CTTCCAGCTCAGCTTCCAGTCATTTCTCAGGTGCAGTCCCAT CCTGGAAAAAGCACGACTT	0.533	0.495
Tdel275	7782357 43	AGGGRAACAARGCGACGTGGAGTTTCGYTGCRACTGTCCACTCTGTGCTTCTAG CTTTGTC[A/G]ACATCCTCAGGGCTCCAGAAAGCGCTCCTCTCCCYAGGCCTC AACCAGCTACGCCCTTTT	0.504	0.483
Tdel276	7782357 47	CTMCCAACATTTCACCTTCTGCACATCCTGAAAGGTTTGCATATCGCCATAGCA ACAGCCT[C/T]TTCTCATCCTCTCCCACTGCTACGATCAAGTTCAACATCTCG TATATAGAAAAGGGGAACC	0.415	0.468
Tdel277	7782357 49	GGTGGGCGGTGGGGAAGCAAGTTAGGGAAGGGAGCAGGGATCTCACAGCTGCCA TCTTTGC[C/T]ATGCTGAGGTTTAGGGGCGCCGCTGTTGGACGAGCTCTGTGG GTGAGCAGCGTCCATGGCA	0.207	0.312
Tdel278	7782357 50	GTCGTTCCATGTTAGCCWGGTAGTTATTGAAGCTCAGCATGCATTTCATCAACA AAGCCAT[T/C]GTGTAAATTACAGACATGAAGAACATTGCACTTCWGTTGCTG AGGGCTTTCATCGTGTAAA	0.619	0.456
Tdel279	7782357 53	ATRTGTCGCCCGGAGCTCTGGGTGTGCAGACGGAAAGTTCCGGGAGGCGACATT CCTCGGA[T/C]GAGAACGGACACGCCGACAGTCCTGATGAGACTCTGTCCCCG ACCTCGGTGATCTACTTCA	0.480	0.494
Tdel280	7782357 55	CTCTGTAACCGTCATTAGACATCCAGCACATGGCTCCGTCTGCAGAACCGGCTT TCACCAA[T/C]GGTGGCGACTTGTGAGTCACTCCTCCCTGGTTGTCATTAACA GCGAGCACATAGGACGGAG	0.342	0.457
Tdel281	7782357 57	GGGAACGAGGGAAGAACTGTCGTGACAGTGTGACAAGGAGACAGGAGAGAGATG GTGTCTA[C/T]TTTATGGGTGAGTCCAGGATTTCTTCATATTCTTCCCTTTGG CGCCGCCCGCCACCTCGAG	0.545	0.491
Tdel282	7782357 62	TCTTCTTCAGTGTACAGTACGGGAATCTGATCGGAGGACATTCAGGGTGGACTT CAATCTC[C/T]TTCACCAAAAGATAAGTTCATATGGTTTTATTAYTTGTCGGA CAACAGGAAAACATGGAGA	0.632	0.461
Tdel283	7782357 65	AGAGGYCGAGGCTCCTAGGGACCAAGCAGACGCAGAGGCCGAGTCACTGAAGGA CCCTCCA[G/A]CGGCCAGAAAGAAACTTGTTGAGCAAGATATGAACAAGCACA TTGAAGGAGAAGCAGGACT	0.244	0.345
Tdel286	7782357 67	GCCATGCCGGCTTTATGTTTAGCAGGCTGATTTTAGTGAGTCGAGTTTTTATTA CGTGTT[A/G]AAATGAAGCACACACAGTAGTTAAAGATTGCAGTTCATTGGAT AAACATTCCAAATCGATG	0.693	0.413
Tdel287	7782357 70	AGCCACGCCTTGTGGARAGGGACTTTTKAATTTTGAGAGTGTTTGACGTGGCCG CTACTTT[G/A]ACGAACACCTGGCTCATAA CCATCGACAGGACTCGTGTTTATTTACAGAGGGGAGGAGCCT	0.414	0.472
Tdel288	7782357 73	ACATGAATAGATGATCACAACMTGGCCTAACATRAGTACACACACTCACGTCTG TTCCTCT[T/C]TGTGTGTGTTTCACTCATCAATGTTTTCCTCTGCTCAAACTG GCCCAAAGTCCTCTGATTG	0.312	0.433
Tdel289	7782357 78	AGCAGCCTCCATCMCATCCTCCTCATCCTCCTCCACCCTGACATCAGGACCTGA GACCTCC[T/C]GTTCCTCACAGAGGCAAAGACCATTGTCAGCTGCCTTTCTTC CACCAGGAAGTCCAGTTCA	0.293	0.411
Tdel290	7782357 79	TAAAGCTCAGAGTACTGTCAGAGTGGTTCAGACGGGCAAAAAGAAACAAGCTGA GAAGTCG[C/G]TCCCGGATTCAACGGTGGAGAAGATCCGTCCATCKAGTAAAA AACCAACAGACCCCGTGAG	0.731	0.378
Tdel293 *	7782357 87	GAGGTTTTCTTTTTTCCCGTCCACGTTCCCATCATTCCTTACACGATCCAACCT TTGCAAG[G/A]TGGTAAAAACTTCATTTTGTTGTTGAAAAGCAGCTAAAAAGT TRTCGTAGCAACTAATTTT	0.301	0.380
Tdel294	7782357 89	CGGTGAACCGTTCCTCCCTGCTGGTCCCCGAACAGGCGAGCAAACCACAGGAAG TGAACGG[G/C]ACCAACGGGAGGACGCCGCTCCCTGTGTCTGAGACGGTCAAA CAGAGCGAGAACTCCACCG	0.513	0.502
Tdel295	7782357 90	CTCTCTCACGCCTTTCCCTCTCTTCTTTCTCTCTCTGCAGGCGCTCTTGCTCCA GGGCTCG[C/T]TGTCTTTCGAGCTCCTTCTGCCTCTCCAGCTCCAAGGCCCTT TCTCGCRCTAAGGCTTGCT	0.665	0.449
Tdel296	7782357 92	CCTTATACAGCGGAGCCGGTTTCTTGGACAAACTGCTGGTGCTGCTGTTGGTCT CAGAGAC[A/G]TCGTAGGACAGAGGGTCCTCGCAATCTGASAGCCGCTCCTCA CTCTCCCCGTTCTTTATAG	0.679	0.434
Tdel298	7782357 97	GACACAGAKGCTCTYTATTCCCAGGACAGTCCTCTCGGGCTAATGTGCATCATT AGTCTCT[A/C]ATCTCGTGAACTGATCATTTTAAGGTTTGAATCCTTCAGGAG ACCTCACRAAGYGGCCAGT	0.871	0.228

Tdel299	7782358 01	AAAGCCTCTSWGAGAGACCTGTGAAAACCGTAGACCGCGTGGCTCCACACGGAG AAAGTGC[T/C]GCCATCCACTCTTTTCTCTTTTCTGTCGCTTTGACCTTTTCC ACTCGCTTTCCTTTCTTCT	0.253	0.351
Tdel300	7782358 03	GCCAGAGCTCCTCAGGCAGAGCAACGCWGGCCCGATCCTTCTCCCGCTTATCAG AGGAACG[G/A]CATCGAGACAGGGAGAGAGAGAGGGAGAAGGACAGGGGCAGG GAGAGAGAGAGGGATAGAG	0.380	0.478
Tdel301	7782358 04	ACCCACATTCCACAACAGACCAGTGGTACGTTCTGCATCTGAGGCTGCAGACTC AACATCA[C/A]TGGTGTGGGTTGACCCAGCACTTGTTACTGAGCTGTTCTGCG TGGGGTTGGGGTTTCGGAG	0.535	0.522
Tdel304	7782358 05	CTTGAGTTCAAAAAGGAAAAGCTGATGGTTTTCGTGGGGGGTCTTGAAGCTGAT TCAACCA[T/A]CCTTCAGCGCCCGATTCCCACCTTCTCTCTCTATCTSTAGTA TGGATGGTTAATGCCGTGA	0.230	0.344
Tdel305	7782358 06	AACCAGGGTCCATGAAGGGCAACCTTCCGTCCGTGGACTACGAGTCCAGTGAGG AGGAAGA[A/G]GAGGAGATGGAGGAGGACGAGGGAGTGGTTGTCAATGAATTG GCCAAGACAAGCTCCAAAA	0.611	0.451
Tdel306	7782358 11	GGGGGCCCGACTGAGGCAGCAGAGACAGAGAAGTTGACTGAGTCTTCTCCAGGA GTTTGTC[T/G]GCTTTCTCTATTTTGTCTCTTTCCATCTTGACCTCTGCTGGG AGGTTCTGTCCACCCAGCT	0.125	0.179
Tdel307	7782358 14	GTCGTACCTGTGGTAGATGTGGTAGTAGTTGAAATAGCAGCAACAACAATAATG ACTTTGC[G/A]TGGAGCAAAACCTGAGTCAAATCCTGCCCGACAACGTGAGGT GTGTWGGATTTCTGCTCAC	0.363	0.451
Tdel309	7782358 16	CCAGCCCATCATCTGAGAACGGATTCAYGYGTAAACAATCGATTTTGGTGAAGA GTGTCAT[T/C]GTTGATTGCTCTGCAGTGATAATCGTTTCTATCTCTCCTCAT CTCTCTCTGTGTTTCTAAC	0.406	0.478
Tdel311	7782358 17	TGATCCTCACAGCCAGAACGCTGCTATGACAAACATCTGGAACCAGGCGCTACA GCTGCTC[C/T]TTAGGAGGAGAAGCAGCAGGTCAGAGACAGCCACACTTCCAC CTGCTACCTTCCCAGCTCT	0.732	0.362
Tdel312	7782358 18	TGGACCCCAGACTCATCTACCAGCTYATCCAGATCTCVCAGAGGTACCARATCC CCCCCGAGGACCTGCTGGACATGCTG[A/C]GGATGGGAGAGAAGAGCAGCCAG AGCACGTTGAGGAAGAGCAGTCAGATATCCAACAAGTT	0.635	0.450
Tdel314	7782358 19	GCAAGAAACTAGGGTCTGTTCTCGTTACTGATCTGTTGCATTTCCTCAGACATC GTTGTGG[C/G]TTAAACGGAAACCAGAAGGAGCGACGCTGTGGCACCRTGACA CAAGCCAACRTACCAGACA	0.741	0.379
Tdel317	7782358 20	GGATTCAGGATGTCCTCCRCTTTGCTTTTGCAATGCATAGTTTGCAACTCTAAA CTAGAAT[A/G]TGTGGTCTGAACCTTTGAATGAGCATGCAGAAGACTTTGTAA CAGCTGTCAAACGACAGGG	0.747	0.373
Tdel319	7782358 21	CTAGAAATAACACAGACAGAACWTCTGCAGCTTCGACCCTCATTTGGTGCTCTG TAACCTC[A/C]GATCGAGCACTTCCAGCTGCAGAACCAGGTTGTTTTTTTGTT TTGCGTTTACAAAGACACC	0.753	0.339
Tdel320	7782358 22	AACAATGGGGGGGGGGCGACGAAGGAGAGACTCTTTCTGGCCAACGAGAAGCAA GCCAGTC[C/T]GTGTTTCTTTTCTTCTGGTTCAGATGGTCGCAGTGAGAGACA AACTTACTGTGTCCTCTGA	0.684	0.418
Tdel321	7782358 23	TCTTTCTATATCTTTGCAATYATTTTTACTTTGTTTTTTGCTAATTGTTCTTCT TGTTGAT[G/A]ATCACTGTTACTGTTGGCACCGGTTTTGCCTCTCTTGTATTT GKTTWTTTGGACTGCGTGG	0.356	0.453
Tdel323	7782358 24	ATTGACTTGCTGCATGAGCTGCCTGCAGTTGAAATGTCACTTTGACTATGAAAC TGTCAAA[T/G]GAACTTCATCTGAAACAGCAGATCCAGTGTCTACCCGTGCAC TTTAACCATTAATGCCTA	0.391	0.485
Tdel324	7782358 26	GCTCTCCATCCCCTCCAGCGTAGCAGCAGCTCTCTATCACGCATCTCTCCATCC TCTGGAC[G/T]GATTTGCCTCCTGTCAATCGTCTAACCCCTGGATCAAGCCTS CCTCTGCTTTCCACTTTGA	0.499	0.518
Tdel326	7782358 30	CTGATAAGGAGGGCTCCACCGACGACGAGTCCCCCTCCTCGTCCTCGTACTCCT CCGACGT[C/T]GGCGGCTACCACCAGAAGACTGAAAGTGGTGCAGGTCAAGCG GAGGTGAAGACCACTGCGC	0.729	0.369
Tdel327	7782358 34	GCACACTTTTTGCCACAATCGTGCTCAGTCTGAGTGGTGAGACCCCGTCTGGTA CATTCCT[C/T]CTGCTTTCCTATATTGCTCGCTTTCTCCTTCCTTTGTCCCTT GTTTCGCACCTCTTCATGG	0.400	0.443
Tdel328	7782358 35	GCAGGAGCGATAGTCCTTCCCTCAGAGGGGTGTCCCGTCCAGGCAGCCTCTCTC AGTTCG[T/C]GCCCCAACTGTCCCAATGCTAACACCCAGGACTACAACAAGCA TGCAGCACAGCGCCTTCC	0.436	0.496
Tdel329	7782358 36	TTCAAGGTAAAAGGGGTGTGTGGTGGTCCCTTGGCTTCTCCCTGGCTGCATCAG AAGATCA[T/A]AAGAGCAGCCTAACGCTTGTGTACTGATGTGACCTCCTAGGC CTCGAAGCGCTTCGGACAG	0.395	0.422

Tdel330	7782358 38	AGCTGTCGGACAAGAGCGGCCTCGCCCTCGATGGCGAACAGTACCGCTGCGAGA GCCAGGA[G/A]CACAGCGAAGTCTTCAGTCCCGTTCAGTCTCCCGCCGCCAAA CAGTCCAAACACACGCCCA	0.344	0.467
Tdel331 *	7782358 42	CAAGGCYTGCACAGTCTGGAAAAATAAAATTGATGCTTATACATTTTGTAATAA CTGAAAA[C/G]AGAGGACTGATCAATTTTAACCAGTAGTGTACTTAAGACTGT ATTCACTCTTTTGAGTTG	0.692	0.324
Tdel332	7782358 46	TGGACATGCAGGATGAGGGGAAGGAGGGCRGAGGAGAGGAAGGAGAGAGGTTCC TAGATGT[C/T]GCAGGGAAGGAAGAGGGGAGCCGTGTGTGTGACTTGTGAGTG AGCTGAGGGAGACGACTTA	0.671	0.418
Tdel333	7782358 47	AAAAGAGATTCCTCTAATCCTGTCAGTTTTGTTGTTTTAAAACAAACGTCAAAC ATGCATC[C/T]GTCAGCCCAGTTATTTCCTTTGTTTTCGCTCATTACTTGAAC GCGGTGGACCTCRCTGATT	0.694	0.408
Tdel334	7782358 48	TGACTGAAAGAATCGTTTTGGAAGAGGAAAGGTGAGTGTGTGTAAAACTCACCA CAGACAC[G/A]GATCATCTATAAAGTCTGAAGCCCTTACAACAAAAAATCATG CTTCAATGTGGAAAAAGAT	0.319	0.435

3.3 Fine scale genetic connectivity via parentage analysis

3.3 Fine scale genetic connectivity via parentage analysis

Publication 5: Retention and fish larval dispersal potential on a highly connected open coast line

Publicación 5: Retención y potencial de dispersión larvaria en una costa abierta

RESUMEN

La conectividad es fundamental para la persistencia y la capacidad de recuperación de las especies marinas, el establecimiento de redes de áreas marinas protegidas (AMP) y la definición de unidades de gestión para la pesca. En el medio marino, la comprensión sobre la conectividad sigue siendo un gran desafío debido a las dificultades técnicas del seguimiento de las larvas. Sin embargo, recientemente, el análisis de paternidad ha proporcionado un medio para hacer frente a esta cuestión con mayor eficacia. Este método requiere un amplio muestreo de los padres, que a menudo no es posible para las especies comerciales, protegidas o de aguas profundas. Mediante análisis de paternidad y de hermandad hemos medido de manera directa la conectividad y la dispersión de las larvas de *Tripterygion delaisi* (N = 1573) con 178 marcadores de SNPs. Es el primer estudio de estas características realizado en un ecosistema marino templado y en una costa abierta en el Mar Mediterráneo. El análisis de paternidad sugiere una conectividad limitada, con una disminución del éxito de la dispersión a 1 km de distancia, y aproximadamente el 15% de los juveniles recogidos se identificaron como auto-reclutas. El análisis de reconstrucción de hermanos, sin embargo, permitió ver que los hermanos en general no se reclutan juntos en la misma ubicación, y que la distancia entre los lugares de reclutamiento fue más extensa (ca.11.5 km). Se argumenta que los análisis de hermandad, en combinación con el análisis de paternidad, proporcionan una imagen más completa del patrón de conectividad y se puede aplicar con mayor facilidad a especies para las cuales el muestreo masivo de los padres no es factible. Tales medidas directas de la dispersión son esenciales para la comprensión de los patrones de conectividad actuales en diferentes hábitats marinos y permiten evaluar el grado de auto-reposición y sostenibilidad de las poblaciones, incluso en especies con flujo génico alto.

Retention and fish larval dispersal potential on a highly connected open coast line

C. Schunter[‡†*], M. Pascual[†], J.C. Garza[§], N. Raventos[‡], E. Macpherson[‡]

[‡]Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Car. Acc. Cala St. Francesc 14, Blanes, 17300 Girona, Spain

[†]Dept. Genètica, Universitat Barcelona, Diagonal 643, 08028 Barcelona, Spain

[§]Southwest Fisheries Science Center, National Marine Fisheries Service and University of California, 110 Shaffer Rd., Santa Cruz 95060, USA.

[*]Corresponding author: Telephone: +34-93-4034850 Fax: +34-93-40434420.

e-*mail address*: celiaschunter@gmail.com

Keywords: Parentage analysis, connectivity, dispersal, recruitment, sibship, *Tripterygion delaisi*.

Short Title: Fish dispersal on a temperate coast line

Abstract

Connectivity is crucial for the persistence and resilience of marine species, the establishment of networks of Marine Protected Areas (MPAs) and the definition of fishery management units. In the marine environment, understanding connectivity still remains a major challenge due to the technical difficulties of tracking larvae. Recently, paternity analysis has provided a means to address this question more effectively. This method requires extensive sampling of parents, which is often not possible with commercial, protected or deep-water species. Here, we directly measure connectivity and larval dispersal for the first time in a temperate marine ecosystem, through genetic analysis with 178 SNP markers and subsequent parentage and sibship reconstruction analyses of the black faced blenny (*Tripterygion delaisi*) from an open coastline in the Mediterranean Sea (N=1573). Parentage analysis suggested limited connectivity, with a decrease in dispersal success over 1 km distance and approximately 15% of the juveniles identified as self-recruits. Sibship reconstruction analysis found that, in general, full siblings did not recruit together to the same location, and that the distance between recruitment locations was much larger than found for parent/offspring pairs (ca.11.5 km). We demonstrate that sibship analysis, in combination with parentage analysis, provides a more complete assessment of connectivity and can be more easily applied to species for which extensive parental sampling is not feasible. Such direct measurements of dispersal are essential to understanding connectivity patterns in different marine habitats and show the degree of self-replenishment and sustainability of populations of marine organisms.

Introduction

Larval dispersal determines the connectivity patterns of many species of marine fish. Connectivity counteracts population fragmentation and is crucial for the persistence and resilience of many species. This has been emphasized in conservation policies, including the design of networks of marine protected areas [1, 2] and in fisheries management [3], and increasing reserve connectivity is a critical step for biodiversity conservation and management [4]. However, due to the technical difficulties in tracking fish through the pelagic larval phase, direct measures of connectivity are scarce and understanding the distribution of dispersal distances and their direction is still a great challenge in marine ecology [5]. One solution to overcome this problem is the use of parentage analysis to study dispersal in species with a relatively stationary adult phase [6]. Such analysis permits the estimation of connectivity, as the detection of parent-offspring pairs can provide direct evidence of offspring dispersal, if adult movement patterns are known [7]. Since the first application of this method in the marine environment, a number of studies have been conducted that differ in scale as well as in study species, but some patterns seem to be general. First, a small proportion of larvae are self-recruits, meaning that they settle close to where they were born. For example, in the panda clownfish only 6% settled in their natal population [8], whereas 17% of coral trout were self-recruits [9]. However, as much as 43% self-recruitment was reported for the orange clownfish in an island environment [6]. The rest of the larvae 'emigrate', with dispersal success rapidly declining over short distances [10]. Even though parentage analyses have found larval export from a marine reserve in the Great Barrier Reef to reach locations ~ 30 km away [9], and in Papua New Guinea up to 35km in distance [6], this could still be considered small scale, especially for species with very discontinuous habitats.

Nevertheless, the successful application of parentage analysis to the study of larval dispersal requires sampling a sufficiently large proportion of the parental population for the successful encounter of parent-offspring pairs [11]. Due to this, studies applying parentage analysis have been restricted to species with confined parental habitats [6, 12] or that form reproductive aggregations [13]. Most commercial and endangered marine species, especially in temperate and cold water ecosystems, do not exhibit such characteristics. In these cases, other genetically related approaches, such as sibship reconstruction, might be useful to identify dispersal and connectivity patterns.

In this study, we performed parentage and sibship reconstruction analyses with the black faced blenny (*Tripterygion delaisi*), a small near-shore rocky reef fish [14], on an open coastline of the North-western Mediterranean Sea using 192 single nucleotide polymorphisms (SNPs). We use the results of these analyses to provide the first direct estimates of the distribution of larval dispersal distances and directionality for a fish species in an open coast temperate marine ecosystem.

Methods

(a) Study species, study area and sampling

The black faced blenny (*Tripterygion delaisi*) is distributed throughout the Mediterranean Sea and the North-eastern Atlantic Ocean [14] in near-shore rocky reef habitat, where they live camouflaged within the rocks or algae for most of the year. When the reproductive period starts in spring, some males, known as territorial or dominant males, undergo coloration changes that result in a black head and bright yellow coloring for the rest of the body, distinguishing them

from other members of the species (Figure 1). These males start protecting a small territory, which is referred to as their nest, against predators and other secondary males [15]. The black faced blenny also displays a high level of homing behavior [16], and adults are not known to disperse even small distances in open water and sandy bottom habitat. Consequently, dispersal, as well as gene flow, is confined to the pelagic larval stage.

The study area is on an open Mediterranean Sea coastline located in North-eastern Spain, near the port of Blanes (Figure 1). It is centered on a rocky shore coastline of ~ 2 km in length, surrounded by sandy bottom habitat that extends for many km in both directions, but especially towards the south-west, and acts as a dispersal barrier (Figure 1).

A total of 946 adult fish were sampled on SCUBA between April and July 2010. Exhaustive searching for all nests in a 2 km section of coastline (Blanes area) was performed and 796 territorial males were sampled from their nesting sites. Additionally, 150 camouflaged adults of unknown sex were sampled. Fish were caught with small nets, body length measured, and a small tissue sample taken from the dorsal fin.

A total of 627 recruits were collected on SCUBA between the months of July and September 2010. Recruits were sampled in the Blanes area, as well as at seven external locations, spanning a distance of 48.2 km, between Arenys and Tossa de Mar (Figure 1). Four of these locations are found to the north-east of Blanes, where after a large sandy bottom stretch, the rocky shore continues, providing habitat for *T. delaisi*. South-west of Blanes, recruits were sampled at all the available sites with rocky habitat.

(b) Genotyping and genetic analyses

All 1573 samples were genotyped using 192 SNP markers developed for *T. delaisi*. SNP genotyping was performed using 96.96 Dynamic SNP Genotyping Arrays on an EP1 genotyping system (Fluidigm Corporation) [17]. Deviations from Hardy-Weinberg (H-W) and linkage disequilibria were evaluated with Genepop 4.2 [18] and loci not in H-W equilibrium as well as the less variable locus of pairs in linkage disequilibrium were eliminated. Genotypes from the remaining 178 SNP loci were used for further analysis. Details of SNP development and of all genotyping assays are described in Schunter *et al.* 2013 [17].

Two relatedness categories were evaluated: parentage and sibship. Parent-offspring matches were established using the software CERVUS [19]. This program allows use of a large number of loci and has been shown to accurately identify parent-offspring pairs in empirical data [20]. The genotyping error rate was set to 1% and a range of different values, between 50% and 95%, for the parameter 'adult proportion sampled' were evaluated. A maximum of three mismatches were allowed and the LOD value cutoff was 2. COLONY [21] was also used to identify parent-offspring pairs, as well as to identify siblings in the juvenile samples. The predefined parameters were used, except that the mating system was set to polygamy. Full sibship was only accepted for values of the posterior probability > 0.75. All of these analyses were repeated with a permuted dataset to evaluate the false positive rate [22]. The program ML-RELATE [23] was also used to identify siblings. Euclidian distances between individual genotypes were established and a principal component analysis performed with these distances using the R package *adegenet* [24]. Geographic distances between individuals were estimated with the R package *fields* [25].

(c) Otolith analysis and wind variables

The body length of all recruits was measured, and the otoliths (lapilli and sagitta) were extracted and mounted on microscope slides. The age of individuals was determined in order to establish

the corresponding dates of hatch and settlement, as well as planktonic larval duration (PLD), following Raventós & Macpherson [26].

Wind characteristics have a strong influence on the inshore circulation pattern in the study area [27]. Considering that *Tripterygion* larvae are always distributed along inshore waters (< 2.5 km from shore) [28], we compared the larval dispersal direction and distance with wind speed and direction to evaluate a possible influence on dispersal patterns. Wind data were obtained from an automated meteorological station belonging to the XMET service (National Weather Service of Catalonia). Wind data were monitored hourly with an anemometer 10 m above the ground. The overall wind regime in the study area revealed a clear diurnal pattern with stronger and more variable winds with solar heating in the daytime and winds blowing from the coast out towards the open sea at night [26]. This pattern makes it necessary to divide the wind data into two different datasets: day (12:00 pm) and night (12:00am). The settlement date of each recruit (as derived from otolith readings) and the corresponding PLD value were used to determine the period over which the wind variables were to be averaged. Day- and night-time daily averages were calculated, and wind speed and direction combined into a single coarse wind variable.

Results

A total of 25 parent-offspring pairs were identified, all in the Blanes area. The same 25 pairs were found with both parentage methods. The proportion of sampled fathers assumed with CERVUS did not affect the results, with all of the same pairs identified regardless of the value used, and no parent-offspring pairs were identified in the permuted datasets. We found that 15% of the juveniles from the Blanes area were self-recruits (Figure 2), with one of these offspring

found directly adjacent to the fathers' nest. Self-recruits did not disperse further than 1.2 km, with 80% settling less than 1 km from their natal nest (Figure 3).

Sibship analysis with COLONY identified 21 pairs of juveniles as full siblings. ML-RELATE identified six of these 21 pairs as full siblings, whereas the other 15 pairs had very similar likelihoods of being full or half siblings. As *T. delaisi* adults do not disperse and males territorially guard one nest, even half siblings will come from either the same nest (paternal) or close nests (maternal). Either way, these 21 identified pairs are clearly siblings and can be assumed to have dispersed from approximately the same location. The geographic distances between siblings were larger than between parent-offspring pairs. The maximum distance between siblings was 11.5 km, but 24% of sibling pairs were separated by > 10 km distance and only 38% dispersed less than 1 km from each other (Figure 3). We found no population structure within our samples, indicating that fish from Blanes and other sampling nearby locations (e.g., Tossa de Mar) belong to the same genetic population (Figure 4).

Dispersal distance and direction did not correlate with the wind variable, which included both wind direction and strength and was calculated for both day- and night-time (Spearman's rho, day component vs. direction: rho = 0.179, p = 0.41; day component vs. distance: rho = 0.665, p = 0.22; night component vs. direction: rho = -0.400, p = 0.06; night component vs. distance: rho = -0.177, p = 0.42). Nevertheless, we found the tendency of dispersal direction to be south-westerly (Figure 2); more self-recruits from the Blanes area moved south-west across the midpoint of the area and more self-recruits that hatched on the south-western side recruited to the south-western part. No recruit collected north-east of Blanes was an offspring of any of the fathers in the Blanes area, which suggests that the larvae do not move north-east.

Discussion

The rate of self-recruitment, settlement of a larva in the location where it hatched, in marine fishes has long been debated [29]. The degree to which a population receives self-recruits has ecological implications, as it influences the level of self-replenishment and population sustainability [30]. Several studies using parentage analysis to study dispersal of coral reef fishes inhabiting embayments in Papua New Guinea revealed high self-recruitment in several species [6, 30], with up to 64% of juveniles recruiting to their natal site. Studies using a similar approach with coral reef fishes but on open coastlines found self-recruitment rates of 17% for coral trout [9] and 10% for clownfish [12]. The rate of self-recruitment for *Tripterygion delaisi* found here (15%) was similar, despite the temperate open coastline environment they inhabit. Variance in the level of self-recruitment could be related to species characteristics, but the rate of larval retention is most likely dependent upon the physical location and oceanographic conditions of the habitat.

Pelagic larval duration (PLD) has been suggested to have a large influence on the rate of self-recruitment [5, 29]. However, this does not seem to be the case in open coastline habitat, as similar proportions of self-recruits were detected for coral reef species with PLDs of almost a month [9] and just days [12], and in the temperate species *T. delaisi* studied here, which has an average PLD of 18 days. Nevertheless, the rate of self-recruitment found in *T. delaisi*, although in agreement with the two other studies in open coastline habitat, might be a slight underestimation, as sneaker males may contribute to the progeny. We collected all the territorial males in the study area and tried to capture the sneaker males, but were largely unsuccessful, as they are camouflaged and quite difficult to see. However, sneaker males only re-occupy 10% of

nest sites after the removal of a dominant male [32], suggesting a generally low reproductive output for this behavioral strategy.

Large-scale investigations that integrate population genetics and oceanography have shown that connectivity can be influenced, or even dominated, by physical processes [33–35]. However, we failed to find a significant correlation between the dispersal direction and geographic distance of self-recruits with a composite wind variable. Similarly, dispersal patterns of the coral reef fish *Amphyprion polymnus* were also found not to be influenced by physical processes, such as currents [8]. In general, near-shore oceanographic features are harder to model, because they are mostly influenced by wind and coastal morphology [27]. Despite the lack of correlation of dispersal and wind, we found the tendency of dispersal direction to be south-western. This pattern is in agreement with the sea-surface circulation in the area, which is dominated by a south-westward current [35, 36]. This tendency of larvae to disperse south-westward, where the coastline is formed by sandy beaches with no available rocky reef habitat for approximately 100 km, suggests that Blanes is a 'dead end' for the black faced blenny population of this coastline, with little successful dispersal occurring beyond this area. Our finding is concordant with a previous study of genetic population structure, which reported the population of Blanes to be genetically distinct from that of the next analyzed locality ~250 km in the south-western direction [38]. However, these populations are not completely disconnected, as isolation by distance was found along the Spanish coast and it appears that a small number of larval recruits must successfully disperse long distances to the south-west. Consequently, the majority of larvae from an open coastline habitat would settle in suitable habitats adjacent to the natal location, but small numbers of larvae allow large-scale connectivity.

Parentage analysis in *T. delaisi* revealed limited dispersal distance. Identified dispersal events declined precipitously beyond 1 km from the natal site, as previously found in reef fish species [39]. This would suggest limited connectivity, and consequently high genetic differentiation, between proximate locations such as Blanes and Tossa de Mar. However, contradictory to this prediction, the entire population of the study area was found to be one genetic unit connected by high gene flow, consistent with a previous study on *T. delaisi* in this area [38].

This high connectivity was, however, reflected in the sibling pairs identified. Sibship analysis has previously been used successfully to elucidate family structure and mating strategies in several other fish species [40–42]. Here, we use it to complement parentage analysis and provide additional information on the dispersal potential of the species. Whereas paternity analysis provides information on both dispersal distance and directionality, identification of siblings recruiting to different locations provides estimates of dispersal distance, because full siblings begin dispersal from the same natal nest. The mean distance between black faced blenny sibling pairs was much larger than the maximum distance between parent-offspring pairs (Figure 3). Nonetheless, sibship analysis could still underestimate dispersal distance, since the natal location is not known. If siblings disperse in opposite directions, then the distance between them will be twice their mean dispersal distance, but if they do not, then the observed distance may be much less than twice their dispersal distance. Indeed, a recent study on a coral reef fish provided evidence that siblings can travel together throughout the entire planktonic larval phase [43]. There was one black faced blenny sibling pair encountered at the same rock. Otolith analysis found that both siblings hatched within two days of each other and spent 19 days in the planktonic larval phase, indicating that they most likely left the nest at the same time and recruited together. In contrast, many groups of recruits were collected simultaneously from the

same locations, but only this one pair of siblings was identified. This indicates that the majority of full siblings do not recruit to the same location, suggesting a high degree of larval mixing and recruitment heterogeneity [43].

The direct measurement of dispersal distance and direction by applying parentage analysis is beginning to dramatically increase our knowledge of larval dispersal in marine organisms. In *T. delaisi*, parent-offspring comparisons indicate that larvae tend to move short distances south-west along the coast. Sibling pairs, however, recruit to locations that can be quite distant from each other, reflecting the high degree of connectivity inferred by the population structure. We show how sibship analysis complements parentage methods and can provide direct insight into both family structure and dispersal potential. It can also be more easily applied to species for which massive collection of the parental population is difficult or unfeasible. Here, the combination of both parentage and sibship analyses provide the first direct measurements of dispersal for a fish in a temperate marine habitat, revealing the rate of self-recruitment and explaining the apparent high gene flow.

Collection and field procedures followed the Spanish Laws (Royal Executive Order, 53/2013) for Animal Experimentation, in accordance with the European Union directive (2010/63/UE).

We are grateful to the Southwest Fisheries Science Center Molecular Ecology and Genetic Analysis Team, especially V. Apkenas, C. Columbus, E. Gilbert-Horvath, E. Anderson, D. Pearse and A. Abadía-Cardoso for support in the lab and helpful comments. We also thank E.

Ballesteros for providing a photo of the black faced blenny. This work was partially funded by the Spanish Ministry of Science and Innovation through the BENTHOMICS (CTM2010-22218-C02-01) project and the FBBVA project (BIOCON 08 – 187／09). The authors CS, MP and EM are part of the research group 2009SGR-636 and 2009SGR-665 of the Generalitat de Catalunya and CS was funded by a JAE-Predoctoral Fellowship.

References

1 Botsford, L. W., White, J. W., Coffroth, M.-A., Paris, C. B., Planes, S., Shearer, T. L., Thorrold, S. R. & Jones, G. P. 2009 Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. *Coral Reefs* **28**, 327–337. (doi:10.1007/s00338-009-0466-z)

2 Palumbi, S. R. 2003 Population genetics, demographic connectivity, and the design of marine reserves. *Ecol. Appl.* **13**, 146–158. (doi:10.1890/1051-0761(2003)013[0146:PGDCAT]2.0.CO;2)

3 Fogarty, M. J. & Botsford, L. W. 2007 Population connectivity and spatial management of marine fisheries. *Oceanography* **20**, 112–123. (doi:10.5670/oceanog.2007.34)

4 Heller, N. E. & Zavaleta, E. S. 2009 Biodiversity management in the face of climate change: a review of 22 years of recommendations. *Biol. Conserv.* **142**, 14–32. (doi:10.1016/j.biocon.2008.10.006)

5 Selkoe, K. & Toonen, R. J. 2011 Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Mar. Ecol. Prog. Ser.* **436**, 291–305. (doi:10.3354/meps09238)

6 Planes, S., Jones, G. P. & Thorrold, S. R. 2009 Larval dispersal connects fish populations in a network of marine protected areas. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5693–7. (doi:10.1073/pnas.0808007106)

7 Hedgecock, D., Barber, P. & Edmands, S. 2007 Genetic approaches to measuring connectivity. *Oceanography* **20**, 70–79. (doi:10.5670/oceanog.2007.30)

8 Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R. & Planes, S. 2011 Connectivity dominates larval replenishment in a coastal reef fish metapopulation. *Proc. Biol. Sci.* **278**, 2954–61. (doi:10.1098/rspb.2010.2780)

9 Harrison, H. B. et al. 2012 Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Curr. Biol.* **22**, 1023–8. (doi:10.1016/j.cub.2012.04.008)

10 Buston, P. M., Jones, G. P., Planes, S. & Thorrold, S. R. 2012 Probability of successful larval dispersal declines fivefold over 1 km in a coral reef fish. *Proc. Biol. Sci.* **279**, 1883–8. (doi:10.1098/rspb.2011.2041)

11 Jones, A. G. & Ardren, W. R. 2003 Methods of parentage analysis in natural populations. *Mol. Ecol.* **12**, 2511–23. (doi:10.1046/j.1365-294X.2003.01928.x)

12 Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R. & Planes, S. 2012 Patterns and persistence of larval retention and connectivity in a marine fish metapopulation. *Mol. Ecol.* **21**, 4695–705. (doi:10.1111/j.1365-294X.2012.05726.x)

13 Almany, G. R. et al. 2013 Dispersal of grouper larvae drives local resource sharing in a coral reef fishery. *Curr. Biol.* **23**, 626–630. (doi:10.1016/j.cub.2013.03.006)

14 Carreras-Carbonell, J., Macpherson, E. & Pascual, M. 2005 Rapid radiation and cryptic speciation in mediterranean triplefin blennies (Pisces: Tripterygiidae) combining multiple genes. *Mol. Phylogenet. Evol.* **37**, 751–61. (doi:10.1016/j.ympev.2005.04.021)

15 Wirtz, P. 1978 The behaviour of the Mediterranean Tripterygion species (Pisces, Blennioidei). *Z. Tierpsychol* **48**, 142–174. (doi:10.1111/j.1439-0310.1978.tb00253.x)

16 Heymer, A. 1977 Experiences subaquatiques sur les performances d'orientation et de retour au gite chez *Tripterygion tripteronotus* et *Tripterygion xanthosoma* (Blennioidei, Tripterygiidae). *Vie Milieu* **27**, 425–435.

17 Schunter, C., Garza, J. C., Macpherson, E. & Pascual, M. 2013 SNP development from RNA-seq data in a non-model fish: how many individuals are needed for accurate allele frequency prediction? *Mol. Ecol. Resour.* **in press**. (doi:10.1111/1755-0998.12155)

18 Rousset, F. 2008 genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–6. (doi:10.1111/j.1471-8286.2007.01931.x)

19 Kalinowski, S. T., Taper, M. L. & Marshall, T. C. 2007 Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **16**, 1099–106. (doi:10.1111/j.1365-294X.2007.03089.x)

20 Slate, J., Marshall, T. & Pemberton, J. 2000 A retrospective assessment of the accuracy of the paternity inference program CERVUS. *Mol. Ecol.* **9**, 801–808. (doi:10.1046/j.1365-294x.2000.00930.x)

21 Jones, O. R. & Wang, J. 2010 COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour.* **10**, 551–5. (doi:10.1111/j.1755-0998.2009.02787.x)

22 Anderson, E. C. 2012 sgm_perm -- Permute a simple genetic matrix. *http://users.soe.ucsc.edu/~eriq/dokuwiki/doku.php?id=sisg:sibship_inference_practical2012*.

23 Kalinowski, S. T., Wagner, A. P. & Taper, M. L. 2006 ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Mol. Ecol. Notes* **6**, 576–579. (doi:10.1111/j.1471-8286.2006.01256.x)

24 Jombart, T. 2008 adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–5. (doi:10.1093/bioinformatics/btn129)

25 Fields Development Team 2006 fields: Tools for Spatial Data. *http://www.cgd.ucar.edu/Software/Fields.*

26 Raventós, N. & Macpherson, E. 2001 Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Mar. Biol.* **138**, 1115–1120. (doi:10.1007/s002270000535)

27 Grifoll, M., Aretxabaleta, A. L., Espino, M. & Warner, J. C. 2012 Along-shelf current variability on the Catalan inner-shelf (NW Mediterranean). *J. Geophys. Res.* **117**, C09027. (doi:10.1029/2012JC008182)

28 Sabates, A., Zabala, M. & García-Rubies, A. 2003 Larval fish communities in the Medes Islands Marine Reserve (North-west Mediterranean). *J. Plankton Res.* **25**, 1035–1046. (doi:10.1093/plankt/25.9.1035)

29 Jones, G. P., Milicich, M. J., Emslie, M. J. & Lunow, C. 1999 Self-recruitment in a coral reef fish population. *Nature* **402**, 802–804. (doi:10.1038/45538)

30 Sponaugle, S. et al. 2002 Predicting self-recruitment in marine populations: biophysical correlates and mechanisms. *Bull. Mar. Sci.* **70**, 341–375.

31 Berumen, M. L., Almany, G. R., Planes, S., Jones, G. P., Saenz-Agudelo, P. & Thorrold, S. R. 2012 Persistence of self-recruitment and patterns of larval connectivity in a marine protected area network. *Ecology and Evolution* **2**, 444–52. (doi:10.1002/ece3.208)

32 Jonge, J. & Videler, J. J. 1989 Differences between the reproductive biologies of *Tripterygion tripteronotus* and *T. delaisi* (Pisces, Perciformes, Tripterygiidae): the adaptive significance of an alternative mating strategy and a red instead of a yellow nuptial colour. *Mar. Biol.* **100**, 431–437. (doi:10.1007/BF00394818)

33 White, C., Selkoe, K., Watson, J., Siegel, D. A., Zacherl, D. C. & Toonen, R. J. 2010 Ocean currents help explain population genetic structure. *Proc. Biol. Sci.* **277**, 1685–1694. (doi:10.1098/rspb.2009.2214)

34 Schunter, C., Carreras-Carbonell, J., Macpherson, E., Tintoré, J., Vidal-Vijande, E., Pascual, A., Guidetti, P. & Pascual, M. 2011 Matching genetics with oceanography: directional gene flow in a Mediterranean fish species. *Mol. Ecol.* **20**, 5167–81. (doi:10.1111/j.1365-294X.2011.05355.x)

35 Selkoe, K. A., Gaines, S. D., Caselle, J. E. & Warner, R. R. 2006 Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology* **87**, 3082–94. (doi:10.1890/0012-9658(2006)87[3082:CSAKAE]2.0.CO;2)

36 Font, J., Garcialadona, E. & Gorriz, E. 1995 The seasonality of mesoscale motion in the northern current of the western mediterranean - several years of evidence. *Oceanologica Acta* **18**, 207–219.

37 André, G., Garreau, P., Garnier, V. & Fraunié, P. 2005 Modelled variability of the sea surface circulation in the North-western Mediterranean Sea and in the Gulf of Lions. *Ocean Dynamics* **55**, 294–308. (doi:10.1007/s10236-005-0013-6)

38 Carreras-Carbonell, J., Macpherson, E. & Pascual, M. 2006 Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. *Mol. Ecol.* **15**, 3527–39. (doi:10.1111/j.1365-294X.2006.03003.x)

39 Buston, P. M., Jones, G. P., Planes, S. & Thorrold, S. R. 2012 Probability of successful larval dispersal declines fivefold over 1 km in a coral reef fish. *Proc. Biol. Sci.* **279**, 1883–8. (doi:10.1098/rspb.2011.2041)

40 De Mestral, L. G., Herbinger, C. M., O'Reilly, P. T. & Taylor, E. B. 2012 Mating structure of an endangered population of wild Atlantic salmon (*Salmo salar*) as determined using sibship reconstruction and a novel method of sex inference. *Canadian Journal of Fisheries and Aquatic Sciences* **69**, 1352–1361. (doi:10.1139/f2012-065)

41 Dibattista, J. D., Feldheim, K. A., Garant, D., Gruber, S. H. & Hendry, A. P. 2009 Evolutionary potential of a large marine vertebrate: quantitative genetic parameters in a wild population. *Evolution* **63**, 1051–67. (doi:10.1111/j.1558-5646.2008.00605.x)

42 Kanno, Y., Vokoun, J. C. & Letcher, B. H. 2010 Sibship reconstruction for inferring mating systems, dispersal and effective population size in headwater brook trout (*Salvelinus fontinalis*) populations. *Conserv. Genet.* **12**, 619–628. (doi:10.1007/s10592-010-0166-9)

43 Bernardi, G., Beldade, R., Holbrook, S. J. & Schmitt, R. J. 2012 Full-sibs in cohorts of newly settled coral reef fishes. *PloS One* **7**, e44953. (doi:10.1371/journal.pone.0044953)

Figures:

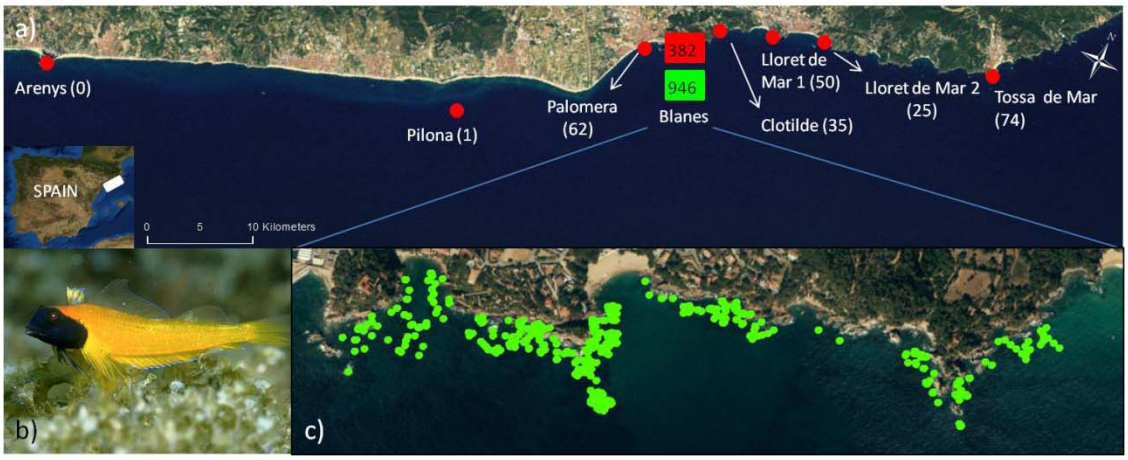

Fig. 1: a) Map of the sampling area with location names and number of recruits sampled in brackets. Note that from la Palomera towards the south-west there are no rocky reefs apart from la Pilona and Arenys. b) Dominant male of *Tripterygion delaisi* c) The Blanes intensive sampling area, where all dominant males encountered were sampled. Nest sites indicated with green dots.

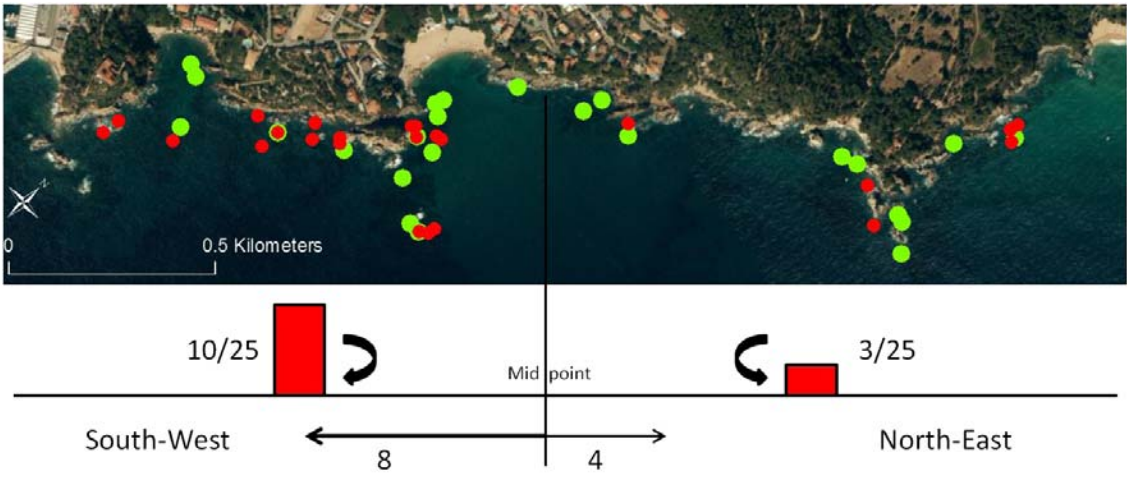

Fig. 2: Map of the Blanes intensive sampling area with parent-offspring pairs indicated. The nesting sites are marked in green and the locations where the self-recruits were caught are marked in red. Red bars beneath the map represent the self-recruits which stayed either on the south-west or the north-east side of the area midpoint. Arrow indicates the directional movement of recruits across the midpoint.

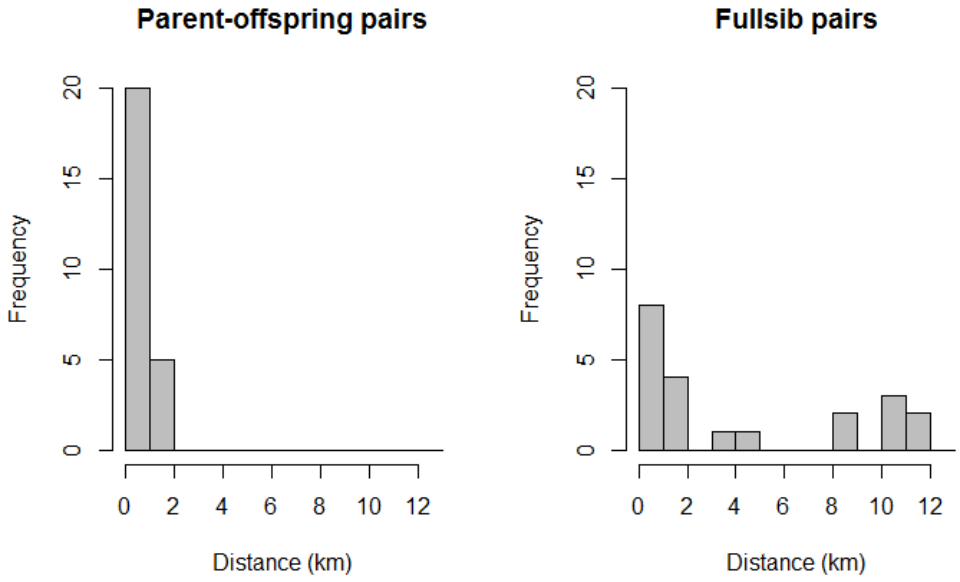

Fig. 3: Geographic distance of dispersal between parent-offspring and full sibling pairs.

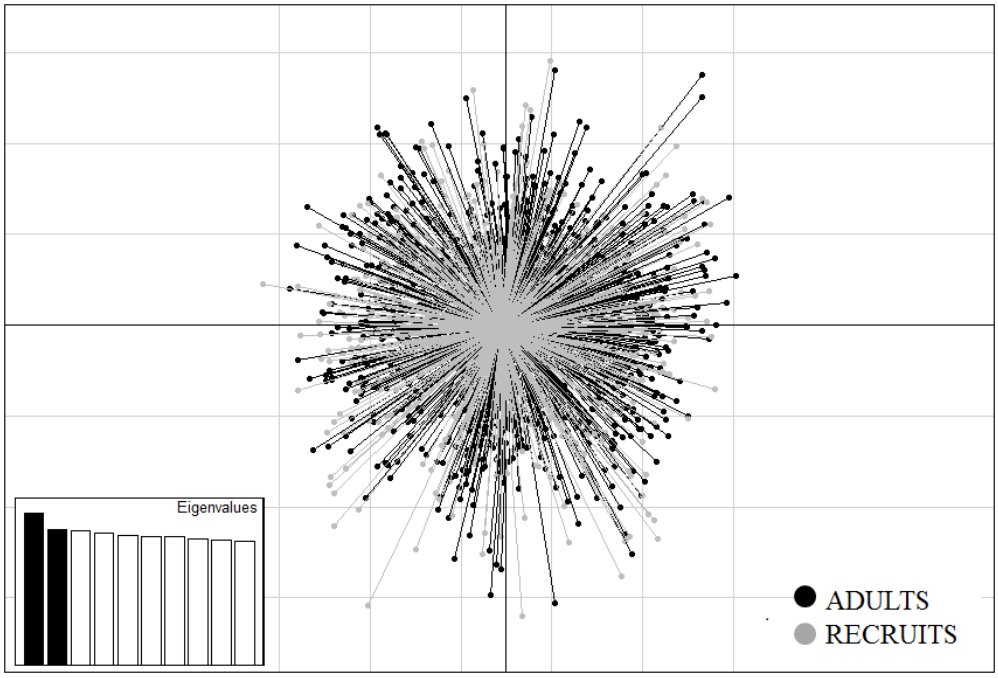

Fig.4: Principal Component analysis of recruit genotypes (grey) and adult genotypes (black).

4. DISCUSSION

DISCUSSION

Marine ecosystems are complex, highly diverse and productive, providing a great variety of resources. Even though humans have been exploiting these resources for centuries, only recently have we begun to understand the extent of exploitation and deterioration in the marine environment, of its fauna and flora, and therefore recognized the need for intensive study. Due to the fact that the marine environment is not readily accessible, studying even basic biological and ecological issues, e.g. abundance, are difficult to accomplish and require technical equipment. Therefore, nowadays most marine species are probably still unknown and answering ecological and evolutionary questions for the marine habitat remains still a major challenge. The application of molecular tools has provided a means to address some of the challenges in studying ecological aspects.

This new study field is referred to as Molecular Ecology and uses genomic material to answer ecological and evolutionary topics also in the marine environment. In this thesis I applied various molecular methodologies to address a broad scope of challenges in marine ecology. With the help of genetic tools it was possible to evaluate a variety of evolutionary processes and answer ecological questions for littoral fishes of the Mediterranean Sea.

Genetic population structure and oceanographic fronts in littoral fish species

One large challenge in marine ecology, which has still not been fully answered, is the understanding of species dispersal capabilities. Dispersal ability is one of the major determinants of the dynamics of local populations and the geographic range of species (Bowler & Benton 2005). *Epinephelus marginatus* and *Serranus cabrilla* are two species of Mediterranean fish with potentially different dispersal capacities. The first species has a large scale distribution, ranging from the Mediterranean Sea all the way across the Atlantic Ocean to the western Atlantic coast of Brazil and Argentina. As this species shows high site fidelity as adults, this distribution patterns would suggest a dispersion of pelagic larvae over very large distances (Heemstra & Randall 1993; Hereu *et al.* 2006). The second species, the comber, although found in the Mediterranean and on the African coast all the way to Southern Africa does not seem to be able to cross the Atlantic Ocean (Guidetti & Cattaneo-Vietti 2002). A large-scale geographic

distribution could suggest that the species has a large potential to disperse and even distant localities are connected with each other. In the case of the comber and the dusky grouper, both revealed complex connectivity patterns with no genetic distinction of fairly distant populations and significant genetic differentiation between neighboring localities. For instance, both species had very low F_{ST} values for pairwise comparisons between Greece and the Balearic Islands, which are located at approximately 1800km distance from each other. Other marine species, such as the blue hake (*Antimora rostrata,* White *et al.* 2011) or the coral reef fish (*Naso hexacanthus,* Horne & Van Herwerden 2013) have shown panmixia over a large scale of several thousand kilometers. For the dusky grouper genetically differentiated locations mainly included an Atlantic locality and thus were mostly geographically distant, revealing that this species might not be panmictic throughout its range. Mediterranean Sea localities show panmixia but excluding Algeria. Such patterns are what have been expected for marine species in the past, due to the apparent continuity of marine waters. In such cases we would expect a slight genetic isolation by distance (IBD) following a stepping stone model of increasing genetic distance with geographic distance (Wright 1969). However, it has become more and more clear that a large variety of marine species do not display such isolation by distance (Bradbury & Bentzen 2007). *Epinephelus marginatus* as well as *Serranus cabrilla* showed a complex pattern, whereas the outcome depended in both cases on the inclusion or exclusion of certain localities. For the dusky grouper, if all studied localities were included in the analysis, then IBD was found due to the distant and genetically distinct samples from the Atlantic Ocean, whereas when excluding these samples there was no significant IBD. Reversely, the comber had no IBD pattern when all samples were included, but a significant pattern was encountered when only analyzing the Spanish coast. This shows that the detection of IBD depends largely on the sample locations included in the study. Also, it shows that the extent of dispersal and gene flow is more complex than previously expected as genetic differentiation as well as gene flow patterns do not solely depend on the distance.

Another factor that has been suggested as a proxy for the dispersal potential of a species is the pelagic larval duration (PLD) (e.g. Sponaugle *et al.* 2002; Selkoe & Toonen 2011). In the review of Selkoe & Toonen (2011) a moderate fit between data on genetic population structure and pelagic larval duration was displayed implying that PLD reflect dispersal potential (but see Galarza *et al.* 2009) . For the dusky grouper and the comber the PLD is around 21-30 days

(Raventós & Macpherson 2001), which would mean that by only considering the PLD they could exhibit the same dispersal potential and in turn a similar genetic structuring. However, these species have a different distribution span and we did not find the same pattern of population differentiation. The comber, on the contrary to the dusky grouper, showed clear genetic structuring for geographically close localities. Shanks (2009) reviewed data on PLD and dispersal potential and found that they are correlated. However, for species with longer PLDs dispersal distances are overestimated with larvae staying closer to the coastline where currents are slower. The author also suggested that life history traits and larval behavior might play a vital role in estimating dispersal capabilities.

Geographical distribution nor isolation by distance nor the PLD seem to be good estimators of dispersal potential and genetic population structure for the dusky grouper and the comber. If none of these parameters work, then what defines the population structure in marine species? Although it is most likely a combination of a variety of factors, there is one component for which there is growing evidence of its importance on the population structure of marine organisms: oceanographic processes and barriers (Galarza *et al.* 2009).

Within the Mediterranean Sea the mean circulation patterns and oceanographic patterns are well described (Millot 1999) and a variety of oceanographic fronts in the western Mediterranean have been documented. After passing through the Gibraltar strait the Atlantic water encounters the higher density Mediterranean water which generates different oceanographic discontinuities, for instance the Almeria-Oran Front (AOF), the Balearic Front (BF) or the Ibiza channel (IC) which can be breaks to gene flow (Patarnello *et al.* 2007; Galarza *et al.* 2009; García-Merchán *et al.* 2012). This allows for the evaluation of the effect of such oceanographic features on the genetic population structure. In the case of the dusky grouper the samples from Algeria were genetically differentiated from the rest of the Mediterranean samples, which could be an effect of a mesoscale phenomenon of coastal eddies slowing down the eastward moving current which are connected to open-sea eddies in the middle basin towards the Balearic Islands (Millot 2005). This could explain why the Algerian population was genetically differentiated due to retention of larvae by the coastal eddies, but also to some extent connected by off-shore eddies to the Balearic Island of Menorca. However, it has to be considered that with a denser sampling in the Alboran Sea it might be possible to discover an effect of the AOF on the dusky grouper. A much

clearer influence of oceanographic processes was shown for *Serranus cabrilla*. With the help of oceanographic particle simulations a significant correlation with genetic structuring and larval migration was detected. All three barriers (AOF, IC and BF) have an effect on the dispersal of comber, which revealed to be mostly mediated by the pelagic larvae. The importance of oceanographic features such as currents and barriers on dispersal in the marine environment is now being clearly recognized (Selkoe *et al.* 2006; White *et al.* 2010). Galarza and coauthors (2009) evaluated these particular barriers for seven different fish species and obtained varying results depending on the species. Different outcomes were found for closely related species with similar life history traits. One reason for such discordance, which has rarely been considered, could be the seasonal changes in current flow or temporal variability in oceanographic processes (Stenseth *et al.* 2006). For the comber we found that the Ibiza Channel hinders larvae to move south across the barrier and we therefore find limited connectivity between population on either side. The dusky grouper however is not affected by this barrier, which is most likely due to the seasonality of this barrier coinciding only with the reproductive season of *Serranus cabrilla* between April and June but not *Epinephelus marginatus*, which reproduces late summer to fall.

The observation that *Epinephelus marginatus* populations are connected on a large-scale, is a positive revelation for the species' protection as this means there is no particular isolated gene pool going extinct. Nonetheless, these data can reflect pattern from the past and the rapid declines in numbers might lead to genetic dis-connectivity. Very recently it has been shown that the current MPAs in the Mediterranean Sea are far from being a true and well connected network and currently 20% of the Mediterranean continental shelf are left without dusky grouper larval supply (Andrello *et al.* 2013). Therefore, we propose three different management or conservation units: Senegal (Atlantic), Algeria (NW African Mediterranean coast) and the rest of the Mediterranean locations. It is fundamental to identify direct connectivity between different management units and crucial in marine reserve designs to ensure the long-term survival of this endangered species. The population structure and gene-flow patterns of *Serranus cabrilla* were influenced by predominant water current patterns and oceanographic processes. Four units were identified, separated by three oceanographic barriers to gene flow: the IC, the AOF and the BF. This shows the necessity to include oceanographic data into population genetic studies to understand the division of the connectivity patterns and also to study directional dispersal patterns.

Comparative transcriptomics

New technologies now provide the means to study and analyze genomes as well as transcriptomes on a genome-wide basis. This, first of all, will improve cost and time-efficient large-scale development of genomic resources such as single nucleotide polymorphism markers also for non-model species. However, many more basic and complex analyses can be undertaken when comparing transcriptomes and such analyses will expand the genomic knowledge of the dynamic properties of organisms at cellular and/or organism levels. It is now clear that the genomic code can vary greatly throughout an organism's body and through time. Understanding how, when and why this variability can be produced is therefore essential to unraveling key questions in evolution as well as ecology. Transcriptomic investigations allow for the better understanding of the relationship between an organism's genome and its phenotype, which in turn is related to the environment. Hence, the study of the differential gene 'use' can give indications on the adaptability of an organism and the capacity of a population or a species to adapt to changing environments.

The transcriptome of an individual varies through time and with external cues, environmental or social. Although the genomic material is the same, different genes are expressed in different tissue types and for different phenotypes for example. This differential expression can shed light on the biological functions required for a certain tissue type or phenotype. Previously it was not possible to look at a genome wide transcriptome for species without the presence of a reference genome, but due to RNA-seq and high throughput sequencers this is now achievable. Wang and coauthors (2009) stated that this method is a revolutionary approach to transcriptomics and due to the advances in technology such next generation sequencing approaches are now also feasible for non-model organisms. *De novo* assembling transcriptomic data is still a major challenge and not a trivial process (Martin & Wang 2011). Our *de novo* assembly of the brain tissue resulted in over 300.000 contigs including alternative spliced isoforms. A great variety of methods and bioinformatic tools have been developed to optimize the assemblies also in the absence of a reference genome, however, the ultimate goal of assembling complete gene transcripts is still not being completely achieved. Assemblers such as Trinity (Grabherr *et al.* 2011) can reconstruct a large proportion of transcripts, but a large quantity still remain highly fragmented. Despite the fact that such fragmentation might lead to a bias in overexpression of larger transcripts,

expression levels with RNAseq seem to be consistent and robust as it has been shown for mammal species (Oshlack & Wakefield 2009). Also, when comparing with microarrays, one Illumina lane represented an equal amount of data than a whole array, enabling identification of differentially expressed genes, while allowing for additional analyses such as detection of low-expressed genes, alternative splice variants, and novel transcripts (Marioni *et al.* 2008). The discovery of new sequences is one of the advantages of RNAseq, however if the transcripts are not in the annotation databases no function will be associated with the genes. Especially for non-model species this is reflected by the low percentage of transcripts with annotations. Such small percentages of successful detection of homologous genes is a common limitation of studying a non-model species at this moment, we found only 30% of brain transcripts to be annotated for the black faced blenny and only 37% had a functional description for wasps (Ferreira *et al.* 2013).

Despite the challenges and limitations, valuable evolutionary results can be drawn from the large quantities of data produced by this methodology and allowed me to analyze genome-wide expression differences between different phenotypes in the context of social dominance.

Alternative mating strategies

We determined the molecular differences among the three reproductive phenotypes of *Tripterygion delaisi*; the territorial male, the sneaker male and the female and found that the comparison between the territorial male and the sneaker male resulted in the highest number of differentially expressed transcripts. It might be surprising that there were more differentially expressed genes between male phenotypes, given the vast amount of studies on many species indicating genetic differences between sexes especially in the brain (Carruth *et al.* 2002; Arnold *et al.* 2004). Regardless, this shows that during the period of reproduction phenotypic plasticity is more important than sexual dimorphism. Phenotypic plasticity, the ability of a genotype to adapt to external conditions by changing its phenotype, has received considerable attention in evolutionary ecology, whereas focus has been laid on adaptive and selective phenotypic changes (Pigliucci 2005; Oliveira 2012). For the black faced blenny we demonstrate phenotypic plasticity on a temporal and reversible scale, as the reproductive period only lasts three months and all males display the same phenotype for the rest of the year.

Interestingly, the territorial male differentially over-expressed more genes than the other two phenotypes, possibly indicating that the dominant male has to go through more changes in these three months of reproductive period. The genes that were differentially expressed in dominant male overall were mostly related to synaptic plasticity. These genes are needed for the ability of a synapse or a neuron to change its conditions and parameters (Zucker & Regehr 2002). In rodents, for example, one of these genes found for *T.delaisi* was shown to be involved in signal transduction that is correlated with behavior (Berger & Roberts 2005). The sneaker overexpressed genes associated with differentiation and development such as for example *Col1a2*, which encodes for Collagen Type I alpha. This gene is a well-studied gene in the context of bone development (Karsenty & Wagner 2002). This result suggests that although this type of male is reproductively active, it might not be fully developed.

Previous studies looking at gene expression in different male phenotypes proposed candidate genes for social dominance which have been more intensively studied (Burmeister *et al.* 2005; Renn *et al.* 2008; Aubin-Horth & Renn 2009; Sneddon *et al.* 2011). These investigations were mostly done using single genes or microarrays. Only one of the seven previously proposed genes, the Somatostatin receptor type 1 (*sstr1*), showed significant differences in expression for *Tripterygion delaisi* males after correction for individual variation. Somatostatin is a neuropeptide also known as a growth hormone-inhibiting hormone and therefore commonly studied in the context of growth. In the African cichlid (*Astatotilapia burtoni*) somatostatin and somatostatin receptors have been shown to play a role in social behavior (Hofmann & Fernald 2000; Trainor & Hofmann 2006, 2007). Another candidate gene was differentially expressed in *Tripterygion delaisi,* the brain aromatase enzyme (*cyp19a1*), for which multiple isoforms were differentially expressed but only before adjustment for individual variation but not after. This shows that individual gene expression variation is highly important as has been indicated also for phenotypic plasticity (Renn *et al.* 2008) and outlier expression might bias the outcome (Barshis *et al.* 2013). This emphasizes the need for biological replicates even in genome wide studies.

The fact that none of the other proposed candidate genes were found to be differentially expressed in *T.delaisi* might be partly a result of the whole brain measurement of gene expression which might mask actual expression differences between different brain regions. However, it is very likely that the different species express different genes as they do not actually

display the same reproductive social system. In African cichlids, the subordinate males have undeveloped testes and need to become territorial to reproduce (Burmeister *et al.* 2005), which is distinct to the sneaker male in *T. delaisi* which has proportionally greater testes and is reproductively active (Jonge *et al.* 1989). In Atlantic Salmon precocious males and adult males show differential expression in some of the candidate genes but the two male types reproduce at different ages and the adult males do not settle down and defend a nest (Aubin-Horth *et al.* 2005). This suggests that there might be no general candidate genes but rather species-specific or context-specific candidate genes for social reproductive behavior.

One proximate consequence of the extreme divergences in phenotypes is the gain in fitness due to an increase in reproductive output. Analysis of species with alternative mating strategies showed differences in fitness for secondary males to be species specific: in European bitterling (*Rhodeus cericeus*) ,sneaker males successfully fertilized 30% of eggs within a nest (Reichard *et al.* 2004), whereas sunfish (*Lepomis gibbosus*) fertilize 15% (Rios-Cardenas & Webster 2008), male rose bitterling (*Rhodeus ocellatus*) 10% (Kanoh 2000) and sand goby sneakers (*Pomatoschistus minutus*) are parents to 5-10% of offspring (Malavasi *et al.* 2001). It would be highly interesting to directly measure the reproductive success of the sneaker male of *Tripterygion delaisi* and measure multipaternity in the nests. Furthermore, according to our findings in brain gene expression, the sneaker male invests in development. As the sneaker male of *T. delaisi* has proportionally greater testes in comparison to the dominant male according to our observations and a previous study (Jonge *et al.* 1989), an additional expression analysis of testis tissue would allow testing if this investment in development is of reproductive nature.

Fine scale genetic connectivity by parentage analysis using SNP markers

SNP development in non-model species

For the correct estimation of direct dispersal patterns via parentage analysis, a large amount of adult and juvenile samples need to be analyzed with a sufficiently large number of genetic markers. Unfortunately almost all marine species can be considered ‘non-model’ species and therefore there are often no genetic markers available and if there are, then the number is limited. Furthermore, the large number of samples to be genotyped requires a method that is ‘high-

throughput' for feasibility. For *Tripterygion delaisi*, the black-faced blenny, we developed 192 single nucleotide markers (SNPs) via transcriptome shotgun sequencing (RNAseq). One of the advantages of using SNPs as a genetic markers is the large density across the whole genome (Brumfield *et al.* 2003) and finding a large amount of SNPs should be relatively easy. However, when applying a pipeline with filtering criteria the number of SNPs reduces very rapidly. When comparing with the scarce published data on the development of SNPs for marine species by using high-throughput sequencing, it seems that the quantity of potential variable sites depended on the sequencing depth of the RNA-seq experiment (e.g. Renaut *et al.* 2010; De Wit & Palumbi 2012). Therefore if there is a need for thousands of SNPs, it is best to deep sequence or perform exome sequencing, such as demonstrated on the Atlantic herring (Lamichhaney *et al.* 2012).

For parentage analysis, which was the aim of the development of SNP markers for *T. delaisi*, the minor allele frequency is more important than a large quantity (Anderson & Garza 2006). A genetic marker is most powerful and useful to determine the relatedness of individuals when the minor allele frequency is high and loci are most informative if the alleles are equifrequent (Anderson & Garza 2006; Baruch & Weller 2008). As a major determinant of the power, SNPs were selected ultimately by this criterion. Such pre-selection of more informative SNP loci which can be critical in terms of funds and time for large genotyping projects, is one of the great advantages of using RNAseq (Garvin *et al.* 2010). Also, the possibility to select certain putative SNP loci helps to identify variant sites which are not true but are artifacts of duplicated genes (Hohenlohe *et al.* 2011). Such loci were eliminated from out dataset, as for a large set of loci we had uncovered an extreme excess of heterozygotes.

One issue when using RNAseq are the intron/exon boundaries (IEB), as only exons are sequenced with no indication of where the introns start. This could lead to amplification failure; however, we found that less than 10% of the putative loci failed due to IEBs in close proximity to the SNP site. One method to avoid such amplification failure is to evaluate the variant with PCR primers (Storer *et al.* 2012), whereas this is costly and time-consuming. We propose to identify homologous sites in genomes of closely related species and evaluate the distance of the putative SNPs site to the IEB. With this technique the amplification failure due to IEB was greatly reduced.

The production of cDNA libraries for RNAseq is time consuming and expensive and it would be optimal to reduce the number of libraries/individual samples to a minimum. By using bootstrap resampling we estimated the optimal sample number for the most efficient SNP development from RNAseq data. To avoid genotyping with monomorphic loci, as seen in the SNP development process for the ringed seal (*Pusa hispida hispida*, Olsen *et al.* 2011), it is best to use seven separately sequenced individuals for the SNP discovery. We also discovered that about twelve separately sequenced individuals represent the population allele frequency similarly well as fifteen individuals do, leaving twelve individuals as the optimal sample size.

Although SNP development for a non-model species can be challenging, if there is sufficient sequencing depth and our proposed filters are applied, valuable and powerful loci can be easily obtained. Despite the many advantages of SNP markers over other types of molecular markers such as microsatellites, one in particular is very important for the ever increasing sample size in ecological studies. SNP genotyping can be high throughput and therefore facilitate large-scale studies with great amounts of samples and markers (Everett *et al.* 2011). Throughout my thesis I have worked with a variety of genetic markers, such as microsatellite loci, mitochondrial genes and SNPs. The development and optimization of microsatellite markers is very time-consuming and the genotyping of this type of marker is often associated with observer-bias (Hauser *et al.* 2011). Notwithstanding, such markers are highly useful and due to their high polymorphism less loci are needed. However, for the evaluation of dispersal patterns via parentage analysis for *T. delaisi*, close to 2000 samples had to be genotyped, clearly requiring a time and resource efficient genotyping method, which was achieved with 192 SNP markers.

Dispersal potential on an open coast line

The black-faced blenny revealed to have no population structuring within the collected samples on the Costa Brava, showing that parents as well as offspring from Blanes and other sampling locations such as Tossa belong to the same genetic population. This supports the finding by a previous study revealing no genetic differences throughout the Costa Brava (Carreras-Carbonell *et al.* 2006). However, migrants from other distinct populations had been detected by assignment tests when assessing the origin of recruits collected in Blanes (Carreras-Carbonell *et al.* 2007) which we could not find in our data. Considering that previous and present results show no genetic structure along the Costa Brava which has an expansion of several hundred kilometers,

gene flow and connectivity within that area must be high. In such a high gene flow scenario Saenz-Agudelo and coauthors (2009) encountered that results from assignment tests were less accurate and that parentage analysis worked well in such conditions.
Also, we find no sweepstake reproduction events for *Tripterygion delaisi*, meaning that there was no difference in genetic variability between the adult and recruit samples (Hedgecock & Pudovkin 2011). In the bicoloured damselfish (*Stegastes partitus*) differences in genetic variability, relatedness and genetic distance were found when comparing the adult and recruit dataset (Christie *et al.* 2010). None of this was the case for the black faced blenny indicating that most likely a large proportion of the adults of the whole genetic population contribute to successful reproduction.

Approximately 15% of the recruits from the Blanes area were self-recruits with one of these offspring found directly adjacent to the fathers' nest. This suggests that even though genetic connectivity is high within the Costa Brava, which has an expansion of several hundred kilometers, there is still a considerable amount of local retention. In other studies much higher self-recruitment rates were found, as for example for anemonefish with up to 64% of juveniles returning to the natal habitat (Almany *et al.* 2007; Planes *et al.* 2009; Berumen *et al.* 2012). Such results lead to believe that self-recruitment is more important than previously considered (Jones *et al.* 2009). However, these studies were undertaken for species in a coral reef environment in an enclosed bay and data from other study systems or habitat types suggest that recruitment is lower. Although still in a coral reef habitat but on an open coastline 10% self-recruitment was documented across years (Saenz-Agudelo *et al.* 2012) and 17% was found for a highly mobile species forming reproductive aggregations (coral trout, *Plectropomus maculatus*, Harrison *et al.* 2012). Cowen and coauthors (2000) proposed that the marine system might not be an open system with high connectivity via large-scale dispersal but a closed one with larval retention. The findings in the last 10 years, including our study on *T. delaisi*, could suggest that self-recruitment, meaning larval retention, might depend mostly on the type of habitat. It seems that in an enclosed area self-recruitment is high, whereas on open coast lines the amount of larval retention lays between 5-20%.

In a system with high connectivity, low genetic structuring and low self-recruitment, paternity analysis only provides limited information, due to low numbers of encountered parent-offspring

pairs. We found that sibship analysis is such cases can provide valuable additional information. The encounter of sibling pairs at a distance of over 11km suggests a larger larval dispersal potential, whereas paternity analysis implicated a decrease in dispersal within a distance of only 1km. Therefore, sibship analysis revealed to be a valuable analytical tool to directly measure the real dispersal potential even on open coast lines. Also, it allows addressing other ecological questions related to larval dispersal. As larvae are almost impossible to track in the plankton, it is not clear if larvae stay with their kin in the plankton or if offspring from the same natal site disperse to different locations. Sibship analysis has been successfully used to reconstruct family structure and mating success in a variety of marine species (Dibattista *et al.* 2009; Kanno *et al.* 2010; de Mestral *et al.* 2012). A recent study on coral reef fish provided evidence that siblings can travel together throughout the whole planktonic larval phase (Bernardi *et al.* 2012). For the black faced blenny patterns show that the vast majority of fullsibs do not recruit to the same location. Although many groups of recruits were collected from the same exact locations, only one fullsib pair was encountered at the exact same rock. According to their otoliths, both siblings were born almost on the same day and spent the same amount of time in the plankton showing that they most likely recruited together. The dispersal patterns of kin can have large implications on the genetic composition and connectivity of a species and the better understanding of larval family structure can increase our knowledge on dispersal patterns.

Another important factor is the direction of larval dispersal. For *T.delaisi* paternity analysis revealed that the directional tendency for offspring was to move south-westward along the coast. Another study measuring direction of dispersal via parentage analysis found no directional consistency across seasons or years and dispersal was found to be multidirectional (Saenz-Agudelo *et al.* 2012). For *Tripterygion delaisi* a south-western dispersal along the coast from the analyzed locality would mean that the larvae will not be able to successfully recruit as the coastline is sandy beach habitat southwards and there is no available habitat to recruit to for approximately 100km. Although this needs further investigation, it demonstrates the importance of available habitat for recruitment success.

Parentage analysis via paternity and sibship analysis can provide direct dispersal measures and therefore help answering some of the great challenges in marine ecology. Due to the requirement of massive sampling up to now studies were mostly restricted to the tropics and to species on confined habitats. We demonstrate that the application of such methodology is possible in a

temperate environment such as the Mediterranean Sea and that it can be applied among others to highly abundant species with low genetic structuring. Such direct measurements of dispersal are essential to understanding true ongoing connectivity patterns in different marine habitats and show the degree of retention and self-replenishment even with apparent high connectivity and gene flow.

Concluding remarks and future directions

Even though certain patterns can be drawn from the different studies on connectivity and larval dispersal, indirect or direct measurements, there is a huge variation between species and study locations. As it is impossible to analyze connectivity patterns for many species and in most locations, results that can be extrapolated or proxies are needed. The proxies that have been studied and used to estimate larval dispersal are mostly of biological nature, e.g. life history traits. With the growing literature and the results documented in this thesis, connectivity whether on a large or a small scale can be determined by genetic markers and is largely influenced by its environment. Oceanographic processes define genetic structuring and dispersal is dependent on the openness of the surrounding habitat. This shows that to completely understand ecological aspects such as larval dispersal, geographical and physical environmental processes need to be considered.

Increased understanding of the importance of the environment on the dispersal and ecology of species demands the improvement of our knowledge and perception of the interaction between the environment and organisms. For a more complete picture and answers to some of the great challenges in ecology and evolution, we cannot only be looking at an organism or a population as isolated elements, but need to consider the interactions and impacts of the environment. Especially in the light of rapidly changing and modifying environmental conditions due to anthropogenic alterations, the study of adaptation can provide a clearer insight into the dynamic evolution, the fitness and survival of organisms. Understanding the species ability to adapt will not only increase our knowledge on the present day ecology but can indicate whether populations and species may adapt in response to environmental change. This type of multidisciplinary

approach will be relevant for the protection and conservation of a valuable and extraordinary ecosystem.

5. CONCLUSIONS

CONCLUSIONS

1. The endangered species *Epinephelus marginatus* showed a week but significant genetic structure and is not panmictic throughout its range. A network of Mediterranean-wide interconnected Marine Protected Areas taking genetic data into account is essential for the protection and long-term survival of the species.

2. The commercially relevant comber (*Serranus cabrilla*) revealed almost no genetic differentiation between localities in the eastern Mediterranean Sea and the central Mediterranean Sea as well as the Balearic Islands and the northern Spanish coast hinting at a large-scale connectivity. Nonetheless, comber samples from southern Spain are genetically distinct from the rest of the Mediterranean with the localities on the southeastern Spanish coast being an admixed population.

3. There was a strong correlation between genetic population structure of comber and oceanographic data in the western Mediterranean sea, indicating that population structure might not depend solely on the length of the pelagic larval phase or other life history factors, but is more determined by the oceanographic features, such as front and currents.

4. Approximately 15% of juveniles of *Tripterygion delaisi* in Blanes are self-recruits, which was revealed by the use of paternity analysis. This shows that even though population structure is low and connectivity high there is considerable retention.

5. For *Tripterygion delaisi*, fullsib analysis allowed for a direct evaluation of the dispersal potential of the species. Siblings were encountered at distances of twelve kilometers from each other, providing direct evidence of connectivity between habitats separated by beach, which adults cannot cross.

6. Resulting from the three studies (of *E. marginatus*, *S. cabrilla* and *T. delaisi*) evaluating connectivity on different scales it has become clear that dispersal is determined by the physical environment. This can be oceanographic features as well as the habitat type and structure in which the species lives in. Hence for a complete understanding of connectivity patterns, the physical environment needs to be considered.

7. A transcriptome de novo assembly of the brain tissue was produced with sequences from high throughput RNA sequencing. Most transcripts in the transcriptome had one isoform and the number of alternatively spliced isoforms increased with transcript length.

8. In *T. delaisi*, more genes were differentially expressed between the two male phenotypes than between males and females suggesting that phenotypic plasticity is a more important factor in differential gene expression during the reproductive period than sexual dimorphism.

9. The territorial male overexpresses genes related to synaptic plasticity and the sneaker male overexpresses genes involved in differentiation and development, which could mean that the sneaker male, although not always smaller in size, is still developing.

10. RNA-seq is a new technique that can be successfully used for the discovery of SNPs and development of markers for large-scale genotyping in non-model species. Due to the high cost, an efficient sample number should be chosen. With the use of seven individuals of *T. delaisi* for RNA sequencing the probability of discarding highly informative SNP loci, due to lack of observed polymorphism, is highly reduced and more than 12 samples does not considerably improve prediction of true population allele frequencies.

6. RESUMEN

Resumen

Introducción general

El Mediterráneo es un mar semi-cerrado, limitado por diferentes masas continentales, y con una costa densamente poblada. La población que habita dichas costas ha explotado y alterado los recursos marinos durante milenios. Por lo tanto, se considera que los impactos antropogénicos y las alteraciones son proporcionalmente más grandes en el Mediterráneo que en otros mares (Coll *et al.* 2010). El Mediterráneo contiene, asimismo, un elevado número de endemismos y una rica diversidad de especies, lo que le convierte en uno de los "hotspots" en biodiversidad del planeta (Myers *et al.* 2000; Macpherson 2002). Debido a tales amenazas la cuenca mediterránea se podría considerar uno de los "hotspots más calientes" de todas las áreas ricas en biodiversidad. Los peligros para este rico entorno son múltiples, pero en la actualidad los que tienen mayor impacto son: la pérdida de hábitat, el impacto de la pesca, la contaminación, la eutrofización y la presencia de especies invasoras. Por otro lado, el desarrollo de infraestructuras, el cambio climático y la actividad agrícola tienen efectos indirectos sobre la degradación de sus aguas.

Hay consenso en que la vida marina en el Mar Mediterráneo y otros mares de todo el mundo está cada vez más alterada y bajo una considerable amenaza por la actividad humana (Hutchings 2000). Unas medidas contra la degradación del medio marino, por ejemplo, son la regulación de la pesca hacia una pesca más sostenible y el establecimiento de redes de Áreas Marinas Protegidas (MPAs). El diseño de las MPAs a menudo se centra en la simple protección de una fracción de los hábitats clave, y se asume que los ecosistemas naturales persistirán en las áreas marinas protegidas y contribuirán a los hábitats fuera de estas zonas. Sin embargo, se ha demostrado que una red de Áreas Marinas Protegidas ofrece más protección que un conjunto de áreas protegidas individuales inconexas (Halpern & Warner 2003). Además, los estudios de modelización demuestran la gran importancia de la configuración espacial de zonas marinas protegidas para la persistencia de sus poblaciones (Kaplan *et al.* 2009). Aunque estas interacciones son evidentes en los resultados de los modelos, la falta de datos sobre la conectividad real entre MPAs impide su evaluación y, por lo tanto, un diseño apropiado de cualquier red. Esta conectividad está principalmente propiciada por la capacidad de dispersión de las larvas, y está influenciada por corrientes y otros procesos oceanográficos. La dificultad en seguir adecuadamente los desplazamientos de las larvas ha ocasionado que la conectividad solo

haya podido cuantificarse en los últimos años (Almany *et al.* 2009). Es por ello que para el establecimiento de redes de MPAs ecológicamente coherentes se necesita una amplia gama de estudios científicos, junto con objetivos alcanzables y claros. Estos objetivos, a menudo son poco exitosos debido a la complejidad del medio marino y de la dinámica de sus poblaciones. Por lo tanto, estudios integrales y multidisciplinarios que combinen modelos oceanográficos, ecología larval y especialmente genética de poblaciones, son esenciales para hacer frente a esta complejidad.

El estudio de la composición genética de las poblaciones biológicas se denomina genética de poblaciones. El concepto básico de la genética de poblaciones es el estudio de la variación genética intra e interpoblacional (Hartl & Clark 2007). Para que una población de individuos tenga éxito debe contener variabilidad genética. Por lo tanto, para obtener una visión general de la población, se analiza la variabilidad de los diferentes fenotipos y genotipos. La herramienta fundamental de este enfoque es el análisis y la cuantificación de las frecuencias alélicas. Éstas miden la frecuencia relativa de un alelo de un locus dado en la población. La cuantificación del número de alelos y su carácter común en una población puede proporcionar información sobre su diversidad genética, así como la riqueza de su acervo genético.

Por otra parte, las estimas de las frecuencias alélicas en diferentes áreas geográficas puede mostrar su grado de diferenciación genética. Si las frecuencias alélicas no son similares entre los individuos de dos áreas podemos considerar estas localidades como dos poblaciones genéticamente distintas. Esta disimilitud se mide como distancia genética y se establece habitualmente mediante el uso de los estadísticos F de Wright. Simplificado, si dos poblaciones son genéticamente distantes la una de la otra, mostrarán diferentes frecuencias de alelos y el valor de F_{ST} será alto (Hudson *et al.* 1992). En el caso de que dos poblaciones sean genéticamente distintas se asume que el intercambio de material genético o flujo génico es bajo.

La dispersión es un proceso ecológico fundamental que puede determinar la estructura genética de las poblaciones, así como la diversidad poblacional tanto a través de las escalas de tiempo evolutivas como ecológicas (Waples & Gaggiotti 2006). En escalas de tiempo evolutivo, al mitigar los efectos de la deriva genética y la reducción del lastre genético en las poblaciones pequeñas, la dispersión puede reducir el riesgo de extinción. En una escala de tiempo ecológico, la dispersión limita la dispersión espacial y por eso es un factor clave en la determinación de la

estructura espacial de las poblaciones. A pesar de su importancia, la comprensión de los patrones de dispersión sigue siendo un reto importante, especialmente en ecología marina. La medida en que las poblaciones están vinculadas entre ellas por el movimiento de los individuos se denomina conectividad (Palumbi 2003; Cowen & Sponaugle 2009a). El movimiento de una población a otra se puede lograr a través de la dispersión por larvas pelágicas o por la migración juvenil o adulta. Sin embargo, en el medio marino se ha demostrado que muchas especies, especialmente peces no-pelágicos, muestran un comportamiento de fidelidad como adultos al sitio que se han asentado y no migran. Por lo tanto, es sobre todo la etapa pelágica de las larvas la que permite la dispersión y el movimiento entre las poblaciones, manteniendo así la conectividad (Planes 2002).

Anteriormente, se asumía que, en general, las poblaciones marinas presentaban un alto flujo genético, favorecido por la inexistencia de barreras físicas, por lo que la conectividad se determinaba principalmente por la capacidad de dispersión de cada especie. Sin embargo, más recientemente se ha demostrado que diversas especies presentan una diferenciación espacial mayor de la esperada cuando tan sólo se consideran sus capacidades dispersivas potenciales (Calderón *et al.* 2007; Palero *et al.* 2008). Los patrones de conectividad han demostrado estar influenciados por otros factores, como las características oceanográficas (White *et al.* 2010), el comportamiento de las larvas (Gerlach *et al.* 2007) y la distribución del hábitat (Pinsky *et al.* 2012). En el Mediterráneo noroccidental, se ha observado recientemente que los procesos oceanográficos, como los patrones de corrientes, discontinuidades oceanográficas y frentes, son factores cruciales que influyen en la conectividad de las poblaciones (Galarza *et al.* 2009). Por lo tanto, para comprender los patrones de conectividad a gran escala, es importante tener en cuenta los procesos físicos que potencialmente influyen en el flujo génico entre poblaciones.

El enfoque de inferir la conectividad de las poblaciones mediante la diferenciación genética de poblaciones es eficaz para la identificación de las principales barreras biogeográficas y oceanográficas (Planes 2002) y permite la medida del flujo génico en escalas de tiempo evolutivo (Waples & Gaggiotti 2006). Por eso, la dispersión a gran escala y el intercambio entre las poblaciones en el medio marino se miden principalmente por métodos indirectos, como la genética de poblaciones, la ecología del comportamiento de las larvas o por modelos biofísicos (Cowen *et al.* 2007). Esto significa que cuando medimos la conectividad a través de la diferenciación de las poblaciones, realmente estamos midiendo procesos del pasado y del

presente, y la conectividad detectada podría verse comprometida por los rápidos cambios de hoy en día, tales como alteraciones antropogénicas (Cowen & Sponaugle 2009b). Por lo tanto, también es esencial comprender cómo y si la conectividad se mantiene en el tiempo presente. Esto se puede lograr mediante la medición directa de la dispersión larvaria, observando la cantidad de larvas que se dispersan y hacia dónde. Sin embargo, la obtención de estimas directas del movimiento de las larvas en el medio marino es difícil, debido a las dificultades técnicas en el seguimiento de las minúsculas larvas en el plancton. Por dichas razones, las trayectorias larvarias, en gran parte, son aún desconocidas, y comprender los patrones de dispersión actuales sigue siendo un gran desafío. El análisis de paternidad permite la estima directa de la conectividad, ya que la detección de juveniles y de sus padres permite, en muchos casos, la reconstrucción del movimiento de la larva, proporcionando una evidencia directa de la dispersión (Hedgecock *et al.* 2007). Sin embargo, la aplicación exitosa del análisis de paternidad requiere un muestreo masivo de una gran proporción de la población parental para el encuentro de pares de crías y padres, especialmente en las especies con una alta tasa reproductiva (Jones & Ardren 2003). Por esta razón, sólo unos pocos estudios han aplicado esta metodología en el medio marino (por ejemplo, Planes *et al.* 2009; Saenz-Agudelo *et al.* 2011; Berumen *et al.* 2012), y gracias a ellos, el conocimiento de los patrones de dispersión ha mejorado. No obstante, estos estudios se llevaron a cabo en ambientes de arrecifes de coral tropicales y, sobre todo en hábitats cerrados, como bahías. Es importante, por lo tanto, ampliar estas investigaciones a otros tipos de hábitat como el mar templado del Mediterráneo, y en áreas abiertas.

La conectividad, tanto desde el punto de vista evolutivo como ecológico, es esencial en la regulación de las poblaciones naturales. En consecuencia, las decisiones sobre la conservación correcta de nuestros ecosistemas marinos deben incluir datos de genética de poblaciones y procesos oceanográficos, así como medidas directas de la dispersión larvaria para una comprensión más completa sobre la dinámica poblacional de las especies (Almany *et al.* 2009).

Cada organismo tiene la capacidad de alterar características, tales como las características morfológicas o su ciclo de vida, en respuesta a condiciones externas. El material genómico de un individuo o genoma, se compone tanto de regiones codificantes (cuyas secuencias pueden ser transcritas) como no codificantes. La ‘activación’ de ciertos genes que codifican para proteínas está regulada y es diferente en cada momento. Esta activación de los genes que conduce a una

transcripción de secuencias de RNA y, posiblemente, a una traducción en una proteína, se denomina expresión génica. La expresión de los genes es lo que proporciona el control celular sobre la estructura y función y es la base para algunos procesos biológicos, tales como la morfogénesis, la diferenciación celular y la capacidad de adaptación de un organismo. La expresión de genes puede ayudar a comprender la fisiología básica, así como algunos procesos evolutivos como la adaptación. En particular, este último puede dar indicaciones sobre la capacidad de recuperación de las especies mediante la evaluación de su potencial de adaptación. Por otra parte, la expresión génica y el análisis de adaptación de los individuos también pueden ayudarnos a entender las diferencias entre poblaciones y cómo se están produciendo tales diferencias.

La expresión génica es dinámica, y el mismo gen puede 'actuar' de diferentes formas en diferentes circunstancias. Este puede ser el caso del mismo gen en dos organismos diferentes con genotipos similares, pero diferentes fenotipos. La plasticidad fenotípica es el cambio en el fenotipo de un genotipo en función del cambio producido en el medio. En otras palabras, la plasticidad permite la activación o desactivación de ciertos genes. Aunque su importancia se ha subestimado durante mucho tiempo, ahora está claro que es una parte determinante de cómo los organismos se desarrollan e interactúan con su entorno. Varios estudios demuestran claramente la influencia de la plasticidad en los procesos ecológicos y evolutivos de muchas especies (Huey & Kingsolver 1989; Pigliucci 2005; Ghalambor *et al.* 2007). Por otro lado, un mayor conocimiento sobre la plasticidad fenotípica ha cambiado las ideas de los paradigmas básicos en la biología evolutiva (Pigliucci 2007). Estos nuevos conocimientos sobre la variabilidad de la expresión génica parecen estar en mi opinión bajo el lema: "La vida no es sólo lo que (genéticamente) se recibe, sino que también es lo que se hace con ello".

Un cambio del fenotipo en los peces se observa a menudo en el momento de la reproducción, en el que dos tipos de machos, que presentan genotipos similares, cambian de comportamiento y/o fenotipo. Por la creación de una jerarquía social temporal, los machos presentan estrategias de apareamiento alternativas. Por ejemplo, un macho secundario puede entrar en el nido de los machos territoriales y fertilizar los huevos que ha puesto la hembra sin invertir en la protección del nido o atraer a la hembra (Immler *et al.* 2004). Aquí, se trata de una señal social externa, que provoca el cambio del fenotipo en el macho territorial que, a menudo, muestra una coloración

corporal diferente durante el período reproductivo. La estructuración social, en su mayoría, ha sido estudiada en el cerebro de cíclidos, debido a su diversidad fenotípica y a la facilidad para su manipulación y mantenimiento en el laboratorio (Taborsky 2008). Además, el salmón atlántico (*Salmo salar*) exhibe tácticas reproductivas alternativas y se ha investigado la plasticidad en el desarrollo de estos fenotipos reproductivos (Aubin-Horth & Renn 2009; Guiry *et al.* 2010). Los patrones de expresión de genes relacionados con la plasticidad fenotípica en las diferentes estrategias de apareamiento han sido analizados, ya sea por genes individuales o a través de análisis de microarrays (Aubin-Horth *et al.* 2005; Burmeister *et al.* 2005; Renn *et al.* 2008), pero no se ha llevado a cabo ningún intento para caracterizar la plasticidad fenotípica mediante un enfoque de todo el genoma.

Para una evaluación amplia y la comprensión de las dinámicas poblacionales tanto como para mirar la plasticidad fenotípica, he optado por trabajar con tres tipos de especies de peces: el mero (*Epinephelus marginatus*), la cabrilla (*Serranus cabrilla*) y un pez bleniforme (*Tripterygion delaisi*). Estas tres especies muestran rasgos muy distintos en su biología, incluyendo una diferente longevidad. Mientras que las tres especies forman una parte importante del ecosistema marino del Mediterráneo, cada especie tiene un estado de conservación diferente y una relevancia distinta para los humanos.

Epinephelus marginatus (Pisces: Serranidae, subfamily Epinephelinae), 'el mero', es una especie que vive en el Mar Mediterráneo y en el Atlántico oriental. También se puede encontrar en el Atlántico occidental, en aguas de América del Sur y en el Océano Índico occidental. Especialmente en el mar Mediterráneo, pero también en la mayoría de los otros lugares, el mero es una especie importante y de gran valor para las pesquerías costeras (Heemstra & Randall 1993). Debido a que es un pez de consumo habitual y una especie comercial está sobreexplotada y se considera como una especie bandera en la conservación y en el establecimiento de áreas marinas protegidas (MPAs) de la región mediterránea (Maggio *et al.* 2006). La especie es de gran tamaño, longeva (hasta 60 años), y crece lentamente con un inicio tardío de su madurez. Además, el mero tiene una reproducción compleja ya que es hermafrodita y cambia de sexo, de hembra a macho (Reñones *et al.* 2007). Habita en arrecifes de la costa rocosa, hasta una profundidad de unos 50 metros en la costa mediterránea (Heemstra & Randall 1993). Se ha observado su comportamiento reproductivo en la Reserva Marina de las islas Medes, en España,

y en otros lugares en los que el mero forma agregaciones reproductivas, mostrando además una fuerte fidelidad al lugar (Zabala *et al.* 1997; Hereu *et al.* 2006). Aunque se han documentado migraciones en muchas especies de la familia Serranidae, *Epinephelus marginatus* es una especie con una fuerte fidelidad territorial, por lo que la dispersión de las larvas pelágicas es previsiblemente el factor más importante en el flujo génico entre poblaciones (Sadovy *et al.* 1994). La larva del mero rara vez se ha capturado en el plancton, pero se sabe que permanece durante 21-30 días en la etapa pelágica (Macpherson & Raventós 2006). Igual que muchas otras especies marinas comercialmente sobreexplotadas y en peligro (Jones *et al.* 2007), los datos sobre su capacidad real de dispersión son escasos.

Serranus cabrilla es una especie común, que habita el Atlántico oriental y el mar Mediterráneo y se encuentra en praderas de fanerógamas y fondos rocosos, con un amplio rango batimétrico (5 - 500 m). El serrano se considera uno de los depredadores más importantes de peces juveniles e invertebrados vágiles (Guidetti & Cattaneo-Vietti 2002). El serrano es económicamente relevante y está incluido en los catálogos de la FAO como especie de interés para la pesca en el Atlántico centro-oriental, el Mediterráneo y el Mar Negro. Sus larvas permanecen en el plancton durante 21-28 días (Raventós & Macpherson 2001) y se han recogido no sólo cerca de la costa, sino también sobre la plataforma continental, a una distancia considerable del hábitat de los adultos (Sabates *et al.* 2003). Se conocen muchos aspectos de su ecología y biología (Torcu-Koc *et al.* 2004), sin embargo, no hay información sobre su estructura genética poblacional ni sobre el grado de conectividad entre poblaciones.

Tripterygion delaisi, también llamado moma amarilla, es un pequeño pez de la costa rocosa del mar Mediterráneo y de la costa atlántica oriental (Carreras-Carbonell *et al.* 2007; Domingues *et al.* 2007). Estos bleniformes viven camuflados en la roca o en las algas en las que habitan la mayor parte del año. En primavera, cuando se inicia el período reproductivo, algunos machos cambian su color, con una cabeza negra y una coloración amarilla brillante en el resto del cuerpo. Estos machos empiezan a proteger un territorio pequeño, lo que se conoce como su nido, contra depredadores y otros machos secundarios (Jonge & Videler 1989). Los machos secundarios y las hembras no cambian su coloración y muestran el mismo fenotipo a lo largo de todo el año. El macho dominante intenta atraer a la hembra, mediante movimientos de cortejo por encima del nido. Posteriormente, la hembra pone los huevos directamente sobre la roca, esponja o algas,

donde se encuentra el nido y el macho dominante fertiliza los huevos expulsando los espermatozoides directamente sobre ellos (Wirtz 1978). Sin embargo, el macho secundario puede pasar por el nido, a cierta distancia, y expulsar el esperma y por lo tanto, intervenir en la reproducción.

La estructura genética de las poblaciones de este bleniforme ha sido evaluada en estudios previos, que proporcionaron un primer análisis sobre la conectividad entre diferentes localidades del Mediterráneo y Atlántico oriental, (Carreras-Carbonell *et al.* 2006, 2007). Debido al proceso de acoplamiento particular de esta especie, especialmente porque los machos dominantes construyen nidos y se convierten en territoriales, y por el cambio de color, que los hace fáciles de observar, esta especie es un candidato ideal para analizar el grado de parentesco y de dispersión en cualquier población. Por otra parte, ya que la especie muestra diferentes fenotipos durante el período reproductivo y los peces de uno u otro fenotipo son fáciles de atrapar, *Tripterygion delaisi* también es óptimo para la investigación genómica de la plasticidad fenotípica.

Objetivos

1. Estudiar la estructura genética poblacional de una especie amenazada, *Epinephelus marginatus*, y proporcionar información para el establecimiento de unidades de gestión.

2. Analizar la estructura genética de una especie de interés comercial (*Serranus cabrilla*), en el mar Mediterráneo y evaluar la conectividad entre sus poblaciones.

3. Establecer la relación entre la estructura genética de *Serranus cabrilla* y los procesos oceanográficos del Mediterráneo occidental, tales como frentes oceanográficos y corrientes, y como afectan al flujo génico y a la conectividad entre poblaciones.

4. Medir directamente la dispersión de las larvas de *Tripterygion delaisi* y, a su vez, identificar el patrón de conectividad a pequeña escala, mediante un estudio de paternidad.

6. Construir un ensamblaje *de novo* del transcriptoma a partir de tejido cerebral de *Tripterygion delaisi* y determinar las principales funciones biológicas que se expresan en el genoma.

5. Averiguar el perfil de la expresión génica diferencial entre los cerebros de los machos territoriales, machos secundarios y hembras de *Tripterygion delaisi* para estudiar las bases moleculares del dimorfismo masculino.

6. Desarrollar marcadores de polimorfismo de un solo nucleótido (SNP) a partir de técnicas de secuenciación masiva y evaluar el número óptimo de individuos necesarios para un desarrollo eficiente de este tipo de marcadores en especies no modelo.

Resultados y discusión general

Uno de los retos más importantes en ecología marina es la comprensión de las capacidades de dispersión de las especies. Esta capacidad es uno de los principales determinantes de la dinámica de las poblaciones y del rango geográfico de las especies (Bowler & Benton 2005). *Epinephelus marginatus* y *Serranus cabrilla* son dos especies de peces del Mediterráneo con diferentes capacidades de dispersión. La primera especie tiene una amplia distribución, que incluye el Mar Mediterráneo, el Océano Atlántico y llega hasta las costas de Brasil y Argentina en el Atlántico occidental (Heemstra & Randall 1993). Como los adultos de esta especie muestran una alta fidelidad territorial, la extensión de su distribución geográfica sugiere una dispersión de larvas pelágicas a muy largas distancias (Heemstra & Randall 1993; Hereu *et al.* 2006). La segunda especie, el serrano, a pesar de que se encuentra en el Mediterráneo y en la costa africana hasta el sur de África, no parece ser capaz de cruzar el Océano Atlántico (Guidetti & Cattaneo-Vietti 2002). Una especie con una distribución geográfica amplia podría sugerir un gran potencial de dispersión, por lo que localidades distantes deberían estar conectadas genéticamente. En el caso del serrano y el mero, no encontramos una estructuración genética diferente entre localidades distantes. Por ejemplo, ambas especies tenían valores muy bajos de F_{ST} al comparar la muestra de Grecia con la de las Islas Baleares, que se encuentran a una distancia de 1800 kilómetros la una de la otra. En el mero se observó que sólo aquella poblaciones situadas a mayor distancia se diferenciaban genéticamente, sobre todo con las comparaciones incluyendo la muestra de Senegal, revelando que esta especie no es panmítica en todo su rango de distribución: En cambio sí que muestra panmixia en la mayor parte del mar Mediterráneo, con la excepción de Argelia.

En el caso del serrano, se identificaron dos grupos genéticamente distintos con una división clara localizada entre las discontinuidades oceanográficas del canal de Ibiza (IC) y el frente del Almeria-Oran (AOF), revelando una población genéticamente mezclada entre ambos. Aunque había una baja diferenciación genética entre poblaciones con mayor influencia del agua mediterránea, en la costa española se observó un flujo génico direccional en *S. cabrilla* con una dispersión del oeste al este a través del AOF, del norte hacia el sur en el IC y desde el sur del IC hacia las Islas Baleares. Las correlaciones entre los datos genéticos y oceanográficos fueron altamente significativas.

La existencia de panmixia en la estructura poblacional de las especies marinas era, hasta hace unos años, el patrón más esperable, dado la "homogeneidad" de las aguas como medio de dispersión. En estos casos, se esperaría un ligero aislamiento genético por distancia (IBD) siguiendo un modelo de piedras de paso donde la distancia genética aumenta con la distancia geográfica (Wright 1969). Sin embargo, cada vez está más claro que una gran variedad de especies marinas no muestran ese aislamiento por distancia (Bradbury & Bentzen 2007). *Epinephelus marginatus*, así como *Serranus cabrilla,* mostraron dicho patrón, pero el resultado dependía en ambos casos de la inclusión o la exclusión de ciertas localidades de muestreo. En el mero, si se incluían todas las localidades estudiadas, se encontraba IBD debido a las muestras distantes y genéticamente distintas del océano Atlántico, mientras que excluyendo estas muestras no había IBD significativo. Del mismo modo, pero a la inversa, el serrano no tenía patrón de IBD cuando se incluyeron todas las muestras, pero se encontró un patrón significativo cuando el análisis sólo incluía muestras de la costa española. Ello indica que el IBD depende, en gran parte, de los puntos de muestreo que se incluyen en el estudio. Además, demuestra que el grado de la dispersión y el flujo génico son más complejos de lo que se esperaba, porque la diferenciación genética y los patrones de flujo genético no dependen sólo de la distancia.

Otro factor que se ha sugerido como indicador de la capacidad de dispersión de una especie es la duración del período de la larva pelágica (PLD) (Sponaugle *et al.* 2002; Macpherson & Raventós 2006; Selkoe & Toonen 2011). En la revisión de Selkoe & Toonen (2011) se muestra una correlación moderada entre los datos de la estructura poblacional y el PLD, lo que implicaría que PLD y medidas de diferenciación genéticas reflejan el potencial de dispersión. El mero y el serrano tienen un PLD alrededor de 21-30 días (Raventós & Macpherson 2001), lo que

significaría que, al considerar sólo el PLD, podrían presentar el mismo potencial de dispersión y, a su vez, una estructura genética similar. Sin embargo, estas especies tienen una distribución diferente y no encontramos el mismo patrón de diferenciación poblacional. El serrano, al contrario que el mero, mostró estructuración genética clara también en localidades geográficamente cercanas. Shanks (2009) revisó datos sobre PLD y el potencial de dispersión y encontró que se correlacionan. Sin embargo, en las especies con PLDs más largos, las distancias de dispersión pueden estar sobreestimadas si las larvas se quedan más cerca de la costa, donde las corrientes son más lentas. Por lo tanto, las características del ciclo biológico y el comportamiento larval podrían desempeñar un papel fundamental para estimar adecuadamente la capacidad de dispersión.

Ni la distribución geográfica, ni el aislamiento por distancia, ni el PLD parecen ser buenos estimadores del potencial de dispersión y que permitan explicar la estructura genética detectada en las poblaciones del mero y del serrano. Si ninguno de estos parámetros es adecuado, entonces ¿qué es lo que define la estructura poblacional de las especies marinas? Aunque lo más probable es una combinación de varios factores se ha puesto también de manifiesto la importancia de los procesos y las barreras oceanográficas para explicar la estructura de las poblaciones de organismos marinos: (Galarza *et al.* 2009).

En el Mediterráneo, los patrones de circulación de corrientes y otros procesos oceanográficos están bien descritos (Millot 1999) y se han documentado varios frentes oceanográficos en el Mediterráneo occidental. Después de pasar por el estrecho de Gibraltar, el agua del Atlántico se encuentra con el agua del Mediterráneo, de mayor densidad, que genera diferentes discontinuidades oceanográficas. Por ejemplo, el Frente Almería-Orán (AOF), el Frente Balear (BF) o el canal de Ibiza (IC), donde puede haber barreras al flujo génico (Patarnello *et al.* 2007; Galarza *et al.* 2009; García-Merchán *et al.* 2012). Esto permite la evaluación de los efectos de dichas características oceanográficas sobre la estructura genética de la población. En el caso de los meros no se observó ningún efecto de los procesos oceanográficos mencionados sobre la estructura genética. Sin embargo, las muestras de Argelia se diferenciaron genéticamente del resto del Mediterráneo, lo que podría estar ocasionado por los remolinos de la zona que ralentizan la corriente hacia el este, la cual, a su vez, está conectada con los remolinos en mar abierto que llegan a las Islas Baleares (Millot 2005). Esto podría explicar por qué la población de

Argelia se diferencia genéticamente del resto, pero está genéticamente más conectada con la población de Menorca. Sin embargo, se debe tener en cuenta que con un muestreo más amplio en el mar de Alborán quizás se podría encontrarse un efecto del AOF en el mero. Se observó una influencia más clara de los procesos oceanográficos en *Serranus cabrilla* ya que las tres barreras (AOF, IC y BF) se ha detectado que tienen un efecto disminuyendo la conectividad genética del serrano. Mediante simulaciones de dispersión de partículas, se detectó una correlación significativa entre la estructuración genética y la dispersión de las larvas por las corrientes. La importancia de características oceanográficas, tales como las corrientes y frentes oceanográficos, está siendo claramente reconocidos en el medio marino (Selkoe *et al.* 2006; White *et al.* 2010). Galarza y coautores (2009) evaluaron estas barreras, en concreto para siete especies diferentes de peces, y obtuvieron resultados distintos dependiendo de la especie. Se encontraron distintos resultados incluso para especies estrechamente relacionadas, con similares rasgos biológicos. Algunas de las razones para tales diferencias, que rara vez se han considerado, podrían ser los cambios estacionales en el flujo de las corrientes o la variabilidad temporal en los procesos oceanográficos (Stenseth *et al.* 2006). En el serrano, se observó que el canal de Ibiza dificulta el movimiento de las larvas hacia el sur y, por lo tanto, nos encontramos con conectividad limitada entre las poblaciones de cada lado. El mero, sin embargo, no se ve afectado por esta barrera, lo que, probablemente, es debido a la estacionalidad de la misma, que coincide sólo con la época reproductiva de *Serranus cabrilla* (primavera), pero no de *Epinephelus marginatus* (verano).

La estructura poblacional de ambas especies es distinta, como se ha visto en los diferentes estudios. Las poblaciones de *Epinephelus marginatus* están conectadas a una mayor escala, que es un aspecto positivo en la protección de la especie, ya que esto significa que a pesar del declive de sus poblaciones no se está extinguiendo un acervo genético particularmente diferenciado. Sin embargo, estos datos pueden reflejar patrones del pasado ya que los rápidos descensos en las poblaciones pueden dar lugar a una disminución de la conectividad genética. Por lo tanto, proponemos tres unidades distintas de gestión o conservación: Senegal (Atlántico), Argelia (costa mediterránea africana del noroeste) y el resto del Mediterráneo. La estructura de la población y los patrones de flujo génico de *Serranus cabrilla* están influenciados por las corrientes predominantes y por los procesos oceanográficos. Se identificaron cuatro unidades de gestión, separadas por tres barreras oceanográficas: IC, AOF y BF. Para asegurar la supervivencia a largo plazo de las especies es fundamental identificar la conectividad entre las

diferentes unidades de gestión e implementarlo en el diseño de redes de reservas marinas. Nuestros resultados demuestran además la necesidad de incluir datos oceanográficos en los estudios de genética de poblaciones para entender la división de los patrones de conectividad y para estudiar la dirección de la dispersión.

El medioambiente no solo influye sobre la estructura poblacional de las especies sino también sobre los genes que se están expresando, así que el transcriptoma de cada individuo varía en el tiempo y en función de las variables externas, sean ambientales o sociales. Aunque el material genómico sea el mismo, distintos genes se expresan en los diferentes tipos de tejidos y para los diferentes fenotipos. Esta expresión diferencial puede dar una idea de las funciones biológicas de un determinado tipo de tejido o fenotipo. En el pasado, no era posible conocer la expresión del transcriptoma sin la existencia de un genoma de referencia, sin embargo, gracias a la RNAseq y los secuenciadores de alto rendimiento, esta dificultad se ha solucionado. Wang y coautores (2009) señalaron que este método es un enfoque revolucionario para la transcriptómica y, debido a los avances en la tecnología de secuenciación, ahora también es factible en organismos no-modelo. A pesar de todos los avances, el ensamblaje *de novo* de un transcriptoma todavía es un reto importante y no es un proceso trivial (Martin & Wang 2011). El ensamblaje *de novo* del cerebro de *Tripterygion delaisi* resultó tener más de 300.000 contigs, incluyendo isoformas empalmadas diferencialmente. Se ha desarrollado una gran variedad de métodos y herramientas bioinformáticas para optimizar los ensamblajes en ausencia de un genoma de referencia. Sin embargo, el objetivo final es la construcción de genes completos, lo que no está resuelto totalmente. Algunos ensambladores, como Trinity (Grabherr *et al.* 2011), pueden reconstruir completamente una gran proporción de los transcritos, pero una gran cantidad todavía permanece fragmentado. A pesar de que esta fragmentación puede dar lugar a un sesgo en la sobreexpresión de los transcritos más grandes, los niveles de expresión con RNAseq parecen ser consistentes y robustos (Oshlack & Wakefield 2009). Además, cuando se compara la metodología con microarrays, sólo un carril de Illumina representa la misma cantidad de datos que un microarray entero, lo que permite la identificación de un mayor número de genes expresados diferencialmente. Sin embargo RNAseq permite hacer, al mismo tiempo, análisis adicionales, tales como la detección de genes de baja expresión, variantes alternativas de corte y empalme y nuevos transcritos (Marioni *et al.* 2008). El descubrimiento de nuevas secuencias es una de las ventajas de RNAseq, sin embargo, si no existen las secuencias anotadas en las bases de datos no

se podrá asociar ninguna función con un gen determinado. El porcentaje de los transcritos anotados es bajo especialmente en las especies que no son modelo. En estos momentos, el reducido porcentaje de detección de genes homólogos es una limitación habitual del estudio en una especie no-modelo, en el cerebro de *Tripterygion delaisi* encontramos sólo el 30% de los transcritos anotados. Estos valores son similares a los encontrados en avispas donde sólo el 37% tenían una descripción funcional (Ferreira *et al.* 2013). A pesar de las dificultades y limitaciones, de las grandes cantidades de datos generados con esta metodología se pueden extraer resultados evolutivos importantes. Para *T. delaisi* esta aproximación permitió analizar las diferencias de expresión del genoma entre diferentes fenotipos en el contexto de dominancia social.

Se determinó la base molecular que diferencia a los tres fenotipos de *Tripterygion delaisi*: macho territorial, macho secundario y hembra. Se observó que la comparación entre machos territoriales y secundarios tenía el mayor número de transcritos diferencialmente expresados. Resulta sorprendente que haya más genes expresados diferencialmente entre los dos fenotipos masculinos, ya que un gran número de estudios indican grandes diferencias genéticas entre los sexos, especialmente en el cerebro (Carruth et al. 2002, Arnold et al. 2004). Esto indica que, durante el período de reproducción, la plasticidad fenotípica juego un papel más importante que el dimorfismo sexual. La plasticidad fenotípica, la capacidad de un genotipo de cambiar su fenotipo para adaptarse a las condiciones externas, ha recibido una considerable atención en ecología evolutiva, pero siempre en menor medida que los cambios fenotípicos adaptativos y selectivos (Pigliucci 2005; Oliveira 2012). En el bleniforme *Tripterygion delaisi* se muestra esta plasticidad fenotípica a una escala temporal, ya que el período reproductivo sólo dura tres meses y todos los varones manifiestan el mismo fenotipo durante el resto del año.

Curiosamente, el macho territorial sobreexpresa más genes diferencialmente que los otros dos fenotipos, lo que posiblemente indica que tiene que pasar por más cambios durante estos tres meses de período reproductivo. Los genes que expresa diferencialmente el macho territorial estaban en su mayoría relacionados con la plasticidad sináptica. Estos genes son necesarios para que las neuronas puedan modificar su capacidad de sinapsis (Zucker & Regehr 2002). En los roedores, por ejemplo, se demostró que uno de los genes expresados en *T.delaisi* está involucrado en la transducción de la señal que condiciona su comportamiento (Berger & Roberts 2005). Los genes sobreexpresados en el macho secundario están relacionados con la

diferenciación y el desarrollo, como por ejemplo COL1A2, que codifica en el colágeno tipo I alfa. Este gen está bien estudiado por su relación con el desarrollo de los huesos (Karsenty & Wagner 2002). Este resultado sugiere que, aunque este tipo de machos se reproduce, no está completamente desarrollado.

Una consecuencia inmediata de las divergencias de fenotipos es la eficacia biológica debido a un aumento en el rendimiento reproductivo. Algunos análisis de especies con estrategias de apareamiento alternativas mostraron que el macho secundario de un pez de aguas dulces (*Rhodeus cericeus*) fertiliza con éxito un 30% de los huevos dentro de un nido (Reichard *et al.* 2004), mientras que el pez luna (*Lepomis gibbosus*) fertiliza un 15% (Rios-Cardenas & Webster 2008) y el macho secundario del góbido *Pomatoschistus minutus* es padre de un 5-10% de la descendencia (Malavasi *et al.* 2001). De acuerdo con nuestros resultados de expresión génica en el cerebro, el macho secundario invierte en genes ligados con su desarrollo. Dado que el macho secundario de *T. delaisi* tiene testículos proporcionalmente mayores en comparación al macho territorial, (Jonge *et al.* 1989 y observaciones personales), un análisis adicional de la expresión del tejido testicular podría revelar si este desarrollo es de carácter reproductivo.

La obtención de un transcriptoma para la especie *Tripterygion delaisi* y el análisis de la expresión diferencial en diferentes individuos nos ha permitido poder aislar marcadores para un estudio de parentesco a gran escala. La estimación directa de los patrones de dispersión mediante el análisis de parentesco debe hacerse a partir de un amplio muestreo de adultos y juveniles y utilizando un número elevado de marcadores genéticos. Lamentablemente, casi todas las especies marinas pueden ser consideradas especies 'no-modelo', de las que a menudo no existen marcadores genéticos disponibles y, si existen, el número es reducido. Por otra parte, genotipar un gran número de muestras requiere un método de alto rendimiento para ahorrar tiempo y recursos. En *Tripterygion delaisi* hemos desarrollado 192 marcadores polimórficos de nucleótidos simples (SNPs) a través de la secuenciación masiva del transcriptoma (RNAseq). Una de las ventajas del uso de SNPs como marcadores genéticos es la gran densidad en todo el genoma (Brumfield *et al.* 2003) y que la búsqueda de una gran cantidad de SNPs es relativamente fácil. Sin embargo, después de la aplicación de unos criterios de filtro, el número de SNPs útiles se reduce muy rápidamente. Cuando se comparan los escasos datos publicados sobre el desarrollo de SNPs en especies marinas, parece que la cantidad de posibles zonas

variables depende de la intensidad de la secuenciación en el experimento RNA-seq (por ejemplo, Renaut *et al.* 2010; De Wit & Palumbi 2012). Por lo tanto, si hay necesidad de miles de SNPs, es mejor secuenciar más intensamente o realizar la secuenciación del exoma, tal como se demuestra en un estudio sobre el arenque del Atlántico (Lamichhaney *et al.* 2012).

Un problema cuando se utiliza RNAseq son los límites entre intrón e exón (IEB), ya que sólo se secuencian los exones, sin ninguna indicación de dónde se inician los intrones. Esto si no se tiene en cuenta en el momento de diseñar los cebadores para identificar los SNPs podría conducir a un fallo de amplificación. En nuestro estudio se observó que menos del 10% de los loci putativos fallaron debido a IEBs en proximidad al lugar de un SNP. Un método para evitar tal fallo de amplificación es evaluar este SNP mediante PCR (Storer *et al.* 2012), pero es costoso y requiere mucho tiempo. Proponemos identificar regiones homólogas en los genomas de especies estrechamente relacionadas y evaluar la distancia del supuesto sitio de SNP a la IEB. Aplicando esta metodología, el fallo de amplificación debido a IEB se eliminó por completo.

La producción de librerías de ADNc para RNAseq es larga y costosa y sería óptimo reducir al mínimo el número de librerías / muestras individuales para el desarrollo de marcadores tipo SNP. Mediante bootstrap se estimó el número de muestras óptimo para el desarrollo de SNPs a partir de los datos RNAseq. Para evitar el genotipado de loci monomórficos se estimó que sólo es necesario utilizar siete individuos secuenciados por separado. También se observó que, con doce individuos, la frecuencia del alelo en la población es igual que con un número superior de individuos, por lo que este número puede considerarse como el máximo necesario para este tipo de estudios.

Aunque el desarrollo de SNPs para una especie no-modelo puede ser un reto, si hay suficiente profundidad de secuenciación y se aplican los filtros que hemos propuesto, se pueden obtener fácilmente loci útiles y potentes. Conviene recordar que entre las muchas ventajas de los marcadores SNP, en comparación con otro tipo de marcadores moleculares como los microsatélites, una de las más importantes es que permite analizar en poco tiempo un gran número de individuos, lo que facilita aquellos estudios ecológicos que requieren un gran número de muestras. El genotipado con SNPs puede ser muy eficiente ya que puede facilitar los estudios a gran escala con grandes cantidades de muestras y / o marcadores (Everett *et al.* 2011). A lo largo de mi tesis, he trabajado con varios marcadores genéticos, como loci microsatélites, genes

mitocondriales y SNPs. El desarrollo y la optimización de marcadores microsatélites es lento y la determinación del genotipo de este tipo de marcador se asocia a menudo con un sesgo de observación (Hauser *et al.* 2011). Tales marcadores son muy útiles y debido a su alto polimorfismo se necesitan menos loci. Sin embargo, para la evaluación de los patrones de dispersión a través del análisis de parentesco de *T. delaisi*, se tuvieron que genotipar cerca de 2.000 muestras, lo que evidentemente requiere un genotipado eficiente en tiempo y recursos, que se logró con los 192 marcadores SNPs aislados a partir de RNAseq.

El bleniforme *Tripterygion delaisi* mostró no tener estructuración poblacional dentro de las muestras recogidas a lo largo de la Costa Brava, lo que demuestra que los padres, así como los descendientes de Blanes y otros lugares de muestreo, como Tossa, pertenecen a la misma población genética. Esto también, avala la conclusión de un estudio anterior con loci microsatélites, que revela una única unidad genética en toda la Costa Brava (Carreras-Carbonell *et al.* 2006). Teniendo en cuenta que los resultados anteriores y actuales no muestran estructura genética en la Costa Brava, con varios cientos de kilómetros de línea de costa, el flujo genético y la conectividad deberían ser elevados. Sin embargo, al analizar con loci microsatélites el origen de reclutas recogidos en Blanes se detectaron mediante test de asignación migrantes de distintas poblaciones alejadas (Carreras-Carbonell *et al.* 2007). En un escenario de elevado flujo génico Saenz-Agudelo y coautores (2009) encontraron que las pruebas de asignación poblacionales, como las usadas en Carreras-Carbonell *et al*. (2007), eran menos precisas que el análisis de parentesco.

Con el estudio de parentesco llevado a cabo en esta tesis con los 192 SNPs, aproximadamente el 15% de los reclutas de la zona de Blanes podemos decir que son autoreclutas e incluso uno de estos descendientes fue encontrado justo al lado del nido de su padre. Esto sugiere que, a pesar de que la conectividad genética es alta dentro de la población, existe una retención considerable. En otros estudios se encontraron tasas de autoreclutamiento mucho más altas, como por ejemplo en el pez payaso, con un máximo de 64% de juveniles que regresaron a su hábitat natal (Almany *et al.* 2007; Planes *et al.* 2009; Berumen *et al.* 2012). Estos resultados llevan a pensar que el autoreclutamiento pueda ser más importante de lo que antes se consideraba (Jones *et al.* 2009). Sin embargo, estos estudios se llevaron a cabo en arrecifes de coral y en una bahía cerrada, y los datos en habitats más abiertos sugieren que el autoreclutamiento es menor. Por ejemplo, en un

arrecife de coral, pero en una costa abierta, el autoreclutamiento que se documentó fue de un 10% a lo largo de los años (Saenz-Agudelo *et al.* 2012) y un 17% en una especie de gran movilidad que forma agregaciones reproductivas (trucha de coral, *Plectropomus maculatus*, Harrison *et al.* 2012). Cowen y coautores (2000) propusieron que el sistema marino podría no ser un sistema abierto con alta conectividad a través de la dispersión a gran escala, sino más bien uno cerrado con mayor retención de las larvas. Los hallazgos de los últimos 10 años, incluyendo nuestro estudio sobre *T. delaisi*, sugerirían que el autoreclutamiento, es decir la retención larval, podría depender sobre todo del tipo de hábitat. Parece ser que, en un lugar cerrado, el auto-reclutamiento es alto, mientras que en las costas abiertas la magnitud de la retención larval puede ser menor y tendría un rango entre el 5% y 20%.

En un sistema con alta conectividad, baja estructuración genética y bajo autoreclutamiento, el análisis de parentesco sólo proporciona información limitada, debido al bajo número de parejas de padres e hijos que se encuentran. En nuestro estudio demostramos que el análisis de hermanos en esos casos puede proporcionar información adicional valiosa. El hallazgo de parejas de hermanos en *Tripterygion delaisi* a una distancia de más de 11 kilómetros sugiere un gran potencial de dispersión de las larvas, mientras que el análisis de parentesco tan solo indica una dispersión con una distancia de 1 km. Por lo tanto, el análisis de hermanos se revela como una herramienta analítica valiosa para medir directamente el potencial de dispersión real, incluso en las líneas de costa abiertas. Además, permite contestar otra cuestión ecológica relacionada con la dispersión de las larvas; como las larvas son casi imposibles de rastrear en el plancton, no está claro si se quedan con sus parientes en el plancton o si los descendientes del mismo lugar de nacimiento se dispersan por diferentes lugares. El análisis de hermanos ha sido utilizado con éxito para reconstruir la estructura familiar y el éxito de apareamiento en varias especies marinas (Dibattista *et al.* 2009; Kanno *et al.* 2010; de Mestral *et al.* 2012). Un estudio reciente sobre los peces de arrecife de coral evidencia que los hermanos pueden viajar juntos a lo largo de toda la fase larval planctónica (Bernardi *et al.* 2012). En *T. delaisi* los patrones muestran que la gran mayoría de hermanos no reclutan en el mismo lugar. Aunque se colectaron muchos grupos de reclutas en el mismo sitio, sólo un par de hermanos se encontraron en la misma roca. A partir del análisis de los otolitos, se observa que ambos hermanos nacieron durante los mismos días y pasaron la misma cantidad de tiempo en el plancton, indicando que se reclutaron a la vez. Los patrones de dispersión de los parientes pueden tener grandes consecuencias en la composición

genética y en la conectividad de cualquier especie lo que puede aumentar nuestro conocimiento sobre los patrones de dispersión de las especies.

Otro factor importante es la dirección de la dispersión de las larvas. En el análisis de parentesco, *Tripterygion delaisi* mostró que la tendencia direccional de los juveniles era moverse hacia el sur-oeste a lo largo de la costa. Otro estudio, que midió la dirección de la dispersión a través del análisis de parentesco, no encontró consistencia direccional a través de las estaciones o años y se encontró que la dispersión era multidireccional (Saenz-Agudelo *et al.* 2012). Ello implicaría que sólo se retienen pocas larvas en las costas abiertas y que la gran mayoría se dispersa a lugares adyacentes (Saenz-Agudelo *et al.* 2012). Para *T. delaisi* una dispersión al sur-oeste de la costa significaría que las larvas no serían capaces de reclutar con éxito, ya que en esa dirección la costa es arenosa y no hay ningún hábitat apropiado disponible en decenas kilómetros. Aunque este aspecto requiere estudios posteriores, se observa que la disponibilidad del hábitat puede tener consecuencias importantes en la estructura poblacional de esta especie.

Los análisis de paternidad y hermandad pueden proporcionar medidas directas de dispersión y, por lo tanto, ayudar a responder a algunas de las grandes cuestiones de la ecología marina. Debido a la exigencia de muestreos masivos, los estudios han sido en su mayoría restringidos a los trópicos y a especies con un hábitat semicerrado. Nuestro estudio muestra resultados en un ambiente templado, como el mar Mediterráneo, y que se puede aplicar en especies muy abundantes con poca estructuración genética. Asimismo, se concluye que las medidas directas de la dispersión son esenciales para la comprensión de los patrones de conectividad y muestran el grado de retención y autoreposición incluso con alta conectividad y flujo génico.

Observaciones finales y orientaciones futuras

A pesar de que se pueden extraer ciertos patrones de los distintos estudios sobre conectividad y dispersión de las larvas, hay una gran variación entre las especies y los lugares de estudio. Como es imposible analizar los patrones de conectividad de muchas especies y en muchos lugares, se necesitan resultados o metodologías que los representen y que puedan ser extrapolables. La mayor parte de los resultados de dispersión larvaria se han relacionado con características biológicas de las especies. No obstante, nuestros resultados y los de otros autores indican que la

conectividad a gran o pequeña escala está determinada, en gran medida, por su entorno. Los procesos oceanográficos que definen la estructuración genética y la dispersión dependen de la orientación y características del hábitat circundante. Esto demuestra que, para entender completamente el potencial de dispersión larval, se deben considerar aspectos geográficos y de oceanografía física.

La importancia del entorno en la dispersión de las especies exige mejorar nuestro conocimiento de la interacción entre dicho entorno y los organismos. Para obtener una imagen más completa y respuestas a algunos de los grandes desafíos de la ecología y la evolución, no sólo debemos mirar un organismo o una población como elementos aislados, sino que hay que tener en cuenta las interacciones y los efectos del medio ambiente. Especialmente, a la luz de la rápida modificación de las condiciones ambientales debido a alteraciones antropogénicas, el estudio de la adaptación puede proporcionar una visión más clara de la dinámica, evolución, y eficacia biológica de los organismos. La comprensión de esta capacidad de adaptación aumentará nuestro conocimiento sobre las respuestas de las poblaciones y de los individuos a los cambios del entorno. Este tipo de enfoque multidisciplinario será relevante para la comprensión completa de la ecología de las especies marinas y la protección y la conservación de un ecosistema extraordinario.

Conclusiones

1. El mero (*Epinephelus marginatus*) mostró una débil, pero significativa, estructura genética y no es panmíctico en todo su rango de distribución. Una red de áreas marinas protegidas interconectadas por todo el Mediterráneo que consideren los datos genéticos es esencial para la protección y la supervivencia a largo plazo de la especie.

2. El serrano (*Serranus cabrilla*), que es una especie comercialmente relevante, mostró una escasa diferenciación genética entre el Mediterráneo oriental, central, Islas Baleares y la costa norte española, lo que implica una amplia conectividad. Sin embargo, las muestras del serrano del sur de España son genéticamente distintas al resto del Mediterráneo y la población de la costa del sureste español pertenece a una población genéticamente mezclada.

3. Se observó una fuerte correlación entre la estructura genética de las poblaciones del serrano y la oceanografía de la zona en el Mediterráneo occidental, lo que indica que la estructura de las especies marinas podría no depender tanto de la duración de la fase larvaria u otras características biológicas, sino más bien de procesos oceanográficos, tales como las corrientes y las barreras oceanográficas.

4. Mediante el análisis de paternidad se demostró que aproximadamente el 15% de los juveniles de *Tripterygion delaisi* en Blanes son autoreclutas. Esto demuestra que, a pesar de que la estructuración de la población es baja y la conectividad es alta, hay una retención larvaria considerable. Las medidas directas de dispersión son esenciales para la comprensión de los patrones de conectividad en diferentes hábitats marinos y muestran el grado de retención y auto-reposición incluso en especies con alta conectividad y flujo génico.

5. Para *Tripterygion delaisi*, el análisis de hermandad permitió una evaluación directa del potencial de dispersión de la especie. Los hermanos se pueden encontrar a unos 12 kilómetros de distancia, proporcionando una evidencia directa de la conectividad entre localidades separadas por playas, que son habitats que los adultos no pueden cruzar.

6. Como resultado de los tres estudios (en *E. marginatus, S. cabrilla* y *T. delaisi*) que midieron la conectividad a diferentes escalas, se ha hecho evidente que la dispersión está determinada por el entorno físico y principalmente por las características oceanográficas y geograficas de la zona.

7. Se produjo un ensamblaje *de novo* de un transcriptoma a partir de tejido cerebral de *T. delaisi* mediante ultrasecuenciación. La mayoría de los transcritos en el transcriptoma tenían una isoforma y el número de isoformas aumentó con la longitud del transcrito.

8. En *T. delaisi*, se expresaron diferencialmente más genes entre los dos fenotipos de machos que entre los machos y las hembras, lo que sugiere que la plasticidad fenotípica es un factor más importante durante el período reproductivo que el dimorfismo sexual.

9. El macho territorial sobreexpresa genes relacionados con la plasticidad sináptica y el macho secundario sobreexpresa genes implicados en la diferenciación y el desarrollo, lo que

podría significar que el macho secundario, generalmente de menor tamaño, aún no está del todo desarrollado.

10. RNA-seq es una nueva técnica que se puede utilizar para la detección de SNPs y el desarrollo de marcadores para el genotipado a gran escala en especies no-modelo. Debido a su alto coste, se debe elegir un tamaño de muestra eficiente. Se ha demostrado que la probabilidad de descartar loci informativos, debido a la falta de polimorfismo, se reduce con el uso de siete individuos de *T. delaisi* en RNAseq y que más de 12 individuos no mejora considerablemente la predicción de frecuencias alélicas reales en la población y es, por lo tanto, el tamaño óptimo de muestra.

7. REFERENCES

REFERENCES

Almany GR, Berumen ML, Thorrold SR, Planes S, Jones GP (2007) Local replenishment of coral reef fish populations in a marine reserve. *Science (New York, N.Y.)*, **316**, 742–4.

Almany GR, Connolly SR, Heath DD, Hogan JD, Jones GP, McCook LJ, Mills M, Pressey RL, Williamson DH (2009) Connectivity, biodiversity conservation and the design of marine reserve networks for coral reefs. *Coral Reefs*, **28**, 339–351.

Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, **172**, 2567–82.

Andrello M, Mouillot D, Beuvier J, Albouy C, Thuiller W, Manel S (2013) Low Connectivity between Mediterranean Marine Protected Areas: A Biophysical Modeling Approach for the Dusky Grouper *Epinephelus marginatus* (JG Hiddink, Ed,). *PLoS ONE*, **8**, e68564.

Arnold AP, Xu J, Grisham W, Chen X, Kim Y-H, Itoh Y (2004) Minireview: Sex chromosomes and brain sexual differentiation. *Endocrinology*, **145**, 1057–62.

Aubin-Horth N, Landry CR, Letcher BH, Hofmann H a (2005) Alternative life histories shape brain gene expression profiles in males of the same population. *Proceedings. Biological sciences / The Royal Society*, **272**, 1655–62.

Aubin-Horth N, Renn SCP (2009) Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. *Molecular Ecology*, **18**, 3763–80.

Bar-Or C, Czosnek H, Koltai H (2007) Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends in Genetics : TIG*, **23**, 200–7.

Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR (2013) Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 1387–92.

Baruch E, Weller JI (2008) Estimation of the number of SNP genetic markers required for parentage verification. *Animal Genetics*, **39**, 474–9.

Berger A, Roberts MA (2005) Dietary Effects of Arachidonate-Rich Fungal Oil and Fish Oil on Murine Hippocampal Gene Expression. In: *Unraveling Lipid Metabolism with Microarrays* (eds Berger A, Roberts MA), pp. 69–94. Marcel Dekker, New York.

Bernardi G, Beldade R, Holbrook SJ, Schmitt RJ (2012) Full-sibs in cohorts of newly settled coral reef fishes. (SCA Ferse, Ed,). *PloS One*, **7**, e44953.

Berumen ML, Almany GR, Planes S, Jones GP, Saenz-Agudelo P, Thorrold SR (2012) Persistence of self-recruitment and patterns of larval connectivity in a marine protected area network. *Ecology and Evolution*, **2**, 444–52.

Botsford LW (1997) The Management of Fisheries and Marine Ecosystems. *Science*, **277**, 509–515.

Botsford LW, White JW, Coffroth M-A, Paris CB, Planes S, Shearer TL, Thorrold SR, Jones GP (2009) Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. *Coral Reefs*, **28**, 327–337.

Bowler DE, Benton TG (2005) Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews*, **80**, 205–225.

Bradbury IR, Bentzen P (2007) Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations. *Marine Ecology Progress Series*, **340**, 245–257.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, **18**, 249–256.

Burmeister SS, Jarvis ED, Fernald RD (2005) Rapid behavioral and genomic responses to social opportunity. *PLoS biology*, **3**, e363.

Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology*, **25**, 169–93.

Calderón I, Ortega N, Duran S, Becerro M, Pascual M, Turon X (2007) Finding the relevant scale: clonality and genetic structure in a marine invertebrate (Crambe crambe, Porifera). *Molecular Ecology*, **16**, 1799–810.

Carreras-Carbonell J, Macpherson E, Pascual M (2005) Rapid radiation and cryptic speciation in mediterranean triplefin blennies (Pisces: Tripterygiidae) combining multiple genes. *Molecular Phylogenetics and Evolution*, **37**, 751–61.

Carreras-Carbonell J, Macpherson E, Pascual M (2006) Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. *Molecular Ecology*, **15**, 3527–39.

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.

Carruth LL, Reisert I, Arnold AP (2002) Sex chromosome genes directly affect brain sexual differentiation. *Nature Neuroscience*, **5**, 933–4.

Christie MR, Johnson DW, Stallings CD, Hixon M a (2010) Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. *Molecular Ecology*, 1042–1057.

Coll M, Piroddi C, Steenbeek J, Kaschner K, Ben Rais Lasram F, Aguzzi J, Ballesteros E, Bianchi CN, Corbera J, Dailianis T, Danovaro R, Estrada M, Froglia C, Galil BS, Gasol JM, Gertwagen R, Gil J, Guilhaumon F *et al.* (2010) The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. *PloS One*, **5**, e11842.

Cowen RK (2000) Connectivity of Marine Populations: Open or Closed? *Science*, **287**, 857–859.

Cowen RK, Gawarkiewicz GG, Pineda J, Thorrold SR, Werner FE (2007) Population connectivity in marine systems: an overview. *Oceanography*, **20**, 14–21.

Cowen RK, Sponaugle S (2009b) Larval Dispersal and Marine Population Connectivity. *Annual Review of Marine Science*, **1**, 443–466.

Dibattista JD, Feldheim KA, Garant D, Gruber SH, Hendry AP (2009) Evolutionary potential of a large marine vertebrate: quantitative genetic parameters in a wild population. *Evolution; International Journal of Organic Evolution*, **63**, 1051–67.

Domingues VS, Faria C, Stefanni S, Santos RS, Brito A, Almada VC (2007) Genetic divergence in the Atlantic-Mediterranean Montagu's blenny, *Coryphoblennius galerita* (Linnaeus 1758) revealed by molecular and morphological characters. *Molecular Ecology*, **16**, 3592–605.

Everett M V, Grau ED, Seeb JE (2011) Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. *Molecular Ecology Resources*, **11 Suppl 1**, 93–108.

Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldon T, Guigo R, Sumner S (2013) Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome biology*, **14**, R20.

Galarza JA, Carreras-Carbonell J, Macpherson E, Pascual M, Roques S, Turner GF, Rico C (2009) The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1473–8.

García-Merchán VH, Robainas-Barcia A, Abelló P, Macpherson E, Palero F, García-Rodríguez M, Gil de Sola L, Pascual M (2012) Phylogeographic patterns of decapod crustaceans at the Atlantic–Mediterranean transition. *Molecular Phylogenetics and Evolution*, **62**, 664–672.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources*, **10**, 915–34.

Gerlach G, Atema J, Kingsford MJ, Black KP, Miller-Sims V (2007) Smelling home can prevent dispersal of reef fish larvae. *Proceedings of the National Academy of Sciences*, **104**, 858–863.

Ghalambor K, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, **21**, 394–407.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–52.

Guidetti P, Cattaneo-Vietti R (2002) Can mineralogical features influence distribution patterns of fish? A case study in shallow Mediterranean rocky reefs. *Journal of the Marine Biological Association of the UK*, **82**, 1043–1044.

Guiry A, Flynn D, Hubert S, O'Keeffe AM, LeProvost O, White SL, Forde PF, Davoren P, Houeix B, Smith TJ, Cotter D, Wilkins NP, Cairns MT (2010) Testes and brain gene expression in precocious male and adult maturing Atlantic salmon (*Salmo salar*). *BMC Genomics*, **11**, 211.

Halpern BS, Warner RR (2003) Matching marine reserve design to reserve objectives. *Proceedings. Biological sciences / The Royal Society*, **270**, 1871–8.

Harrison HB, Williamson DH, Evans RD, Almany GR, Thorrold SR, Russ GR, Feldheim KA, van Herwerden L, Planes S, Srinivasan M, Berumen ML, Jones GP (2012) Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Current Biology*, **22**, 1023–8.

Hartl DL, Clark AG (2007) *Principles of Population Genetics*. Sunderland: Sinauer associates.

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources*, **11 Suppl 1**, 150–61.

Hedgecock D, Barber P, Edmands S (2007) Genetic approaches to measuring connectivity. *Oceanography*, **20**, 70–79.

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

Hedrick P (2009) *Population Genetics and Ecology*. Princeton University Press.

Heemstra PC, Randall JE (1993) Groupers of the World (Family Serranidae, Subfamily Epinephelinae). An annotated and illustrated catalogue of the grouper, rockcod, hind, coral

grouper and lyretail species known to date. FAO Species Catalogue Vol. 16. *FAO Fish. Synop*, 125.

Hereu B, Diaz D, Pasqual J, Zabala M, Sala E (2006) Temporal patterns of spawning of the dusky grouper *Epinephelus marginatus* in relation to environmental factors. *Marine Ecology Progress Series*, **325**, 187–194.

Hofmann HA, Fernald RD (2000) Social status controls somatostatin neuron size and growth. *The Journal of Neuroscience : the official Journal of the Society for Neuroscience*, **20**, 4740–4.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11 Suppl 1**, 117–22.

Horne JB, van Herwerden L (2013) Long-term panmixia in a cosmopolitan Indo-Pacific coral reef fish and a nebulous genetic boundary with its broadly sympatric sister species. *Journal of Evolutionary Biology*, **26**, 783–99.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–9.

Huey RB, Kingsolver JG (1989) Evolution of thermal sensitivity of ectotherm performance. *Trends in Ecology & Evolution*, **4**, 131–5.

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings. Biological sciences / The Royal Society*, **272**, 1525–34.

Hutchings JA (2000) Collapse and recovery of marine fishes. *Nature*, **406**, 882–5.

Immler S, Mazzoldi C, Rasotto MB (2004) From sneaker to parental male: change of reproductive traits in the black goby, *Gobius niger* (Teleostei, Gobiidae). *Journal of Experimental Zoology. Part A, Comparative experimental biology*, **301**, 177–85.

Jones GP, Almany GR, Russ GR, Sale PF, Steneck RS, Oppen MJH, Willis BL (2009) Larval retention and connectivity among populations of corals and reef fishes: history, advances and challenges. *Coral Reefs*, **28**, 307–325.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, **12**, 2511–23.

Jones GP, Srinivasan M, Almany GR (2007) Population Connectivity and Conservation of Marine Biodiversity. *Oceanography*, **20**, 100.

Jonge J, Ruiter AJH, Hurk R (1989) Testis-testicular gland complex of two Tripterygion species (Blennioidei, Teleostei): differences between territorial and non-territorial males. *Journal of Fish Biology*, **35**, 497–508.

Jonge J, Videler JJ (1989) Differences between the reproductive biologies of *Tripterygion tripteronotus* and *T. delaisi* (Pisces, Perciformes, Tripterygiidae): the adaptive significance of an alternative mating strategy and a red instead of a yellow nuptial colour. *Marine Biology*, **100**, 431–437.

Kanno Y, Vokoun JC, Letcher BH (2010) Sibship reconstruction for inferring mating systems, dispersal and effective population size in headwater brook trout (*Salvelinus fontinalis*) populations. *Conservation Genetics*, **12**, 619–628.

Kanoh Y (2000) Reproductive success associated with territoriality , sneaking , and grouping in male rose bitterlings , *Rhodeus ocellatus* (Pisces : Cyprinidae). *Environmental Biology of Fishes*, **57**, 143–154.

Kaplan DM, Botsford LW, O'Farrell MR, Gaines SD, Jorgensen S (2009) Model-based assessment of persistence in proposed marine protected area designs. *Ecological Applications*, **19**, 433–448.

Karsenty G, Wagner EF (2002) Reaching a genetic and molecular understanding of skeletal development. *Developmental Cell*, **2**, 389–406.

Lamichhaney S, Martinez Barrio A, Rafati N, Sundström G, Rubin C-J, Gilbert ER, Berglund J, Wetterbom A, Laikre L, Webster MT, Grabherr M, Ryman N, Andersson L (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 19345–50.

Macpherson E (2002) Large-scale species-richness gradients in the Atlantic Ocean. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **269**, 1715–1720.

Macpherson E, Raventós N (2006) Relationship between pelagic larval duration and geographic distribution in Mediterranean littoral fishes. *Marine Ecology Progress Series*, **327**, 257–265.

Maggio T, Andaloro F, Arculeo M (2006) Genetic population structure of *Epinephelus marginatus* (Pisces, Serranidae) revealed by two molecular markers. *Italian Journal of Zoology*, **73**, 275–283.

Malavasi S, Lindström K, Sundström L (2001) Behaviour and success of sneaker males in the sand goby, *Pomatoschistus minutus. Acta ethologica*, **4**, 3–9.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, **18**, 1509–1517.

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nature Reviews Genetics*, **12**, 671–82.

De Mestral LG, Herbinger CM, O'Reilly PT, Taylor EB (2012) Mating structure of an endangered population of wild Atlantic salmon (*Salmo salar*) as determined using sibship reconstruction and a novel method of sex inference. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**, 1352–1361.

Millot C (1999) Circulation in the Western Mediterranean Sea. *Journal of Marine Systems*, **20**, 423–442.

Millot C (2005) Circulation in the Mediterranean Sea. *Circulation*, **69**, 5–21.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**, 208–216.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–8.

Oliveira RF (2012) Social plasticity in fish: integrating mechanisms and function. *Journal of Fish Biology*, **81**, 2127–2150.

Olsen MT, Volny VH, Bérubé M, Dietz R, Lydersen C, Kovacs KM, Dodd RS, Palsbøll PJ (2011) A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Arctic ringed seal (*Pusa hispida hispida*). *Molecular Ecology Resources*, **11 Suppl 1**, 9–19.

Oshlack A, Wakefield MJ (2009) Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, **4**, 14.

Palero F, Abelló P, Macpherson E, Gristina M, Pascual M (2008) Phylogeography of the European spiny lobster (*Palinurus elephas*): Influence of current oceanographical features and historical processes. *Molecular Phylogenetics and Evolution*, **48**, 708–17.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, **13**, 146–158.

Patarnello T, Volckaert F a MJ, Castilho R (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, **16**, 4426–44.

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST Databases. *Genome Research*, **9**, 167–174.

Pigliucci M (2005) Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology & Evolution*, **20**, 481–6.

Pigliucci M (2007) Do we need an extended evolutionary synthesis? *Evolution; International Journal of Organic Evolution*, **61**, 2743–9.

Pinsky ML, Palumbi SR, Andréfouët S, Purkis SJ (2012) Open and closed seascapes: Where does habitat patchiness create populations with high fractions of self-recruitment? *Ecological Applications*, **22**, 1257–1267.

Planes S (2002) Biogeography and larval dispersal inferred from population genetic analysis. In: *Coral Reef Fishes. Dynamics and Diversity in a Complex Ecosystem* , pp. 201–220. Academic Press.

Planes S, Jones GP, Thorrold SR (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5693–7.

Raventós N, Macpherson E (2001) Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Marine Biology*, **138**, 1115–1120.

Reichard M, Smith C, Jordan WC (2004) Genetic evidence reveals density-dependent mediated success of alternative mating behaviours in the European bitterling (*Rhodeus sericeus*). *Molecular Ecology*, **13**, 1569–78.

Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (Coregonus spp. Salmonidae). *Molecular Ecology*, **19 Suppl 1**, 115–131.

Renn SCP, Aubin-Horth N, Hofmann H a (2008) Fish and chips: functional genomics of social plasticity in an African cichlid fish. *The Journal of Experimental Biology*, **211**, 3041–56.

Reñones O, Piñeiro C, Mas X, Goñi R (2007) Age and growth of the dusky grouper *Epinephelus marginatus* (Lowe 1834) in an exploited population of the western Mediterranean Sea. *Journal of Fish Biology*, **71**, 346–362.

Rios-Cardenas O, Webster MS (2008) A molecular genetic examination of the mating system of pumpkinseed sunfish reveals high pay-offs for specialized sneakers. *Molecular Ecology*, **17**, 2310–20.

Roberge C, Blanchet S, Dodson JJ, Guderley H, Bernatchez L (2008) Disturbance of social hierarchy by an invasive species: a gene transcription study. *PloS One*, **3**, e2408.

Sabates A, Zabala M, García-Rubies A (2003) Larval fish communities in the Medes Islands Marine Reserve (North-west Mediterranean). *Journal of Plankton Research*, **25**, 1035–1046.

Sadovy Y, Colin PL, Domeier ML (1994) Aggregation and spawning in the tiger grouper, *Mycteroperca tigris* (Pisces: Serranidae). *Copeia*, **2**, 511–516.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2009) Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. *Molecular Ecology*, **18**, 1765–76.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2011) Connectivity dominates larval replenishment in a coastal reef fish metapopulation. *Proceedings. Biological sciences / The Royal Society*, **278**, 2954–61.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2012) Patterns and persistence of larval retention and connectivity in a marine fish metapopulation. *Molecular Ecology*, **21**, 4695–705.

Schulze A, Downward J (2001) Navigating gene expression using microarrays--a technology review. *Nature Cell Biology*, **3**, E190–5.

Selkoe KA, Gaines SD, Caselle JE, Warner RR (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology*, **87**, 3082–94.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–29.

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

Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. *The Biological bulletin*, **216**, 373–85.

Sneddon LU, Schmidt R, Fang Y, Cossins AR (2011) Molecular correlates of social dominance: a novel role for ependymin in aggression. *PloS One*, **6**, e18181.

Sponaugle S, Cowen RK, Shanks A, Morgan SG, Leis JM, Pineda J, Boehlert GW, Kingsford MJ, Lindeman KC, Grimes C, Munro JL (2002) Predicting self-recruitment in marine populations: biophysical correlates and mechanisms. *Bulletin of Marine Science*, **70**, 341–375.

Stenseth NC, Jorde PE, Chan K-S, Hansen E, Knutsen H, André C, Skogen MD, Lekve K (2006) Ecological and genetic impact of Atlantic cod larval drift in the Skagerrak. *Proceedings. Biological sciences / The Royal Society*, **273**, 1085–92.

Storer CG, Pascal CE, Roberts SB, Templin WD, Seeb LW, Seeb JE (2012) Rank and order: evaluating the performance of SNPs for individual assignment in a non-model organism. *PloS One*, **7**, e49018.

Taborsky M (2008) Alternative reproductive tactics in fish. In: *Alternative reproductive tactics: an integrative approach* (ed Oliveira RF, Taborsky M BH), pp. 251–299. Cambridge University Press, Cambridge.

Torcu-Koc H, Turker-Cakir D, Dulcic J (2004) Age, growth and mortality of the comber, Serranus cabrilla (Serranidae) in the Edremit Bay (NW Aegean Sea, Turkey). *Cybium*, **28**, 19–25.

Trainor BC, Hofmann HA (2006) Somatostatin regulates aggressive behavior in an African cichlid fish. *Endocrinology*, **147**, 5119–25.

Trainor BC, Hofmann HA (2007) Somatostatin and somatostatin receptor gene expression in dominant and subordinate males of an African cichlid fish. *Behavioural Brain Research*, **179**, 314–20.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution*, **34**, 275–305.

Wan Q-H, Wu H, Fujihara T, Fang S-G (2004) Which genetic marker for which conservation genetics issue? *Electrophoresis*, **25**, 2165–76.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57–63.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–39.

White TA, Fotherby HA, Stephens PA, Hoelzel AR (2011) Genetic panmixia and demographic dependence across the North Atlantic in the deep-sea fish, blue hake (*Antimora rostrata*). *Heredity*, **106**, 690–9.

White C, Selkoe K, Watson J, Siegel DA, Zacherl DC, Toonen RJ (2010) Ocean currents help explain population genetic structure. *Proceedings. Biological sciences / The Royal Society*, **277**, 1685–1694.

Wirtz P (1978) The behaviour of the Mediterranean Tripterygion species (Pisces, Blennioidei). *Zeitschrift fur Tierpsychologie*, **48**, 142–174.

De Wit P, Palumbi SR (2012) Transcriptome-wide polymorphisms of red abalone (Haliotis rufescens) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, **22**, 2884–97.

Wright S (1969) *Evolution and the genetics of populations: Vol. 2. The theory of gene frequencies*. University of Chicago Press.

Zabala M, Louisy P, García-Rubies A, Gràcia V (1997) Socio-behavioural context of reproduction in the Mediterranean dusky grouper Epinephelus marginatus (Lowe,

1834)(Pisces, Serranidae) in the Medes Islands marine reserve (NW Mediterranean, Spain). *Scientia Marina*, **61**, 79–98.

Zucker RS, Regehr WG (2002) Short-term synaptic plasticity. *Annual Review of Physiology*, **64**, 355–405.

Informe dels directors sobre el factor d'impacte de les publicacions derivades de la tesi presentada per Celia Schunter

Val a dir que CAP INFORMACIÓ d'aquestes publicacions ha estat ni serà utilitzada en altres Tesis Doctorals.

Publicació 1: Genetic connectivity pattern in an endangered species: The dusky grouper (Epinephelus marginatus)

C. Schunter, J. Carreras-Carbonell, S. Planes, E. Sala, E. Ballesteros, M. Zabala, J.G. Harmelin, M. Harmelin-Vivien, E. Macpherson and M. Pascual

Journal of Experimental Marine Biology and Ecology 401: 126-133 (2011)

Tasques de la doctoranda :

En aquest article la doctoranda ha contribuït obtenint mostres de diferents poblacions (Tunisia i Senegal) per tal d'augmentar la zona analitzada de l'espècie. Ha optimitzat i fet totes les amplificacions de tots els loci microsatèl·lits per a totes les poblacions, ha fet les anàlisis estadístiques pertinents i ha redactat la primera versió del manuscrit i les modificacions del mateix demanades pels revisors de la revista.

Journal of Experimental Marine Biology and Ecology té un índex d'impacte de 2.263 (SCI 2012) i es troba al segon quartil (lloc 27 de 100 revistes) a l'àrea de "MARINE & FRESHWATER BIOLOGY".

Publicació 2: Matching genetics with oceanography: directional gene flow in a Mediterranean fish species

C. Schunter, J. Carreras-Carbonell, E. Macpherson, J. Tintoré, E. Vidal-Vijande, A. Pascual, P. Guidetti and M. Pascual

Molecular Ecology 20: 5167-5181 (2011)

Tasques de la doctoranda :

En aquest article la doctoranda ha fet les anàlisis estadístiques pertinents i ha redactat la primera versió del manuscrit i les modificacions del mateix demanades pels revisors de la revista.

Molecular Ecology té un índex d'impacte de 6.275 (SCI 2012) i es troba al primer quartil (lloc 10 de 136 revistes) a l'àrea de "ECOLOGY".

Publicació 3: Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics

C. Schunter, S. Vollmer, E. Macpherson, M. Pascual

Article enviat pendent de decisió en primera revisió

Tasques de la doctoranda :

En aquest article la doctoranda ha contribuït obtenint les mostres, fent les extraccions de RNA, les llibreries i tots els anàlisis bioinformàtics des de l'assemblatge del transcriptoma fins a la quantificació i identificació dels gens expressats de manera diferencial. Ha redactat la primera versió del manuscrit i les modificacions del mateix discutides entre els altres autors.

Publicació 4: SNP development from RNA-seq data in a non-model fish: how many individuals are needed for accurate allele frequency prediction?

C. Schunter, J. C. Garza, E. Macpherson and M. Pascual

Molecular Ecology Resources (2013) doi: 10.1111/1755-0998.12155

Tasques de la doctoranda :

En aquest article la doctoranda ha fet totes les anàlisis de laboratori i bioinformàtiques fins a identificar els SNPs, genotipar tots els individus i fer simulacions per tal de poder avaluar el número d'individus necessari per tal de predir de manera acurada els millors marcadors i la freqüència en les poblacions. Ha redactat la primera versió del manuscrit i les modificacions del mateix demanades pels revisors de la revista.

Molecular Ecology Resources té un índex d'impacte de 7.432 (SCI 2012) i es troba al primer quartil (lloc 7 de 136 revistes) a l'àrea de "ECOLOGY".

Publicació 5: Retention and fish larval dispersal potential on a highly connected open coast line

C. Schunter, M. Pascual, JC. Garza, N. Raventos and E. Macpherson

Article enviat pendent de decisió en primera revisió

Tasques de la doctoranda :

En aquest article la doctoranda ha fet tots els mostrejos, genotipat tots els individus amb SNPs i realitzat les anàlisis estadístiques pertinents. Ha redactat la primera versió del manuscrit i les modificacions del mateix discutides entre els altres autors.

Barcelona, 5 de setembre de 2013

Els Directors de la Tesi:

Marta Pascual Berniola

Enrique Macpherson Mayol